

Applying N-of-1 GAMMs to Quantify Pitch Count Associations with Fastball

Metrics and Biomechanics: A Comparative Modeling Study

BY

Garrett Chaz Fernandez

A Dissertation Submitted to the Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND

SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Biomedical Engineering

May, 2026

Winston-Salem, North Carolina

Approved By:

Kristen F. Nicholson, PhD, Advisor and Chair

Garrett S. Bullock, PT, DPT, DPhil

Patrick Ward, PhD

Tanner Filben, PhD

Ryan S. McGinnis, PhD

Joel D. Stitzel, PhD

DEDICATION

I dedicate this work to my parents, Eddie and Windie Fernandez, whose love, support, and encouragement have shaped every step of this journey; to my brothers, Dakota and Braylon, for always being in my corner; and to my grandfather, Kieth, for helping nurture the curiosity that led me here. Thank you all for always listening, encouraging me to think bigger, and giving me the confidence to chase ideas wherever they led. This dissertation reflects the support, belief, and inspiration you have given me throughout my life.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my advisor and chair, Dr. Kristen F. Nicholson, for her guidance, support, and mentorship throughout this process. Thank you for your time, encouragement, and thoughtful feedback over the course of this work. Your guidance has been invaluable to my development as a researcher and to the completion of this dissertation.

I would also like to extend special thanks to Dr. Garrett S. Bullock for his mentorship, guidance, and continued support throughout this process. Your investment in this work has meant a great deal to me. Thank you for the time, effort, and thought you contributed over the years, and for always being willing to talk through ideas, provide feedback, and help strengthen both the research and my growth as a scholar.

I am also grateful to the other members of my dissertation committee, Dr. Patrick Ward, Dr. Tanner Filben, Dr. Ryan S. McGinnis, and Dr. Joel D. Stitzel. Thank you for your expertise, support, and thoughtful contributions to this work. I greatly appreciate the time and insight each of you provided throughout this process.

I would also like to extend my appreciation to the Wake Forest Analytics team for their help and contributions throughout this project. Thank you for the time, effort, and collaboration that supported this work over the past four years.

Finally, I would like to thank the coaches, players, and staff of the Wake Forest baseball team. This work would not have been possible without your participation, support, and willingness to contribute to this research over the last four years. I am deeply grateful for the trust, effort, and collaboration you gave throughout this process.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xiii
ABSTRACT	xiv
CHAPTER I: INTRODUCTION	1
Background	2
Chapter Summaries	14
References	16
CHAPTER II: N-of-1 Modeling of Rolling Pitch Count Associations with Fastball Performance Metrics in Collegiate Baseball Pitchers	22
Abstract	23
Introduction	24
Methods	26
Study Design	26
Study Participants and Eligibility Criteria	26
Sample Size and Data Sufficiency	27
Data Sources and Variable Collection	28
Variable Definitions	29
Missing Data	30
Data Processing and Feature Construction	31

N-of-1 Model Development.....	31
Model Evaluation, Diagnostics, and Sensitivity Analyses.....	33
Statistical Analysis.....	34
Results.....	35
Descriptive Statistics	35
Overall N-of-1 Model Performance Across Fastball Metrics.....	38
Fastball Velocity Results	40
Fastball Spin Rate Results	43
Fastball Vertical Break Results.....	44
Residual Autocorrelation, Stability, and Sensitivity Analyses	45
Bootstrap Uncertainty Summaries	47
Discussion	48
Limitations.....	51
Applied Implications and Future Directions.....	52
Conclusion	52
References	54
CHAPTER III: N-of-1 Modeling of Rolling Fastball Pitch Count Associations with Pitching Biomechanics in Collegiate Baseball Pitchers	59
Abstract	60
Introduction.....	61
Methods.....	64

Study Design	64
Study Participants and Eligibility Criteria.....	64
Sample Size and Data Sufficiency.....	65
Data Sources and Variable Collection	65
Variable Definitions	66
Missing Data	71
Data Processing and Feature Construction	72
N-of-1 Model Development.....	73
Model Evaluation, Diagnostics, and Sensitivity Analyses.....	75
Statistical Analysis.....	76
Results.....	77
Descriptive Statistics	77
Biomechanics Model Performance Overview	78
Outcome-Specific Biomechanics Model Performance	82
Pitcher-Specific Heterogeneity	84
Residual Autocorrelation, Stability, and Sensitivity Analyses	85
Bootstrap Uncertainty Summaries	89
Discussion	91
Limitations	93
Applied Implications and Future Directions.....	94
Conclusion	94

References	96
CHAPTER IV: Comparative Evaluation of Group-Based and Individualized N-of-1 Models for Pitch Count Associations in Baseball Pitchers	
Abstract	104
Introduction	105
Methods.....	107
Study Design.....	107
Data Sources and Outcome Selection	107
Analytic Unit and Data Structure.....	108
Model Development.....	109
Statistical Analysis for Model Comparison	111
Results.....	112
Comparison Set Overview	112
Velocity Model Comparison	112
Biomechanics Model Comparison	114
Residual Autocorrelation Comparison	117
Discussion	119
Limitations	123
Applied Implications and Future Directions.....	124
Conclusion	125
References	126

CHAPTER V: Practical Applications of Individualized Pitch Count Models for Monitoring Collegiate Baseball Pitchers	131
Abstract	132
Introduction	133
Monitoring Implementation	135
Source of Individualized Predictions	135
Construction of Monitoring Inputs	136
CUSUM Implementation and Tuning	137
Signal Definition	139
Applied Monitoring Results	140
Cohort-level Monitoring Summary	140
Representative Pitcher Examples	142
Discussion	145
Limitations	148
Future Directions	149
Conclusion	150
References	151
CHAPTER VI: SUMMARY OF RESEARCH	154
Appendix A – Performance Models Prediction Curves.....	156
Appendix B - Biomechanic Outcomes Descriptive Statistics	260
Appendix C – Additional CUSUM Plots.....	261

Appendix D – N-of-1 Demo Code	274
Appendix E – Diagnostic Plot Sample	319
CURRICULUM VITAE	355
Education.....	355
Research Experience.....	355
Publications	356
Manuscripts Submitted / Under Journal Consideration.....	356
Manuscripts in Preparation	356
Conference and Research Presentations.....	357
Technical Skills	357
Professional Memberships.....	358
Honors and Awards	358

LIST OF TABLES

Table I: Descriptive Statistics for Fastball Dataset by Season	24
Table II: Overall N-of-1 Model Performance Across Fastball Metrics	39
Table III: Residual autocorrelation summary for individualized fastball models	45
Table IV: Summary of RMSE-Based Stability and Sensitivity Across Fastball Outcomes	47
Table V: Bootstrap uncertainty summaries across pitcher-specific N-of-1 performance models by outcome.	48
Table VI: Mean performance of individualized biomechanics models by outcome and season.....	79
Table VI (continued): Mean performance of individualized biomechanics models by outcome and season	80
Table VII: Residual autocorrelation summaries for final individualized biomechanics models by outcome	86
Table VIII: Stability and sensitivity summaries for final individualized biomechanics models by outcome.....	88
Table IX: Bootstrap uncertainty summaries across pitcher-specific N-of-1 biomechanics models by outcome.	90
Table X: Comparison of velocity model performance for individualized N-of-1 and group- based models.	113
Table XI: Comparison of biomechanics model performance for individualized N-of-1 and group-based models across selected representative outcomes.....	115
Table XII: Cohort-level summary of the final appearance-level CUSUM monitoring sample.	141

LIST OF FIGURES

Figure 1: Distribution of rolling pitch count across the 2023–2025 seasons.....	37
Figure 2: Fastball pitches per appearance by pitcher role and setting.....	38
Figure 3: Distribution of adjusted R^2 across individualized fastball performance models.....	40
Figure 4: Season-specific distribution of adjusted R^2 for individualized fastball velocity models.....	41
Figure 5: Individualized fastball velocity curves across rolling pitch count.....	42
Figure 6: Season-specific distribution of adjusted R^2 for individualized fastball spin rate models.....	43
Figure 7: Season-specific distribution of adjusted R^2 for individualized fastball vertical break models.....	44
Figure 8. Mean AR(1) values by pitcher-season for individualized fastball velocity models.	45
Figure 9: Relative variability of fastball biomechanics outcomes across appearances.....	78
Figure 10: Overall performance of individualized biomechanics models by season.....	82
Figure 11: Distribution of adj. R^2 values for final individualized biomechanics models.....	84
Figure 12: Heat map of adjusted R^2 values for each biomechanics outcome.....	85
Figure 13: Pitcher-level velocity model performance comparing individualized N-of-1 and group-based approaches across the 2023, 2024, and 2025 seasons.....	114
Figure 14: Pitcher-level adj. R^2 for selected biomechanics outcomes comparing individualized N-of-1 and group-based approaches across the 2024 and 2025.....	116
Figure 15: Pitcher-level RMSE for selected biomechanics outcomes comparing individualized N-of-1 and group-based approaches across the 2024 and 2025.....	117

Figure 16: Pitcher-level residual autocorrelation for velocity models across modeling stages and seasons.....	118
Figure 17: Pitcher-level residual autocorrelation for maximum elbow varus torque models across modeling stages and seasons.....	119
Figure 18: Appearance-level CUSUM for Player 8 under the final selected monitoring specification	143
Figure 19: Appearance-level CUSUM for Player 13 under the final selected monitoring specification.....	144
Figure 20: Appearance-level CUSUM for Player 25 under the final selected monitoring specification.....	145

LIST OF ABBREVIATIONS

Abbreviation	Definition
ACF	Autocorrelation function
Adj. R^2	Adjusted coefficient of determination
AIC	Akaike information criterion
AR(1)	First-order autoregressive correlation structure
BIC	Bayesian information criterion
BR	Ball release
CI	Confidence interval
CUSUM	Cumulative sum
FC	Foot contact
GAMM	Generalized additive mixed model
HRV	Heart rate variability
IQR	Interquartile range
LASSO	Least absolute shrinkage and selection operator
MER	Maximum external rotation
MPH	Miles per hour
MSE	Mean squared error
PMM	Predictive mean matching
R^2	Coefficient of determination
RMSE	Root mean squared error
RPM	Revolutions per minute
SD	Standard deviation

ABSTRACT

Workload monitoring remains a persistent challenge in baseball, particularly in collegiate pitchers exposed to repeated throwing demands across games, practices, bullpens, and travel. Although pitch count is one of the most widely used workload indicators in applied baseball settings, it does not fully capture the broader context in which performance and biomechanical responses occur, and traditional group-based approaches may obscure meaningful heterogeneity in athlete-specific workload-response patterns. Therefore, the overall objective of this dissertation was to apply individualized N-of-1 generalized additive mixed models to investigate rolling pitch count associations in collegiate baseball pitchers across performance, biomechanics, comparative modeling, and applied monitoring contexts. Across the 2023, 2024, and 2025 seasons, individualized fastball models demonstrated that rolling pitch count associations could be characterized at the pitcher-season level, with velocity providing the strongest and most practically interpretable signal, while spin rate and vertical break functioned as complementary outcomes. Across the 2024 and 2025 seasons, individualized biomechanics models further showed that workload-related response could also be expressed mechanically, although model performance varied across biomechanical outcomes and pitcher-seasons. Comparative analyses demonstrated that individualized N-of-1 models outperformed group-based models across both performance and biomechanics domains, supporting pitcher-specific modeling when the goal is to characterize within-athlete workload-response patterns. Finally, individualized velocity models were translated into an appearance-level CUSUM monitoring framework, showing that pitcher-specific expected-versus-observed surveillance could identify heterogeneous in-season deviation patterns. Collectively, these findings support individualized modeling as a more informative framework for evaluating

workload-response patterns in collegiate baseball pitchers than reliance on population-average approaches alone.

This research is comprised of four parts:

- The first part aims to quantify within-pitcher associations between rolling pitch count and fastball performance metrics, including velocity, spin rate, and vertical break, in collegiate baseball pitchers.
- The second part aims to quantify within-pitcher associations between rolling pitch count and key pitching biomechanics in collegiate baseball pitchers.
- The third part aims to compare individualized N-of-1 models with traditional population-level mixed-effects models in explanatory performance and residual independence.
- The fourth part aims to apply cumulative-sum (CUSUM) monitoring procedures to individualized model outputs to detect early, systematic declines in performance across a season.

CHAPTER I: INTRODUCTION

Baseball pitching is a high-demand, repetitive athletic task performed within a dynamic competitive environment shaped by games, practices, bullpen sessions, recovery periods, and travel.¹⁻³ Across the course of a season, pitchers are exposed to repeated throwing loads that may influence performance, movement patterns, and overall readiness to compete. In applied baseball settings, coaches, sports medicine clinicians, and performance staff are therefore tasked with making frequent decisions about workload, recovery, and return to play. These decisions are often made with the intent of balancing performance optimization with athlete health, yet the process is rarely straightforward. Although throwing exposure can be tracked in several ways, accurately determining how a pitcher is responding to that exposure remains a persistent challenge.¹⁻⁴ The same workload may be well tolerated by one pitcher, but associated with underperformance, altered mechanics, or delayed recovery in another.

This reflects a broader challenge in sport science and athlete management: it is generally easier to quantify how much work an athlete has performed than to determine how that athlete has responded to it. Measures of exposure, such as pitch count, provide useful information about how much throwing has occurred, but they do not directly explain how the athlete is adapting to that demand. In baseball, terms such as fatigue, readiness, and recovery are often used to interpret observed changes in pitcher status, yet these constructs are not directly interchangeable and are not easily captured by any single variable. A pitcher's response to accumulated throwing exposure may depend on multiple interacting factors, including recent workload history, spacing between appearances, internal physiological state, travel-related stress, and the broader competitive context. As a result, practical workload questions in baseball rarely concern exposure alone. More

often, the central question is whether a given pitcher is responding normally, favorably, or unfavorably to the demands being placed upon him.

These challenges are particularly relevant in collegiate baseball, where pitchers operate within dense schedules and individualized roles, but where workload decisions must still be made quickly and with imperfect information. Traditional workload approaches have often relied on population-average logic or generalized thresholds, which can be useful for broad guidance but may not fully reflect the response pattern of a specific athlete.^{2,5,6} At the same time, modern baseball environments generate increasingly rich longitudinal data, including pitch-level ball metrics, biomechanical measures, wearable-derived recovery information, and contextual variables related to rest and travel. This creates an opportunity to move beyond one-size-fits-all interpretations and toward individualized modeling strategies that better reflect pitcher-specific workload-response patterns.⁶⁻⁹ Therefore, the purpose of this dissertation was to apply individualized N-of-1 generalized additive mixed models to examine associations between rolling pitch count and key performance and biomechanics outcomes in collegiate baseball pitchers, compare those individualized models with traditional group-based approaches, and explore how individualized model outputs may support applied monitoring frameworks.

Background

Workload Monitoring in Baseball Pitchers

Workload monitoring has become a central part of modern baseball performance management.¹⁻³ At its core, workload monitoring refers to the process of tracking the demands placed on athletes and evaluating how they respond over time. For pitchers, this process is especially important because throwing is both highly repetitive and biomechanically demanding, with exposure accumulating not only in games but also in practices, bullpens, live sessions, and other structured throwing activities.^{1,2,10,11} Across a

season, pitchers may experience substantial fluctuations in workload depending on role, schedule, and team context, making it difficult to rely on a single static assessment of workload status.

In applied settings, workload monitoring is often driven by two related objectives.^{1,4} The first is to support performance by identifying conditions under which pitchers are most likely to maintain or maximize desired output. The second is to reduce the likelihood of maladaptive responses, such as persistent underperformance, poor recovery, or potentially unfavorable biomechanical changes. Although these goals are often discussed together, they are not identical. A pitcher may maintain acceptable visible performance while accumulating stress in a way that emerges later. They may also show subtle performance or mechanical changes before any major issue becomes obvious. For this reason, workload monitoring is not simply about documenting exposure, but about interpreting exposure within a broader framework of response.

A major challenge within this process is that exposure is typically easier to quantify than adaptation.^{1,4,12} Coaches and practitioners can often identify how much a pitcher has thrown but may be less certain how that pitcher has tolerated the demand. This distinction is especially important in baseball because workload response is unlikely to be uniform across athletes. Individual differences in physical capacity, mechanics, recovery behavior, role expectations, and environmental context may all shape how workload is expressed. Accordingly, the practical value of workload monitoring depends not only on how well exposure is measured, but on how effectively athlete-specific responses can be interpreted.

Pitch Count as a Workload Metric

Among available workload indicators in baseball, pitch count remains one of the most common and practical.^{1,2,12} Its widespread use is largely due to its simplicity. Pitch

count is easy to record, easy to communicate, and already embedded within baseball decision-making at many levels of play. It provides a straightforward estimate of throwing volume and can be applied consistently across games and practices. As a result, pitch count often serves as the default workload metric in baseball, particularly when practitioners need a rapid and operationally feasible way to summarize recent throwing exposure.

However, the practical strengths of pitch count also highlight its limitations. Pitch count reflects how many pitches were thrown but does not directly explain how the pitcher responded to those pitches.¹⁻³ It does not, by itself, account for appearance spacing, prior workload history, recovery status, throwing effort, physiological readiness, travel demands, or differences between competition and practice environments. Two appearances with identical pitch counts may impose substantially different practical demands depending on the context in which they occurred. In that sense, pitch count is better understood as an external workload proxy than as a direct indicator of fatigue, recovery, or athlete status.¹³⁻¹⁵

These limitations do not make pitch count unhelpful; rather, they clarify the way it should be studied. Pitch count remains highly relevant because it is one of the few workload variables already available at scale in most baseball environments. The key issue is not whether pitch count should be ignored, but whether it can be interpreted more effectively when considered alongside broader contextual information. Rolling pitch count may be particularly valuable in this regard because it reflects accumulated throwing exposure across time rather than treating each appearance in isolation.^{3,12} Within a more comprehensive workload framework, pitch count can therefore remain the central exposure variable while still being interpreted as one part of a larger response system.

Fatigue in Baseball Pitching

The broader concept underlying many workload discussions in baseball is fatigue.^{4,11,16,17} In both applied and research settings, fatigue is frequently used to explain changes in performance, alterations in mechanics, or the perceived need for additional recovery. Despite its common use, however, fatigue is difficult to define precisely and even more difficult to measure directly.⁴ It is not a single observable event or universally consistent state. Rather, fatigue is better understood as a multifactorial process that may involve physiological, neuromuscular, perceptual, and behavioral components, each of which may be expressed differently depending on the athlete and context.

In baseball pitching, this complexity is especially important because fatigue is rarely identified through one definitive marker.^{1,11,16,17} Instead, fatigue-related responses are often inferred from observed changes, such as loss of velocity, altered command, changes in pitch characteristics, modified mechanics, or perceived decreases in readiness. Yet these indicators do not always occur together, and their absence does not necessarily imply that the athlete is unaffected by workload. A pitcher may maintain visible performance while compensating mechanically or may experience subtle recovery-related changes that are not immediately evident in box-score outcomes.^{10,11,18} This makes fatigue difficult to treat as a singular endpoint and suggests that its expression may depend on which outcome domain is being monitored.

The challenge becomes greater when considering individual variability. Similar workloads may not produce similar fatigue-related responses across pitchers.^{5,14} Differences in training history, role, recovery strategies, physical characteristics, and tolerance to repeated throwing may all influence how accumulated load is expressed. Even within the same pitcher, responses may vary across seasons or competitive periods.^{11,16} As a result, fatigue in pitchers should not be framed as a uniform or predictable decline that follows a single pattern. A more appropriate perspective is that

fatigue-related responses may emerge in multiple ways, and that monitoring systems must be capable of detecting a range of possible expressions.

For this dissertation, the importance of fatigue lies less in treating it as a directly measured variable and more in recognizing it as a conceptual driver of workload-response analysis. Because fatigue cannot be fully captured by any single measure, workload-related patterns must instead be inferred through repeated observation of meaningful outcomes within context. This supports the broader need for individualized, longitudinal modeling approaches that focus not only on how much a pitcher has thrown, but how that pitcher's performance and mechanics behave over time in relation to accumulated exposure.

Internal and External Load in Sport

A useful framework for understanding workload response is the distinction between internal and external load.^{13–15} In sport science, external load generally refers to the work completed or exposure accumulated by the athlete, while internal load refers to the athlete's response to that work. External load in baseball may include variables such as pitch count, throwing frequency, plate appearances, innings played or other measures of workload volume.^{1,11,13} Internal load may include physiological, perceptual, or recovery-related responses that reflect how the athlete is tolerating those demands. Although this distinction is conceptually helpful, the relationship between internal and external load is not always straightforward.

For pitchers, pitch count is a practical example of external load because it quantifies throwing exposure in an accessible way. However, an external load value does not inherently reveal whether the pitcher is well recovered, physiologically strained, or perform normally relative to expectation. Internal responses may differ even when external exposures appear similar.^{14,15} This is one reason why external load metrics alone are often

insufficient for fully characterizing workload status. They describe what the athlete has done, but not necessarily how the athlete is responding.

The distinction between internal and external load is particularly relevant in baseball because many practical questions arise at the intersection of the two. Staff may know how many pitches a pitcher has thrown but still need to determine whether that workload has been tolerated appropriately. They may also observe a change in performance and need to consider whether it reflects accumulated exposure, insufficient workload (e.g., detraining), poor recovery, contextual stressors, or normal variation. For that reason, the present dissertation treats pitch count as the central external exposure while also considering broader contextual indicators that may help explain response patterns within pitchers.

Recovery, HRV, and Wearable Monitoring

Recovery is a critical part of the workload-response process because athlete performance is influenced not only by the demands imposed, but by how effectively the athlete adapts between exposures. In baseball, recovery may be shaped by appearance frequency, game role, sleep, travel, recent workload, and broader physiological state. Because recovery is difficult to observe directly, applied sport environments increasingly rely on monitoring tools intended to provide additional insight into athlete readiness and recovery-related status. Among these tools, wearable technologies have become increasingly prominent because they allow repeated data collection outside traditional laboratory settings.

Heart rate variability (HRV) has received particular attention as one possible indicator related to autonomic function, readiness, or recovery.^{19–21} In applied settings, HRV is often interpreted as a marker that may reflect how an athlete is responding to recent training or competition demands. Wearable platforms such as WHOOP have made

this type of monitoring more accessible by embedding data collection within routine athlete use.¹⁹ For baseball organizations, this creates the possibility of incorporating recovery-related information into broader workload models without requiring extensive separate testing infrastructure.

At the same time, HRV should be interpreted cautiously.^{19,20} It is not a direct measure of fatigue, and its relationship to performance or readiness is context dependent. HRV can be influenced by multiple factors unrelated to throwing exposure, including sleep, illness, stress, measurement timing, and compliance. In addition, wearable-derived metrics may rely on proprietary processing methods that limit methodological transparency. For these reasons, HRV is best understood as one potentially informative contextual variable rather than a definitive standalone indicator. Within this dissertation, HRV was considered part of the broader workload environment when available, but not as a singular explanation for pitcher response.

Travel and the Broader Competitive Context

Workload response in collegiate baseball occurs within a broader competitive context that extends beyond throwing volume alone. One important component of this context is travel.^{1,22,23} Travel may influence athlete response through disruption of routine, altered sleep opportunities, reduced recovery quality, and changes in physical or psychological preparation.^{1,22,23} In collegiate settings, away competition can introduce additional sources of strain that are not fully captured by pitch count or appearance totals. These effects may be especially relevant when travel coincides with dense schedules, limited rest, or repeated competition demands.

Travel also highlights a broader point: workload does not occur in isolation. The same nominal throwing exposure may have different consequences depending on whether it occurs in a game or practice setting, after extended rest or short rest, at home

or away, or within a relatively stable versus disrupted weekly schedule. This means that contextual variables should not be treated as secondary details, but as part of the environment in which workload-response patterns emerge. Including travel-related and competition-related context in workload models provides a more realistic framework for interpreting pitcher responses over time.

Performance Outcomes as Signals of Workload Response

In applied baseball environments, performance outcomes are often the most visible signals of how a pitcher may be responding to workload. For the purposes of this dissertation, fastball velocity, spin rate, and vertical break were selected as the primary performance outcomes because they are routinely tracked, practically relevant, and directly tied to pitch characteristics.^{1,24,25} These measures offer a useful response domain because they are already familiar to coaches and analysts, and because changes in these outcomes may reflect meaningful shifts in pitcher output over time.

Fastball velocity serves as the clearest practical anchor among these outcomes.^{24,26} It is easy to interpret, widely used in decision making, and often the first metric noticed when discussing pitcher performance. However, velocity alone does not fully characterize pitch quality. Spin rate and vertical break provide additional information about how the pitch behaves and may reveal changes not captured by speed alone.^{1,25,27} A pitcher may show relatively stable velocity while demonstrating shifts in movement or spin-related characteristics, suggesting that different performance outcomes may vary in their sensitivity to workload-related change.

This distinction supports the use of multiple performance outcomes rather than a single endpoint. Workload response may not emerge uniformly across metrics, and relying on one measure alone may obscure meaningful heterogeneity. By modeling velocity, spin rate, and vertical break separately, the present dissertation sought to examine whether

rolling pitch count associations differed not only across pitchers, but across performance domains within pitchers. This framing aligns with the broader premise that workload response is multidimensional and should not be assumed to follow a single visible pattern.

Pitching Biomechanics as a Complementary Response Domain

Although ball-flight outcomes are highly relevant in applied baseball, they do not fully explain how a pitch was thrown. Pitching biomechanics provide a complementary domain for evaluating workload response because they reflect the movement strategies and loading patterns underlying pitch execution.^{10,18,28,29} Advances in motion-capture systems and applied biomechanics analysis have made it increasingly feasible to quantify variables related to joint loading, segment positioning, and kinetic-chain organization during pitching.^{10,18,30} These measures provide information that is not directly available from performance outcomes alone.

This distinction is especially important in the context of workload. A pitcher may maintain visible performance despite experiencing meaningful mechanical adaptation or compensation.^{10,11,16–18} For example, the athlete may preserve velocity while redistributing mechanical demands across the upper extremity or lower body, or while altering sequencing and positional characteristics within the motion. In these cases, performance metrics alone may suggest stability, whereas biomechanics may reveal a different workload-response pattern. As a result, biomechanics are not merely an optional extension of performance analysis, but a separate and potentially critical response domain.

The present dissertation therefore treated biomechanics as a key complementary outcome space. Variables related to elbow loading, shoulder loading, upper-extremity positioning, lead-leg mechanics, and whole-body coordination were examined to evaluate whether workload-related patterns were expressed mechanically within pitchers.^{10,18,28,31}

This approach reflects the view that pitcher response to accumulated throwing exposure may emerge in more than one form, and that a fuller understanding of workload requires attention to both what the pitch does and how the pitcher produces it.

Limitations of Group-Based Workload Research

Much of the workload literature in sport has relied on group-based approaches that estimate average relationships across athletes.^{5,6,32} These methods are valuable because they can identify broad trends, summarize shared patterns, and support population-level inference. In baseball, group-based models may be useful for describing general associations between pitch count and performance or biomechanics across a roster or season.

More recently, mixed-effects models have been increasingly adopted to account for repeated measures and within-athlete correlation across observations. However, these models are often interpreted primarily through their fixed effects, with comparatively less attention given to the variability captured by random effects or to how individual athletes deviate from the group-level relationship. As a result, the inferential target in much of the literature remains the average response rather than the fully characterized individual athlete.⁷⁻⁹

This distinction matters because meaningful heterogeneity is likely present in pitcher workload-response patterns. The average relationship across a group may not adequately represent the way any particular pitcher responds to accumulated exposure.^{5,9,14} A pitcher with one type of response curve may be averaged together with others who show weaker, stronger, or qualitatively different patterns. In such cases, the resulting group estimate may be statistically informative at the population level while still being of limited use for athlete-specific decision making, particularly when individual variability is not explicitly examined or reported.

These limitations are especially relevant in applied high-level baseball settings, where coaches and clinicians do not make decisions for a hypothetical average pitcher. They make decisions for the specific athlete currently preparing, recovering, or competing. Accordingly, the question is not whether group-based models remain informative, but whether they are sufficient when the goal is to understand workload response at the individual level. This question becomes central when athlete-specific monitoring and intervention are the practical targets of interest.

Rationale for Individualized and N-of-1 Approaches

An individualized analytic framework may be better suited for studying pitcher workload response when substantial heterogeneity is expected. N-of-1 methods shift the unit of inference from the group to the individual, using repeated measures within a single subject to estimate within-subject associations over time.^{6-9,32} This approach is especially relevant when the scientific question concerns how a particular athlete responds to exposure rather than how an average population behaves. In baseball, where pitch-level observations can accumulate longitudinally across a season, N-of-1 modeling is particularly well aligned with the underlying data structure.^{6,9}

While mixed-effects models have been increasingly used to account for repeated measures and allow athletes to deviate from group-level effects, these models are often interpreted primarily through their fixed effects.⁵⁻⁹ In many applied sport science settings, limited attention is given to the magnitude or meaning of individual differences captured by random effects, and individualized response patterns are not explicitly characterized. As a result, although mixed models permit heterogeneity, they do not always provide a clear understanding of how specific athletes respond to workload over time.⁵⁻⁹

The value of individualized modeling in this context comes from both its statistical and practical fit. Statistically, repeated observations within pitchers allow workload-

response patterns to be estimated without assuming that all pitchers share the same exposure-response relationship. Practically, individualized models align more closely with the way decisions are made in baseball environments. Coaches and practitioners need to understand whether a specific pitcher is behaving as expected under a given workload context, not simply whether the roster-average trend suggests a modest change. This makes individualized modeling especially attractive when responses are expected to vary across pitchers, seasons, and outcome types.

Importantly, individualized and group-based models should not be viewed as interchangeable versions of the same objective.^{5,7} They answer related but distinct questions. Group-based models estimate population-level associations while accounting for clustered observations. N-of-1 models estimate within-athlete associations and prioritize the individual as the inferential target. The distinction between these approaches becomes practically meaningful when the resulting conclusions differ. If individualized models capture structure that is obscured in group averages, then their added value extends beyond methodological preference and becomes directly relevant to athlete monitoring and management.

For these reasons, the present dissertation adopted an individualized N-of-1 generalized additive mixed modeling framework to evaluate rolling pitch count associations within collegiate baseball pitchers. This framework was chosen not simply because it allows flexible modeling, but because it better matches the conceptual and applied problem. Workload response in pitchers is unlikely to be fully represented by a single average effect, and the increasing availability of repeated pitch-level data creates an opportunity to estimate more athlete-specific associations across time.

Monitoring Applications in Sport

A major practical motivation for individualized modeling is that it may support monitoring applications rather than only retrospective description.^{1,4,6} In sport settings, practitioners often need to determine whether an athlete is behaving within an expected range or showing meaningful deviation from that expectation. This type of question is inherently individualized because the relevant baseline is not the team average, but the athlete's own typical response pattern under comparable conditions.

Individualized models create a framework for this type of monitoring by generating expected outcomes that can be compared with observed outcomes over time. When repeated deviations occur, those departures may signal a meaningful change in performance status or workload response. Monitoring tools such as cumulative sum control charts provide one way of formalizing this process by evaluating whether small but persistent deviations are accumulating across time.³³⁻³⁵ In the present dissertation, this applied perspective was used to extend individualized modeling beyond explanatory analysis and toward practical surveillance concepts that may be useful in real-world baseball settings.

Chapter Summaries

CHAPTER II: N-of-1 Modeling of Rolling Pitch Count Associations with Fastball Performance Metrics in Collegiate Baseball Pitchers - examines individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers using an N-of-1 modeling framework. Velocity, spin rate, and vertical break are evaluated as the primary outcomes, with emphasis placed on velocity as the clearest practical anchor. This chapter focuses on whether rolling pitch count is associated with measurable performance changes within pitchers and whether those associations vary across athletes and across ball-flight metric outcomes.

CHAPTER III: N-of-1 Modeling of Rolling Fastball Pitch Count Associations with Pitching Biomechanics in Collegiate Baseball Pitchers - extends the individualized modeling framework to pitching biomechanics. This chapter evaluates whether rolling pitch count is associated with biomechanical outcomes reflecting joint loading, segment positioning, and kinetic-chain organization, and whether workload-related patterns may emerge mechanically even when they are not fully evident in performance metrics alone

CHAPTER IV: Comparative Evaluation of Group-Based and Individualized N-of-1 Models for Pitch Count Associations in Baseball Pitchers - directly compares individualized N-of-1 models with traditional group-based models. This chapter evaluates whether individualized approaches provide information that differs from or improves upon population-average hierarchical approaches when quantifying pitch count associations. The comparison focuses on fastball velocity as the primary performance outcome and selected biomechanics outcomes as a secondary comparison domain.

CHAPTER V: Practical Applications of Individualized Pitch Count Models for Monitoring Collegiate Baseball Pitchers - apply individualized model outputs to practical monitoring concepts. Using individualized expectation-based frameworks, this chapter explores how pitcher-specific models may be used to detect meaningful deviation across time and to support applied workload surveillance. This chapter emphasizes practical interpretation and monitoring utility rather than only explanatory model performance.

Chapter VI: Summary of Research - is a brief overview of the completed research

References

1. Dowling B, McNally MP, Chaudhari AMW, Oñate JA. A Review of Workload-Monitoring Considerations for Baseball Pitchers. *J Athl Train*. 2020;55(9):911-917. doi:10.4085/1062-6050-0511-19
2. Bakshi NK, Inclan PM, Kirsch JM, Bedi A, Agresta C, Freehill MT. Current Workload Recommendations in Baseball Pitchers: A Systematic Review. *Am J Sports Med*. 2020;48(1):229-241. doi:10.1177/0363546519831010
3. Reinold MM, Dowling B, Fleisig GS, et al. An Interval Throwing Program for Baseball Pitchers Based upon Workload Data. *IJSPT*. 2024;19(3):326-336. doi:10.26603/001c.94146
4. Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. *Sports Med*. 2014;44(S2):139-147. doi:10.1007/s40279-014-0253-z
5. Araujo A, Julious S, Senn S. Understanding Variation in Sets of N-of-1 Trials. *PLoS One*. 2016;11(12):e0167167. doi:10.1371/journal.pone.0167167
6. Vieira R, McDonald S, Araújo-Soares V, Sniehotta FF, Henderson R. Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies. *Health Psychology Review*. 2017;11(3):222-234. doi:10.1080/17437199.2017.1343680
7. Yang L, Armijo-Olivo S, Gross DP. Single-Case Experimental Design in Rehabilitation: Basic Concepts, Advantages, and Challenges. *Am J Phys Med Rehabil*. 2023;102(5):454-460. doi:10.1097/PHM.0000000000002215

8. Schmid C, Yang J. Bayesian Models for N-of-1 Trials. *Harvard Data Science Review*. 2022;(Special Issue 3). doi:10.1162/99608f92.3f1772ce
9. Daza EJ. Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Track Health Data. *arXiv*. Preprint posted online January 22, 2019:arXiv:1901.03423. doi:10.48550/arXiv.1901.03423
10. Trasolini NA, Nicholson KF, Mylott J, Bullock GS, Hulburt TC, Waterman BR. Biomechanical Analysis of the Throwing Athlete and Its Impact on Return to Sport. *Arthrosc Sports Med Rehabil*. 2022;4(1):e83-e91. doi:10.1016/j.asmr.2021.09.027
11. Birfer R, Sonne MW, Holmes MW. Manifestations of muscle fatigue in baseball pitchers: a systematic review. *PeerJ*. 2019;7:e7390. doi:10.7717/peerj.7390
12. Shanley E, Bailey L, Sandago MP, Pinkerton A, Singleton SB, Thigpen CA. The use of a pitch count estimator to calculate exposure in collegiate baseball pitchers. *Phys Ther Sport*. 2015;16(4):344-348. doi:10.1016/j.ptsp.2015.02.002
13. Impellizzeri FM, Marcora SM, Coutts AJ. Internal and External Training Load: 15 Years On. *Int J Sports Physiol Perform*. 2019;14(2):270-273. doi:10.1123/ijsp.2018-0935
14. Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, Robertson SJ. Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *Int J Sports Physiol Perform*. 2017;12(2):230-234. doi:10.1123/ijsp.2015-0791
15. McLaren SJ, Macpherson TW, Coutts AJ, Hurst C, Spears IR, Weston M. The Relationships Between Internal and External Measures of Training Load and Intensity

- in Team Sports: A Meta-Analysis. *Sports Medicine*. 2018;48(3):641. doi:10.1007/s40279-017-0830-z
16. Erickson BJ, Sgori T, Chalmers PN, et al. The Impact of Fatigue on Baseball Pitching Mechanics in Adolescent Male Pitchers. *Arthroscopy*. 2016;32(5):762-771. doi:10.1016/j.arthro.2015.11.051
 17. Grantham WJ, Byram IR, Meadows MC, Ahmad CS. The Impact of Fatigue on the Kinematics of Collegiate Baseball Pitchers. *Orthop J Sports Med*. 2014;2(6):2325967114537032. doi:10.1177/2325967114537032
 18. Diffendaffer AZ, Bagwell MS, Fleisig GS, et al. The Clinician's Guide to Baseball Pitching Biomechanics. *Sports Health: A Multidisciplinary Approach*. 2023;15(2):274-281. doi:10.1177/19417381221078537
 19. Mosley E, Laborde S. A scoping review of heart rate variability in sport and exercise psychology. *International Review of Sport and Exercise Psychology*. 2024;17(2):773-847. doi:10.1080/1750984X.2022.2092884
 20. Gisselman AS, D'Amico M, Smoliga JM. Optimizing Intersession Reliability of Heart Rate Variability—The Effects of Artifact Correction and Breathing Type. *Journal of Strength and Conditioning Research*. 2020;34(11):3199-3207. doi:10.1519/JSC.0000000000002258
 21. Cornell DJ, Paxson JL, Caplinger RA, Seligman JR, Davis NA, Ebersole KT. Resting Heart Rate Variability Among Professional Baseball Starting Pitchers. *Journal of Strength and Conditioning Research*. 2017;31(3):575-581. doi:10.1519/JSC.0000000000001538

22. Rossiter A, Warrington GD, Comyns TM. Effects of Long-Haul Travel on Recovery and Performance in Elite Athletes: A Systematic Review. *The Journal of Strength & Conditioning Research*. 2022;36(11):3234. doi:10.1519/JSC.0000000000004021
23. Janse Van Rensburg DC, Jansen Van Rensburg A, Fowler P, et al. How to manage travel fatigue and jet lag in athletes? A systematic review of interventions. *Br J Sports Med*. 2020;54(16):960-968. doi:10.1136/bjsports-2019-101635
24. Chalmers PN, Erickson BJ, Ball B, Romeo AA, Verma NN. Fastball Pitch Velocity Helps Predict Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers. *Am J Sports Med*. 2016;44(8):2130-2135. doi:10.1177/0363546516634305
25. Nathan AM. The effect of spin on the flight of a baseball. *American Journal of Physics*. 2008;76(2):119-124. doi:10.1119/1.2805242
26. Matsuo T, Escamilla RF, Fleisig GS, Barrentine SW, Andrews JR. Comparison of Kinematic and Temporal Parameters between Different Pitch Velocity Groups. *Journal of Applied Biomechanics*. 2001;17(1):1-13. doi:10.1123/jab.17.1.1
27. Mayo BC, Miller A, Patetta MJ, et al. Preventing Tommy John Surgery: The Identification of Trends in Pitch Selection, Velocity, and Spin Rate Before Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers. *Orthop J Sports Med*. 2021;9(6):23259671211012364. doi:10.1177/23259671211012364

28. Nicholson KF, Mylott JA, Waterman BR, Bullock GS. Pitching Biomechanics Normative Values and Kinetic Differences by Competition Level. *J Surg Orthop Adv*. 2022;31(3):177-180.
29. Agresta C, Freehill MT, Nakamura B, Guadagnino S, Cain SM. Using Sensors for Player Development: Assessing Biomechanical Factors Related to Pitch Command and Velocity. *Sensors (Basel)*. 2022;22(21):8488. doi:10.3390/s22218488
30. Cardinale T, Salle TL, Aguinaldo A. CONCURRENT VALIDITY AND RELIABILITY OF IN-LAB MARKERLESS MOTION CAPTURE IN ESTIMATING JOINT KINEMATICS IN BASEBALL PITCHING. *ISBS Proceedings Archive*. 2024;42(1):144.
31. Bullock GS, Strahm J, Hulburt TC, Beck EC, Waterman BR, Nicholson KF. THE RELATIONSHIP OF RANGE OF MOTION, HIP SHOULDER SEPARATION, AND PITCHING KINEMATICS. *Int J Sports Phys Ther*. 2020;15(6):1119-1128. doi:10.26603/ijsp20201119
32. Daza EJ. Causal Analysis of Self-tracked Time Series Data Using a Counterfactual Framework for N-of-1 Trials*. *Methods Inf Med*. 2018;57(Suppl 1):e10-e21. doi:10.3414/ME16-02-0044
33. Noyez L. Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg*. 2009;9(3):494-499. doi:10.1510/icvts.2009.204768
34. Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. *British Journal of Surgery*. 2008;95(7):925-929. doi:10.1002/bjs.6056

35. Page ES. Continuous Inspection Schemes. *Biometrika*. 1954;41(1/2):100-115.

doi:10.2307/2333009

CHAPTER II: N-of-1 Modeling of Rolling Pitch Count Associations with Fastball Performance Metrics in Collegiate Baseball Pitchers

Garrett Fernandez^{1,2}, Garrett Bullock^{1,3}, Patrick Ward⁴, Joel Stitzel^{1,2}, Ryan McGinnis¹,
Tanner Filben⁵, Kristen Nicholson^{1,3}

¹Department of Biomedical Engineering, Wake Forest School of Medicine, Winston-Salem, NC, ²Virginia Tech-Wake Forest School of Biomedical Engineering and Sciences, Winston-Salem, NC, ³Department of Orthopedic Surgery and Rehabilitation, Wake Forest School of Medicine, Winston Salem, NC, ⁴Seattle Seahawks, Seattle, WA, USA, ⁵Elemance, LLC, Winston-Salem, NC

Abstract

Workload monitoring remains a persistent challenge in baseball, particularly in collegiate pitchers exposed to repeated throwing demands across games, practices, and bullpens. Although pitch count is one of the most widely used workload indicators in applied settings, it does not fully capture the broader context in which performance responses occur. Therefore, the purpose of this chapter was to examine individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers ($n = 35$) using an N-of-1 modeling framework. Velocity, spin rate, and vertical break were evaluated as primary outcomes using a longitudinal observational design with pitcher-season as the analytic unit across the 2023, 2024, and 2025 seasons, totaling 12,497 pitches across 837 appearances. Final models were developed separately for each pitcher-season and outcome using LASSO-based predictor screening followed by generalized additive mixed modeling, with rolling pitch count retained as the primary exposure and contextual covariates related to recovery, travel, and workload included when available. Velocity demonstrated the strongest overall model performance (adjusted $R^2 = 0.54$; RMSE = 0.87 mph), followed by spin rate (adjusted $R^2 = 0.44$; RMSE = 70.1 rpm) and vertical break (adjusted $R^2 = 0.35$; RMSE = 2.28 in). AR(1) residual structures were retained in 77.1% of velocity models, 85.7% of spin rate models, and 68.6% of vertical break models, while interaction terms were retained in 80.0%, 65.7%, and 77.1% of final models, respectively. Individualized curves demonstrated substantial heterogeneity in both magnitude and functional form across pitcher-seasons. Overall, rolling pitch count associations could be characterized at the pitcher-season level across all three fastball outcomes, but the strength and consistency of those associations differed by metric, with velocity providing the clearest and most practically interpretable individualized signal.

Introduction

Workload monitoring remains a persistent challenge in baseball, particularly in collegiate pitchers who are exposed to repeated throwing demands across games, practices, and bullpens.¹⁻³ Among available workload indicators, pitch count is one of the most widely used in applied settings because it provides a simple and easily accessible estimate of throwing volume. However, pitch count alone does not provide a clear or consistent indication of performance response. Pitch count may reflect accumulated throwing exposure, but it does not fully capture the broader internal and external demands that influence how a pitcher responds over time.⁴⁻⁶ A pitcher's response to a given workload may be influenced by recovery status, recent workload history, travel demands, and other contextual factors.⁷⁻⁹

This complexity is closely related to the broader concept of fatigue in sport. Fatigue is frequently invoked when describing performance changes in athletes, yet it remains difficult to define and measure directly. In baseball, fatigue-related responses are often inferred from observable changes in performance rather than identified through any single definitive marker.^{7,9-11} For pitchers, such changes may be reflected in fastball performance metrics including velocity, spin rate, and vertical break.^{7,11,12} These outcomes are practically important because they are routinely tracked, directly relevant to on-field performance, and can be sensitive to changes in pitch characteristics over time. Among them, fastball velocity is often the most visible and interpretable indicator of performance, although spin rate and vertical break provide additional information regarding pitch quality that may not be captured by velocity alone.

Previous workload research in baseball has often relied on group-based approaches that estimate average responses across athletes.^{3,7,9,11} Although such

methods are useful for identifying population-level patterns, they may obscure substantial heterogeneity in workload-related performance responses.^{13–15} Pitchers exposed to similar rolling pitch counts may demonstrate significantly different performance trajectories due to differences in physical characteristics, recovery capacity, mechanical strategies, and competitive context.^{3,7,11} This limitation is particularly relevant in applied settings, where workload decisions are made for individual athletes rather than for an average member of a group.

An individualized analytic framework may therefore be better suited for evaluating these relationships. N-of-1 methods shift the unit of inference from the group to the individual, allowing repeated measures from a single athlete to be used to estimate within-pitcher associations over time.^{14,16–18} This approach is well aligned with baseball performance data, where dense longitudinal observations are often available and where workload-related responses may vary not only across pitchers, but also across outcomes within the same pitcher.^{16–19} In this context, individualized models provide an opportunity to evaluate whether rolling pitch count is associated with consistent changes in fastball performance and whether those associations differ across pitchers and across performance metrics.

Importantly, rolling pitch count should not be evaluated in pure isolation.^{2,4–6} Because pitcher performance is influenced by a broad workload environment, models based solely on pitch count may oversimplify the processes underlying performance change. Measured indicators reflecting internal load, external workload, recovery, and travel-related factors may all contribute to observed performance patterns and may influence how rolling pitch count associations manifest over time.^{5–7,20,21} Incorporating these contextual factors provides a more comprehensive framework for evaluating workload-related performance

patterns, while acknowledging that no single set of observed variables can fully account for all sources of variation.

Therefore, the purpose of this chapter was to examine individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers using an N-of-1 modeling framework. Fastball velocity, spin rate, and vertical break were evaluated as the primary outcomes, with particular emphasis on velocity because of its prominence and ease of interpretation in baseball contexts. By modeling these outcomes within pitchers and utilizing a broader measured workload context, this chapter sought to characterize workload-related performance patterns associated with rolling pitch count, determine the extent to which those patterns varied across pitchers, and assess whether responsiveness differed across fastball performance metrics.

Methods

Study Design

This study used a longitudinal observational design to examine individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers using an N-of-1 modeling framework.^{13,14,19,22–24} The analytic focus was on within-pitcher associations over time rather than population-average effects. Data was drawn from three competitive seasons (2023, 2024, and 2025), with the 2023 season analyzed retrospectively and the 2024 and 2025 seasons collected prospectively. Each pitcher-season was treated as a distinct longitudinal unit to preserve season-specific workload context and to allow workload-performance associations to vary across both pitchers and seasons. This chapter focused on fastball performance outcomes, including velocity, spin rate, and vertical break.

Study Participants and Eligibility Criteria

This study was approved by the Wake Forest University School of Medicine Institutional Review Board. The study sample consisted of collegiate baseball pitchers with available pitch-level data during the study period. Analyses were conducted at the pitcher-season level, such that each eligible pitcher within each season contributed an individualized observational series. This approach was selected to preserve the within-season workload context and to allow workload-performance associations to vary across both pitchers and seasons.

Pitcher-seasons were eligible for inclusion if they contained at least 50 recorded fastballs with corresponding performance metrics and rolling pitch count information sufficient for individualized longitudinal modeling. Because the analytic goal was to estimate within-pitcher associations, pitcher-seasons were excluded when the number of usable pitch-level observations was insufficient for model development. Initial predictor screening was performed using least absolute shrinkage and selection operator (LASSO) regression, and subsequent modeling procedures were applied only to eligible pitcher-seasons meeting these data sufficiency requirements.

Sample Size and Data Sufficiency

Data sufficiency was evaluated using both a simulation-based planning analysis and empirical checks performed within the final individualized modeling framework. During study planning, a simulation-based precision analysis was conducted using repeated-measures data structures informed by the Spring 2023 baseball season.^{25,26} The simulation framework assumed 14 collegiate pitchers, incorporated pitcher-level random intercept and random slope variation and evaluated a series of increasingly complex repeated-measures model specifications across 1,000 simulations per candidate model. Precision was summarized using the estimated coefficient and corresponding 95% confidence interval width for the pitch-count effect. This analysis was used to inform

feasible model complexity during study development rather than to define a single fixed sample-size threshold for the final individualized models.

For the final analyses, data sufficiency was evaluated at the pitcher-season level, consistent with the N-of-1 design. For each pitcher-season and outcome, initial predictor screening was performed only when at least 50 usable observations were available. Observed data sufficiency for individualized modeling was further examined through stabilization analyses, trial-reduction sensitivity analyses, and model-specification sensitivity analyses evaluating the effects of removing selected model components. These procedures were used to assess the extent to which model behavior changed as the amount of available data or model complexity was reduced. This combined approach was used to evaluate data adequacy for individualized longitudinal modeling while accounting for variation in the amount of repeated data available across pitcher-seasons.

Data Sources and Variable Collection

Data was assembled from season-level pitching datasets spanning the 2023, 2024, and 2025 collegiate baseball seasons. The 2023 season was analyzed retrospectively, whereas the 2024 and 2025 seasons were collected prospectively. Pitch-level fastball performance data were obtained from Trackman V3 Game Tracking Unit (Trackman, Scottsdale, AZ) and subsequently compiled by the Wake Forest Baseball Analytics team into the analytic database used for this study.²⁷ This data included ball-flight metrics such as velocity, spin rate, vertical break, pitch type, and pitch count. Additional workload and contextual variables were obtained from the analytics team's records, while travel-related variables were derived from team travel records. Internal load data was obtained using WHOOP, a wrist-worn monitoring device for pitchers who elected to wear them; accordingly, HRV data were available only for a subset of pitchers in the prospectively collected 2024 and 2025 seasons.^{28,29} Pitcher identity, season, and

appearance were retained as indexing variables for longitudinal analysis. Prior to model development, season files were synchronized to ensure consistent variable naming, coding structure, and data type formatting across years while preserving the original pitch-level structure of the observations.

Variable Definitions

Fastball Performance Metrics

Fastball performance metrics refer to measurable characteristics of a baseball pitch following ball release. For the purposes of this study, three commonly used fastball performance metrics were evaluated: velocity, spin rate, and vertical break.^{30–32} Velocity was defined as the speed of the baseball immediately after release from the pitcher's hand and was reported in miles per hour. Spin rate was defined as the rate of baseball rotation during the pitch and was reported in revolutions per minute. Vertical break was defined as the vertical movement profile of the baseball during flight and was reported in inches. Each metric was modeled separately at the pitcher-season level.

Rolling Pitch Count Exposure

Rolling pitch count served as the primary workload exposure in this study. Pitch count was defined as the number of pitches thrown within a given pitching context and was used to represent accumulated throwing workload over time.^{1–3} Pitching appearances were defined as discrete game or practice throwing sessions in which the pitcher recorded eligible fastball observations. Time between appearances was defined as the number of days separating consecutive pitching appearances. Additional workload-history variables, including lagged pitch count measures and prior appearance pitch count, were derived to reflect recent exposure preceding a given observation.

Contextual Covariates

Contextual covariates were included to account for the broader workload environment in which rolling pitch count associations occurred. Recovery-related variables included rest, defined as the number of days between appearances, and prior appearance maximum pitch count, defined as the highest pitch count recorded in the preceding appearance. Internal load was represented by heart rate variability (HRV), obtained from WHOOP wrist-worn devices used among pitchers who elected to participate in wearable monitoring. HRV reflects beat-to-beat variation in cardiac intervals and was included as an available indicator of physiological status within the broader workload context.^{28,29} Because HRV data were only collected during the prospectively observed 2024 and 2025 seasons, this variable was available only for a subset of pitcher-seasons. Game context was represented by an indicator distinguishing game from non-game appearances. Travel-related variables included travel status, distance traveled in kilometers, and method of transportation. Transportation method was categorized according to whether travel occurred by bus or plane, with non-travel observations coded separately during preprocessing.

Missing Data

Missing data was addressed during preprocessing using multiple imputation, consistent with the broader analytic framework used in this project.³³ Prior to imputation, variables not retained for the analytic dataset were removed, and remaining variables were reviewed for missingness. Analytic variables were then formatted as needed for imputation, with numeric conversion applied to retained continuous variables where appropriate. Multiple imputation was subsequently performed using predictive mean matching (PMM) implemented through the mice package in R.³³ Five imputations were generated over 50 iterations, after which a completed analytic dataset was derived for downstream modeling. Selected indexing and identifying variables were then restored

from the original source file where needed. This approach was used to preserve the observed structure of the data while reducing case loss associated with incomplete covariate information. Because HRV was not collected during the retrospective 2023 season and was only available for a subset of pitchers in the prospectively collected 2024 and 2025 seasons, its absence in the 2023 season was treated as structural unavailability rather than ordinary random missingness.

Data Processing and Feature Construction

Following data assembly, season-level datasets were synchronized across 2023, 2024, and 2025 to ensure consistent variable naming, coding structure, and analytic formatting prior to model development. Pitch-level observations were retained as the primary unit of analysis, with pitcher, season, and appearance preserved as indexing variables for longitudinal modeling. Variables required for analysis were reformatted as needed according to their analytic role, with continuous variables represented numerically and categorical variables retained in factor form where appropriate.

Feature construction focused on preparing workload-history and contextual variables for individualized modeling. Derived variables included rolling pitch count, lagged pitch count measures from prior appearances, prior appearance maximum pitch count, and time between appearances. Travel-related variables were also organized for analysis, including travel status, distance traveled, and transportation method. To reduce redundancy during predictor screening, travel-related variables and lagged pitch count variables were each treated as grouped candidate predictor sets, from which only the strongest variable within a group was retained for subsequent model development.

N-of-1 Model Development

Individualized model development was conducted separately for each pitcher-season and for each fastball performance outcome. Model development proceeded in two

stages: predictor screening followed by pitcher-season-specific model fitting. The overall level of allowable model complexity was informed by the earlier simulation-based precision analysis described in the sample size and data sufficiency section. Based on that planning framework, final individualized models were restricted to a maximum of five predictor variables in addition to the primary pitch-count term, with allowance for up to one interaction term where supported by the data and model fit. This constraint was used to balance model flexibility with data sufficiency at the pitcher-season level.

In the predictor-screening stage, LASSO regression was used to identify a reduced set of candidate predictors for each pitcher-season and outcome.^{34,35} Screening was performed using cross-validated glmnet models with a Gaussian family and standardized predictors. Rolling pitch count was forced into all screening models by assigning it a zero penalty, ensuring retention of the primary workload exposure regardless of selection of additional variables. Candidate predictors included recovery-related, workload-history, internal load, game-context, and travel-related variables. When multiple predictors were available from conceptually overlapping groups, including lagged pitch count variables and travel-related variables, only the strongest variable from each group was retained for subsequent model development. After screening, no more than five predictors were carried forward into final model fitting, not including the forced pitch-count term.

In the model-fitting stage, pitcher-season-specific generalized additive mixed models (GAMMs) were fit separately for velocity, spin rate, and vertical break using mgcv::gamm, with nlme correlation structures incorporated when appropriate.³⁶ Rolling pitch count was modeled as a smooth term using $s(\text{PitchCount}, k = 4)$ to allow potentially nonlinear associations with each performance outcome. Predictors retained from the screening stage were incorporated as linear terms unless evaluated as part of a candidate interaction. Interaction terms were assessed one at a time, with no more than one

interaction retained in a final model. For numeric interaction variables, tensor product smooths of the form $\text{te}(\text{PitchCount}, Z, k = c(4,4))$ were used.³⁶ For transportation, interactions were represented using $s(\text{PitchCount}, \text{by} = \text{Transportation}, k = 4) + \text{Transportation}$. When an interaction term was included, the corresponding variable was not simultaneously retained as an additional main linear term.

Within each pitcher-season, multiple candidate models were evaluated, including a base model and interaction-augmented alternatives where applicable. Candidate models were further assessed with and without an AR(1) residual correlation structure.³⁷ The AR(1) formulation was specified within appearance using $\text{corAR1}(\text{form} = \sim \text{PitchCount} \mid \text{Appearance})$, and a random intercept for appearance was included to account for clustering of pitches within appearances. Final model selection was based primarily on Akaike information criterion, with consideration of overall model fit and interpretability.

Model Evaluation, Diagnostics, and Sensitivity Analyses

Several procedures were used to evaluate individualized model performance, residual behavior, and robustness. Candidate models were compared primarily using Akaike information criterion (AIC).³⁸ For each pitcher-season, base and interaction-augmented candidates were evaluated with and without an AR(1) residual correlation structure, and the final model was selected as the candidate with the lowest AIC. Final reported model performance was summarized using root mean square error (RMSE), coefficient of determination (R^2), and adjusted within-subject R^2 . Mean squared error (MSE) and Bayesian information criterion (BIC) were retained as secondary checks during model evaluation, but were not emphasized in the main chapter summaries because RMSE and adjusted R^2 provided the most interpretable indicators of model error and fit for the present results.^{39,40}

Model adequacy was further assessed using residual-based diagnostic procedures, including residuals-versus-fitted plots and quantile-quantile (Q-Q) plots, to examine overall fit and key distributional assumptions. Residual autocorrelation was initially characterized using baseline linear models of velocity on pitch count and appearance, followed by estimation of within-appearance AR(1) values. These summaries were used to characterize the extent of serial dependence across pitcher-seasons and to support model-level comparison of candidate formulations with and without AR(1) correlation structures. Diagnostic outputs also included residual-based visualizations generated for individualized models.

Because model adequacy in N-of-1 settings may be sensitive to both specification choices and data quantity, multiple sensitivity analyses were performed. First, model specification sensitivity analyses compared four simplified configurations reflecting the presence or absence of HRV and AR(1) structures. These analyses were used to quantify changes in RMSE, adjusted R^2 , and residual autocorrelation associated with inclusion or exclusion of HRV and AR(1) structures. Second, stabilization analyses assessed whether models fit using progressively larger early-series subsets converged toward the corresponding full-series velocity curves as additional observations were added. Finally, trial-reduction analyses evaluated how removing increasing proportions of observations, including prefix truncation and other reduced-data scenarios, affected model fit and deviation from the corresponding full-data curves.

Statistical Analysis

Descriptive statistics were calculated across the combined dataset to summarize the observed distributions of workload and fastball performance variables. Individualized analyses were then conducted separately for each pitcher-season and for each fastball performance outcome. Statistical emphasis was placed on estimation of within-pitcher

associations between rolling pitch count and fastball performance metrics, as well as the extent of heterogeneity in these associations across pitcher-seasons, rather than on a single pooled population-average effect.

All data preprocessing, model development, model evaluation, sensitivity analyses, and visualization were performed in R version 4.5.1 (R Core Team, 2025). Key packages included dplyr 1.1.4, readxl 1.4.5, readr 2.1.5, ggplot2 3.5.2, glmnet 4.1.9, mgcv 1.9.3, nlme 3.1.168, purrr 1.1.0, stringr 1.5.1, tidyr 1.3.1, tibble 3.3.0, and car 3.1.3. Sample code illustrating the model development pipeline and overall analytic workflow is provided in Appendix D.

Results

Descriptive Statistics

Descriptive statistics for the fastball dataset are presented in Table I. Across the three study seasons, the analytic sample included 35 pitcher-seasons, 837 appearances, and 12,497 fastballs. Pitch-level volume increased across seasons, whereas mean pitch count per appearance declined. Fastball velocity remained relatively stable across years, spin rate was similar in 2023 and 2024 before decreasing in 2025, and mean vertical break became progressively more negative across seasons. Variability in the fastball performance metrics also differed across years, with the broadest spread generally observed in 2024.

The distribution of rolling pitch count is shown in Figure 1. Rolling pitch count demonstrated a right-skewed distribution, with the highest concentration of observations occurring at lower pitch counts and progressively fewer observations at higher pitch counts. This pattern is consistent with the structure of baseball throwing exposures, where a substantial proportion of pitches arise from shorter practice or bullpen sessions, while fewer observations extend into higher-count game or extended-appearance workloads.

For descriptive visualization purposes, pitchers were also classified by role, with starters defined as pitchers averaging more than 50 pitches per game and all others classified as relievers. Using this rule, three starters were identified in each season. As shown in Figure 2, starters demonstrated higher fastball pitch counts per appearance than relievers, particularly in game settings. This role-based classification was used descriptively and was not part of the primary individualized modeling framework.

Table I: Descriptive Statistics for Fastball Dataset by Season

	2023 Season (Retrospective)	2024 Season	2025 Season
# of Pitchers	10	10	15
# of Appearances	238	273	326
# of Pitches (Fastballs)	3731	4334	4432
Mean Pitch Count (SD)	24.6	22.9	19.7
Mean Velocity (SD) (mph)	92.1 (2.3)	92.1 (3.5)	91.9 (2.1)
Mean Spin Rate (SD) (rpm)	2319.8 (162.7)	2329.3 (218.3)	2215.5 (137.9)
Mean Vertical Break (SD) (in)	-16.8 (5.4)	-18.6 (9.3)	-19.6 (8.0)

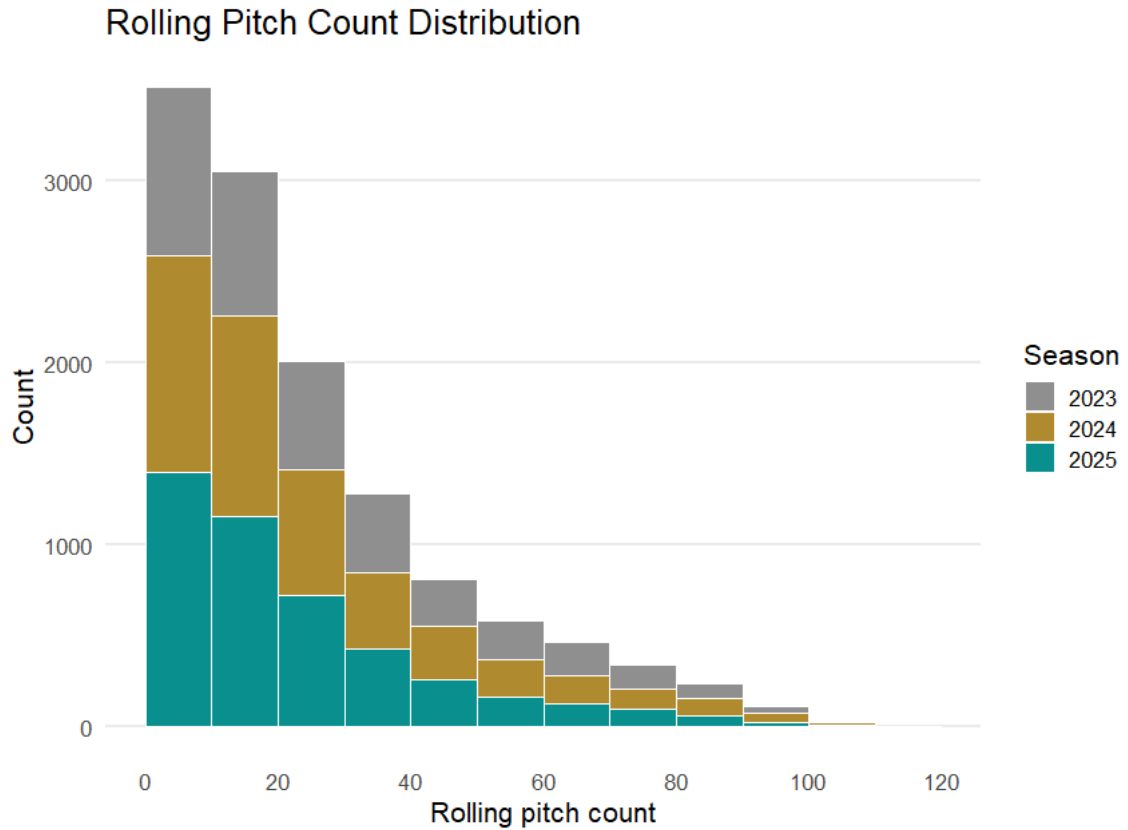


Figure 1: Distribution of rolling pitch count across the 2023–2025 seasons. Histogram of rolling pitch count values pooled across the analytic fastball dataset and stratified by season.

Exposure by Role (Fastballs)

Pitches per appearance stratified by game vs practice

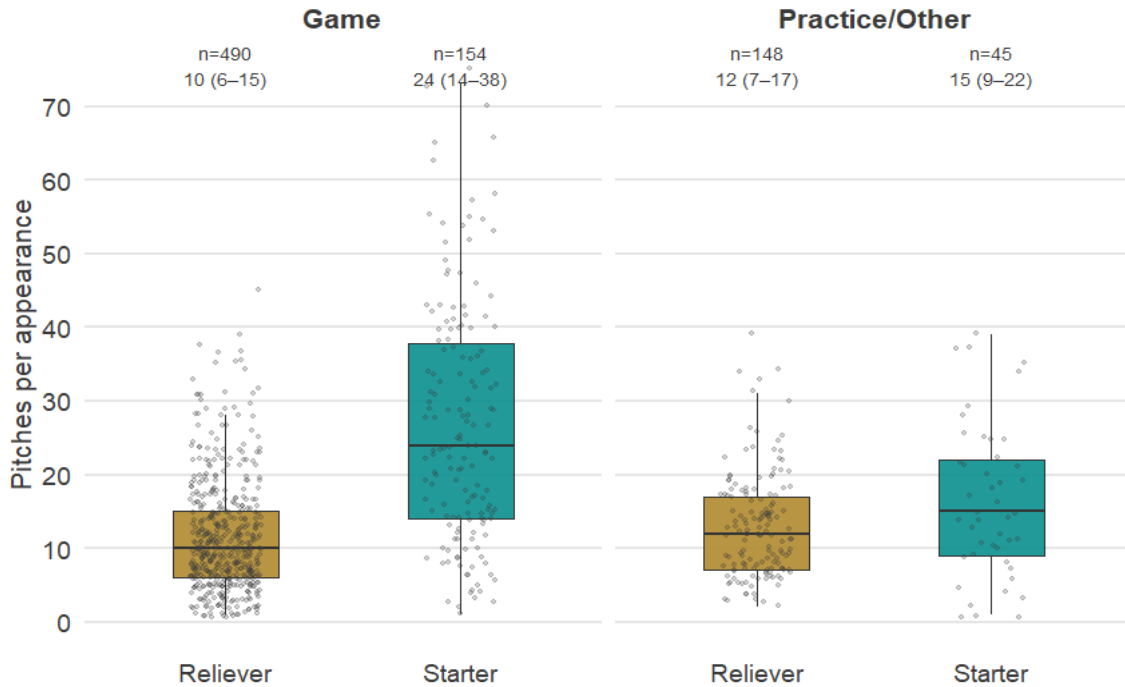


Figure 2: Fastball pitches per appearance by pitcher role and setting. Distribution of fastball pitches per appearance stratified by role classification and by game versus practice/other appearances.

Overall N-of-1 Model Performance Across Fastball Metrics

Overall N-of-1 model performance across the three fastball outcomes is summarized in Table II. Across outcomes, velocity demonstrated the strongest overall model fit, spin rate showed intermediate performance, and vertical break demonstrated the weakest fit. AR(1) residual correlation structures and interaction terms were retained in a majority of final models across all three outcomes. Under the model-selection procedure, AR(1) structures were retained when the AR(1) version of a candidate model had a lower AIC than the corresponding model without AR(1).

Table II: Overall N-of-1 Model Performance Across Fastball Metrics

	Velocity (mph)	Spin Rate (rpm)	Vertical Break (in)
Avg # of Pitches	347.5	347.4	347.3
Mean Adjusted R ²	0.54	0.44	0.35
Mean R ²	0.55	0.45	0.36
Mean RMSE	0.87	70.1	2.28
AR(1) Retained	77.1%	85.7%	68.6%
Interaction Retained	80%	65.7%	77.1%

The distribution of adjusted R² values across outcomes is shown in Figure 3. Velocity models were generally centered at higher adjusted R² values than spin rate and vertical break models, whereas vertical break demonstrated the lowest overall distribution and the greatest between-pitcher variability. Because velocity demonstrated the strongest

overall model performance, subsequent results are presented first for fastball velocity, followed by spin rate and vertical break.

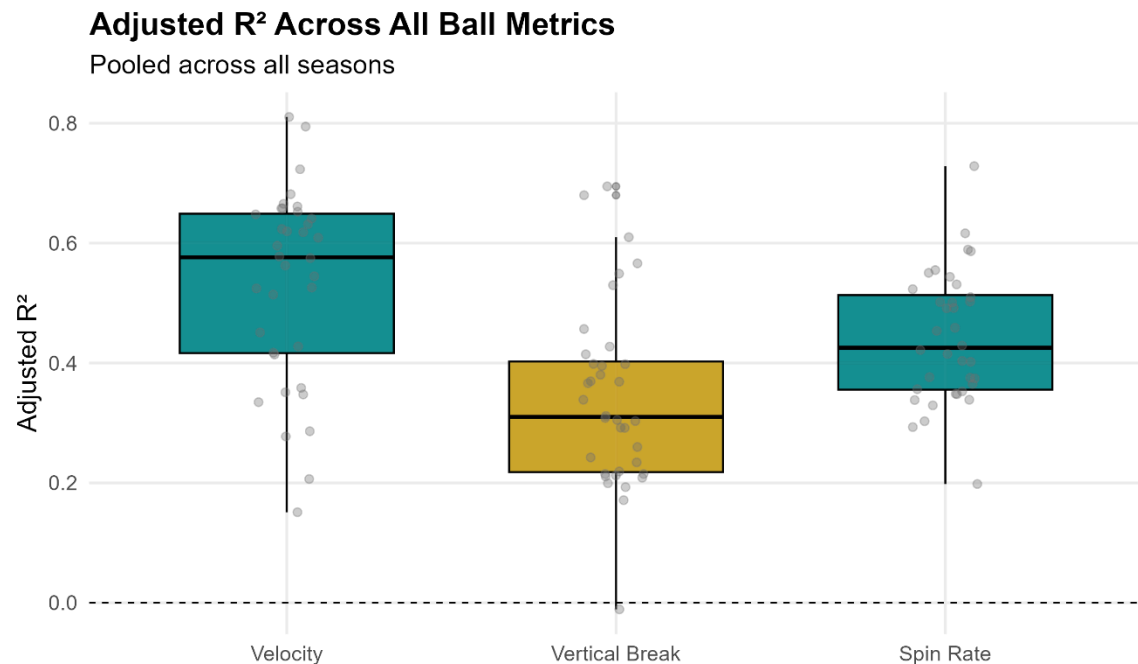


Figure 3: Distribution of adjusted R² across individualized fastball performance models. Boxplots with overlaid pitcher-season values showing the distribution of adjusted R² for N-of-1 models fit to velocity, vertical break, and spin rate.

Fastball Velocity Results

Fastball velocity demonstrated the strongest overall individualized model performance of the three fastball outcomes. As shown in Figure 4, adjusted R² values for the velocity models remained consistently positive across all three seasons, with median values exceeding 0.50 in 2023 and 2025 and somewhat lower values in 2024. Although variability in model fit was present within each season, velocity models demonstrated moderate-to-strong explanatory performance across pitcher-seasons.

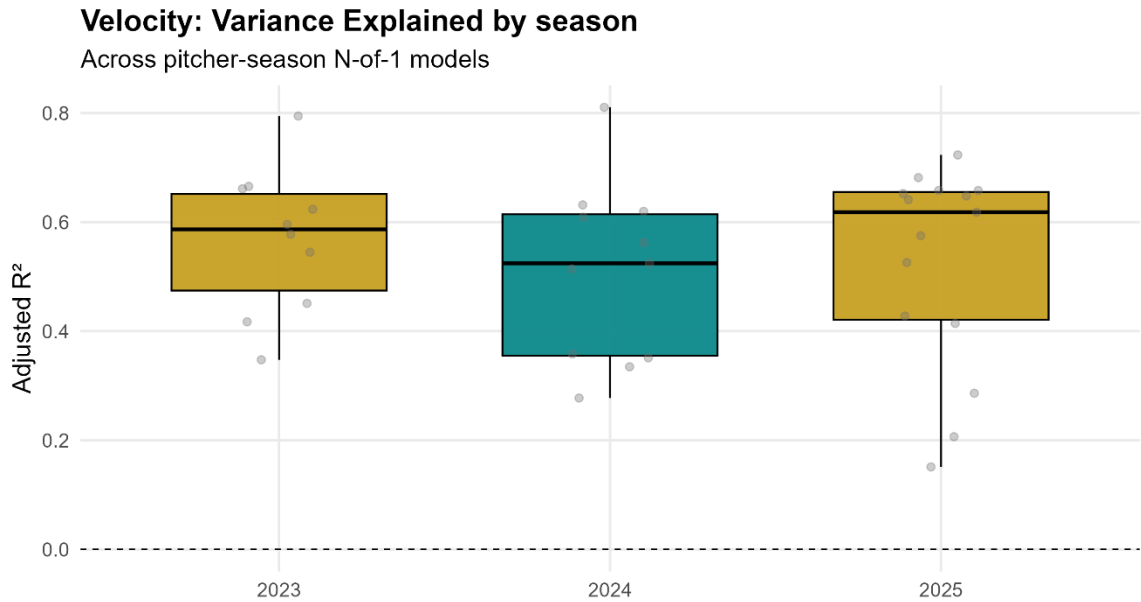


Figure 4: Season-specific distribution of adjusted R^2 for individualized fastball velocity models. Boxplots with overlaid pitcher-season values showing the distribution of adjusted R^2 for N-of-1 velocity models across the 2023, 2024, and 2025 seasons.

The individualized fastball velocity curves are shown in Figure 5. Across pitcher-seasons, the modeled association between rolling pitch count and velocity varied substantially in both magnitude and functional form. Many pitcher-seasons demonstrated declining modeled velocity as rolling pitch count increased, although the steepness of decline differed across individuals. Other pitcher-seasons showed comparatively stable modeled velocity across the observed pitch-count range, whereas several demonstrated more nonlinear patterns, including localized curvature or non-monotonic changes. These curves show substantial heterogeneity in the modeled association between rolling pitch count and fastball velocity across pitchers and seasons. Additional individualized velocity curves are provided in Appendix A. A sample of diagnostic plots is provided in Appendix E. Because more than 1,000 diagnostic plots were generated for this project, not all are included.

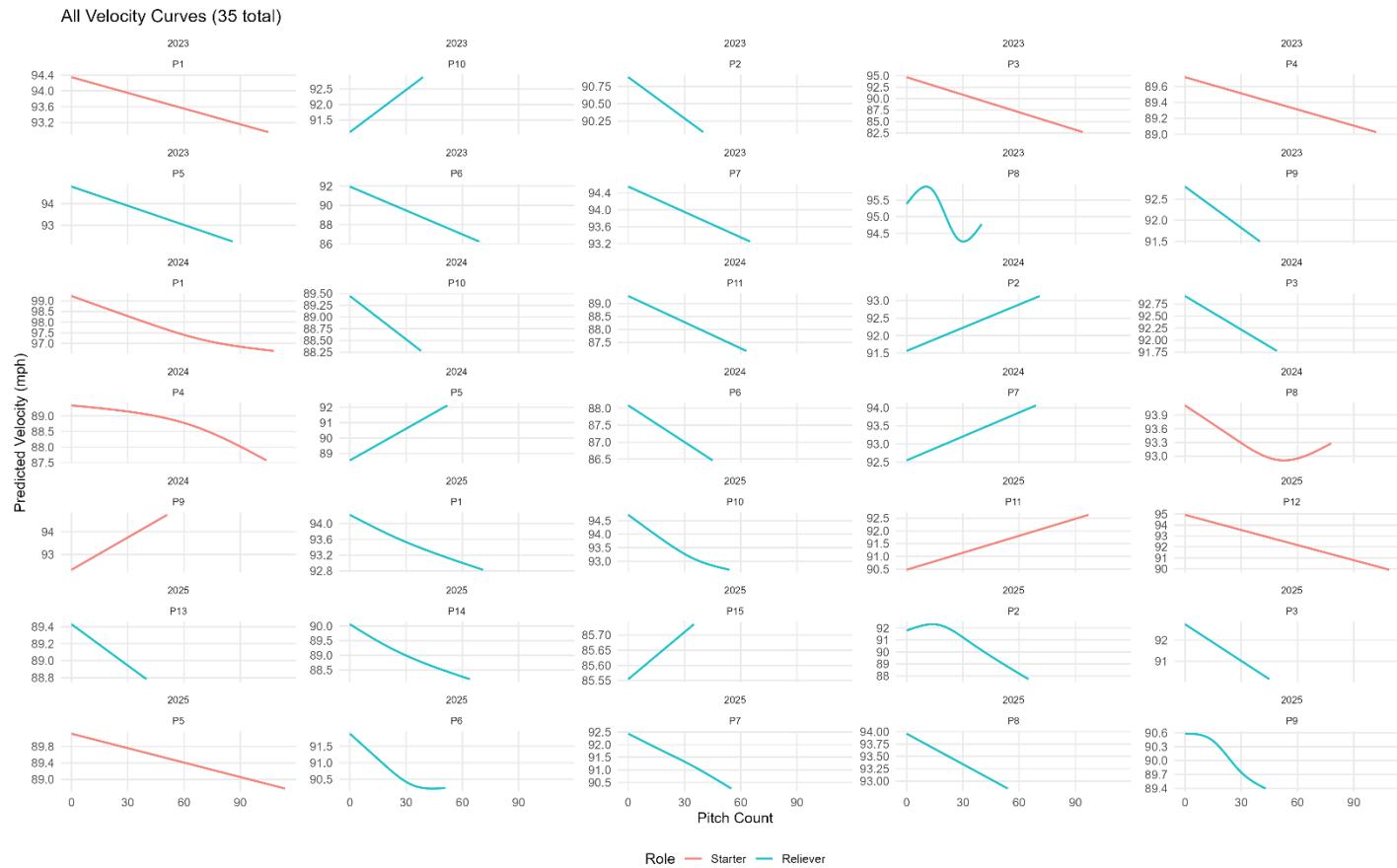


Figure 5: Individualized fastball velocity curves across rolling pitch count. Predicted N-of-1 velocity curves for all pitcher-seasons, illustrating the range of modeled associations between rolling pitch count and fastball velocity.

Fastball Spin Rate Results

Fastball spin rate demonstrated intermediate overall model performance relative to velocity and vertical break. As shown in Figure 6, adjusted R^2 values for the spin rate models were positive across all three seasons, with moderate variability across pitcher-seasons. Median adjusted R^2 appeared lowest in 2023 and somewhat higher in 2024 and 2025. Across seasons, the distribution of adjusted R^2 for spin rate remained below velocity but above vertical break. Although between-pitcher variability was present in each season, most spin rate models clustered within a moderate range of explanatory performance. The 2025 season showed the broadest spread of adjusted R^2 values, whereas 2023 and 2024 were somewhat more compactly distributed. Overall, individualized spin rate model fit was more variable than velocity but remained consistently above the distribution observed for vertical break. Additional individualized spin rate curves are provided in Appendix A.

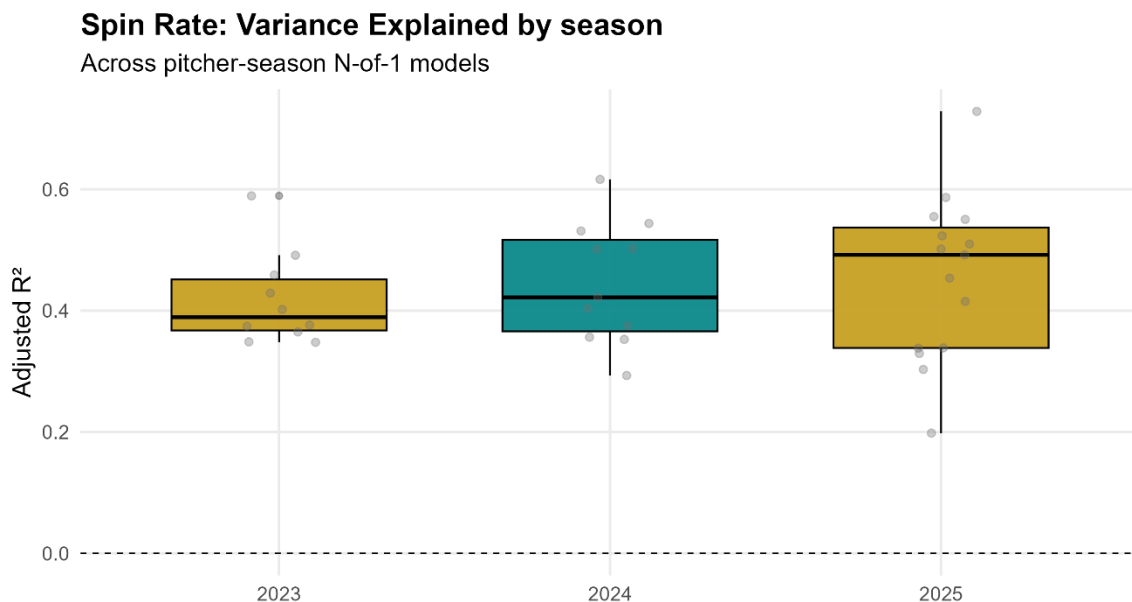


Figure 6: Season-specific distribution of adjusted R^2 for individualized fastball spin rate models. Boxplots with overlaid pitcher-season values showing the distribution of adjusted R^2 for N-of-1 spin rate models across the 2023, 2024, and 2025 seasons.

Fastball Vertical Break Results

Fastball vertical break demonstrated the weakest overall model performance of the three fastball outcomes. As shown in Figure 7, adjusted R^2 values for the vertical break models varied across seasons, with median values lowest in 2023, higher in 2024, and highest in 2025. Although most pitcher-seasons showed positive adjusted R^2 values, a small number of models were near zero or below zero, particularly in 2025. Between-pitcher variability in model fit was present across all three years, with several pitcher-seasons showing comparatively low adjusted R^2 values and others demonstrating more moderate model performance. The 2025 season showed the broadest overall range of adjusted R^2 values, whereas the 2023 distribution was centered at the lowest values. Overall, individualized vertical break model fit remained below that observed for velocity and spin rate across seasons. Additional individualized vertical break curves are provided in Appendix A.

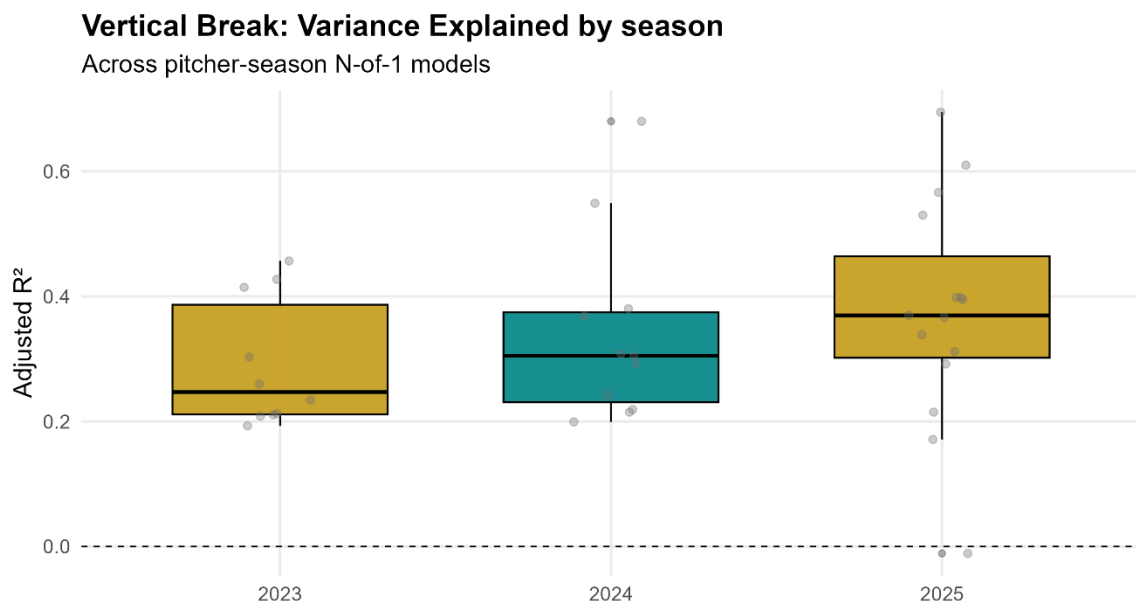


Figure 7: Season-specific distribution of adjusted R^2 for individualized fastball vertical break models. Boxplots with overlaid pitcher-season values showing the distribution of adjusted R^2 for N-of-1 vertical break models across the 2023, 2024, and 2025 seasons.

Residual Autocorrelation, Stability, and Sensitivity Analyses

Model diagnostics and robustness analyses were performed to evaluate residual autocorrelation, model stability, and sensitivity to alternative specifications and reduced-data conditions. Residual autocorrelation summaries are presented in Table III, and specific pitcher-season AR(1) values for the velocity models are shown in Figure 8. Although average AR(1) values were low overall, residual autocorrelation varied across pitcher-seasons and outcomes, with both positive and negative values observed. Positive mean AR(1) values were more common than negative values for all three fastball outcomes.

Table III: Residual autocorrelation summary for individualized fastball models

Outcomes	Average Mean AR(1)	Median AR(1)	Range of Mean AR(1)	% AR(1) Positive/ Negative
Velocity	0.052	0.038	-0.304 to 0.359	65.7% / 34.3%
Spin Rate	0.037	0.041	-0.237 to 0.252	74.3% / 25.7%
Vertical Break	0.010	0.006	-0.235 to 0.261	57.1% / 42.9%

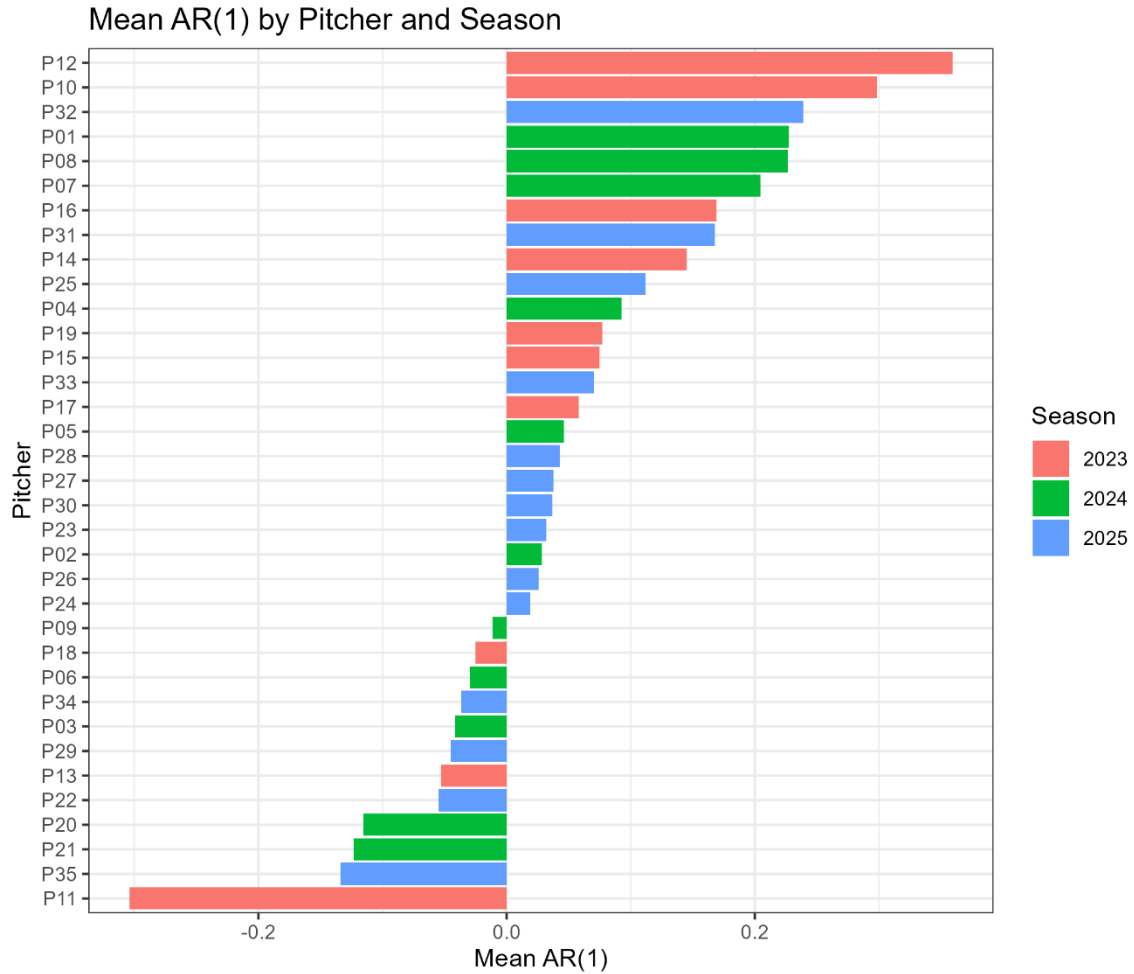


Figure 8: Mean AR(1) values by pitcher-season for individualized fastball velocity models. Mean within-appearance AR(1) values for pitcher-seasons included in the velocity analyses, stratified by season.

Stability and sensitivity summaries are presented in Table IV. Stability analyses assessed whether models based on progressively larger early-series subsets converged toward the corresponding full-series models. Sensitivity analyses evaluated how model behavior changed under alternative model specifications or reduced-data conditions. Across outcomes, stabilization analyses showed lower deviation from the corresponding full-series models as additional early observations were included. Alternative model

specifications produced relatively small changes in RMSE, whereas reduced-data models showed greater deviation from the corresponding full models under trial-reduction conditions. Visual inspection of model diagnostic plots largely supported acceptable residual behavior across pitcher-seasons, including residuals-versus-fitted patterns, Q-Q plots, residuals across pitch order, and residual autocorrelation functions.

Table IV: Summary of RMSE-Based Stability and Sensitivity Across Fastball Outcomes

Analyses	Velocity	Spin Rate	Vertical Break
Curve stabilization			
▲RMSE (prefix vs full)	0.64 [0.33, 1.02]	36.86 [18.64, 62.92]	1.34 [0.74, 2.38]
Alternate			
Specification vs Baseline ▲RMSE	0.03 [0.01, 0.06]	0.97 [0.29, 3.35]	0.04 [0.01, 0.08]
Trial Sensitivity			
(30% removal) ▲RMSE	0.15 [0.11, 0.32]	12.26 [7.86, 18.49]	0.48 [0.31, 0.83]

Bootstrap Uncertainty Summaries

Bootstrap uncertainty summaries are presented in Table V. Velocity showed the narrowest bootstrap uncertainty intervals, followed by spin rate, whereas vertical break demonstrated the widest intervals overall. This ordering was consistent with the broader

model-performance results and suggests that individualized predictions were estimated most precisely for velocity and least precisely for vertical break.

Table V: Bootstrap uncertainty summaries across pitcher-specific N-of-1 performance models by outcome.

Outcome	Mean CI width	Median CI width
		[IQR]
Velocity	3.92	2.96 [1.83]
Spin Rate	4.82	4.42 [2.73]
Vertical Break	6.36	5.65 [3.22]

Discussion

The purpose of this chapter was to examine individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers using an N-of-1 modeling framework. Overall, rolling pitch count associations could be characterized at the pitcher-season level across all three fastball outcomes, although model performance differed by metric. Velocity demonstrated the strongest and most consistent individualized model performance (mean adjusted $R^2 = 0.54$), spin rate showed intermediate performance (mean adjusted $R^2 = 0.44$), and vertical break demonstrated the weakest overall fit (mean adjusted $R^2 = 0.35$). In addition, the modeled associations between rolling pitch count and fastball performance were heterogeneous across pitcher-seasons, particularly for velocity, where individualized curves varied in both magnitude and shape.

Rather than assessing whether rolling pitch count was associated with a single uniform performance response, the present analyses were designed to determine whether those associations could be identified at the individual pitcher level within a broader workload context. The results suggest that they could, but not in a consistent or interchangeable manner across metrics. Across all three outcomes, individualized models captured meaningful variation in fastball performance, yet the extent of that captured variation differed across outcomes and pitcher-seasons. This is important because it suggests that rolling pitch count associations are not equally expressed across all performance metrics, and that some outcomes may be more informative than others when evaluating workload-related performance patterns.

Among the three performance outcomes, velocity emerged as the most consistently captured metric. Velocity models showed the strongest overall fit, the highest adjusted R^2 values, and the clearest individualized curve patterns. Season-specific velocity results indicated that model fit remained moderate-to-strong across all three seasons, while the individualized curve panel demonstrated substantial variability in the shape of the modeled association between rolling pitch count and fastball velocity. Many pitcher-seasons showed declining modeled velocity as rolling pitch count increased, whereas others showed relatively stable, nonlinear, or even increasing patterns across pitch counts. In some cases, increases in velocity may reflect progressive warm-up within outings, competitive ramping, or selective continuation of higher-performing appearances, although these interpretations remain speculative. Collectively, these findings indicate that velocity may be the most practically responsive fastball performance outcome within this modeling framework, not because it follows a uniform directional pattern, but because it is consistently sensitive to workload-related changes across pitchers. This pattern was also

reflected in the bootstrap uncertainty summaries, in which velocity showed narrower uncertainty intervals than spin rate and vertical break.

Spin rate and vertical break provided additional insight, but the associations were less consistently captured than for velocity. Spin rate demonstrated intermediate model performance, with adjusted R^2 values generally lower than velocity but higher than vertical break. Vertical break showed the weakest overall fit and the greatest inconsistency across pitcher-seasons, including a small number of models with adjusted R^2 values near or below zero. This does not necessarily mean that spin rate or vertical break are unimportant workload-related outcomes. Rather, it suggests that their relationship with rolling pitch count may be less direct, more weakly expressed, or more influenced by other measured or unmeasured factors than velocity within the present analytic framework. As a result, velocity appears to provide the clearest individualized signal of performance change, while spin rate and vertical break may function more appropriately as complementary outcomes rather than equally informative primary indicators.

An important finding of this chapter was the degree of heterogeneity observed across pitcher-seasons. Even within the same outcome, the modeled associations between rolling pitch count and performance were not uniform. This variability was evident across the individualized velocity curves and in the spread of adjusted R^2 values across all three outcomes. From the standpoint of the chapter purpose, this reinforces that rolling pitch count associations should not be assumed to operate in the same way for every pitcher. Instead, the findings support the view that pitcher-specific workload-performance relationships may vary substantially in both direction and magnitude, even when the same general exposure metric is used. Heterogeneity was also reflected in predictor selection, as the variables retained during screening were not uniform across pitcher-seasons, further indicating that workload-performance associations differed across individuals.

These findings should also be interpreted within the broader workload context in which the models were developed. Rolling pitch count was the primary exposure of interest, but the models were not intended to treat pitch count as an isolated explanation of performance change. Recovery-related variables, prior workload, game context, travel-related variables, and HRV when available were all considered within the broader modeling framework. This is particularly relevant given that fatigue is often discussed in baseball as a possible contributor to performance changes yet remains difficult to define and measure directly. In the present chapter, the observed performance patterns may be consistent with fatigue-related responses in some pitcher-seasons, but the models do not directly quantify fatigue itself. A more appropriate interpretation is that they characterize individualized workload-related performance patterns associated with rolling pitch count within a broader measured context.

Limitations

Several limitations should be considered when interpreting these findings. First, this study used an observational design, and the modeled associations should not be interpreted as causal effects of rolling pitch count on fastball performance. Second, the chapter focused specifically on fastball performance metrics, and the findings may not generalize to other pitch types or broader aspects of pitcher performance. Third, predictor availability was not identical across seasons, most notably because HRV was only available in the prospectively collected 2024 and 2025 seasons and only among pitchers who participated in wearable monitoring. Relatedly, HRV was obtained from WHOOP, a wrist-worn sensor that derives HRV during sleep, which has been shown to be a less optimal method of HRV assessment than more controlled measurement approaches. Fourth, although the models incorporated a broader workload context, unmeasured factors likely still influenced performance responses. Finally, model performance was not

equally strong across outcomes, particularly for vertical break, which indicates that some performance metrics may be more difficult to characterize reliably within this individualized framework.

Applied Implications and Future Directions

From an applied standpoint, these results suggest that velocity may be the most immediately interpretable fastball performance metric for individualized workload monitoring in this setting. Because velocity demonstrated the strongest and most consistent model performance (mean adjusted $R^2 = 0.54$), it may provide the clearest signal when evaluating how a pitcher's performance changes across rolling pitch count exposure. However, the chapter results also show that no single metric fully captures all workload-related changes. Spin rate and vertical break, despite weaker overall model performance (mean adjusted $R^2 = 0.44$ and 0.35 , respectively), may still contribute useful complementary information in some pitcher-seasons. Accordingly, practical monitoring approaches based on pitch count should likely consider multiple performance outcomes and broader contextual factors rather than assuming that one uniform threshold or one single metric will apply equally across all pitchers.

Future work should build directly from the findings of this chapter by refining how individualized workload-performance patterns are translated into practical monitoring strategies. Additional research is also needed to determine whether alternative workload definitions, expanded contextual covariates, or other performance indicators improve model performance for outcomes such as spin rate and vertical break.

Conclusion

This chapter examined individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers using an N-of-1 modeling framework. Overall, rolling pitch count associations could be characterized at the pitcher-

season level across velocity, spin rate, and vertical break, although model performance differed by outcome. Velocity demonstrated the strongest and most consistent model fit, whereas spin rate showed intermediate performance and vertical break demonstrated weaker and more variable fit. Across outcomes, the modeled associations were heterogeneous across pitcher-seasons, indicating that rolling pitch count was not associated with a single uniform performance response. These findings support the use of individualized modeling to characterize pitcher-specific workload-related performance patterns within a broader measured workload context.

References

1. Reinold MM, Dowling B, Fleisig GS, et al. An Interval Throwing Program for Baseball Pitchers Based upon Workload Data. *IJSPT*. 2024;19(3):326-336. doi:10.26603/001c.94146
2. Shanley E, Bailey L, Sandago MP, Pinkerton A, Singleton SB, Thigpen CA. The use of a pitch count estimator to calculate exposure in collegiate baseball pitchers. *Phys Ther Sport*. 2015;16(4):344-348. doi:10.1016/j.ptsp.2015.02.002
3. Zaremski JL, Pazik M, Vasilopoulos T, Horodyski M. Workload Risk Factors for Pitching-Related Injuries in High School Baseball Pitchers. *Am J Sports Med*. 2024;52(7):1685-1691. doi:10.1177/03635465241246559
4. Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, Robertson SJ. Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *Int J Sports Physiol Perform*. 2017;12(2):230-234. doi:10.1123/ijsp.2015-0791
5. Impellizzeri FM, Marcora SM, Coutts AJ. Internal and External Training Load: 15 Years On. *Int J Sports Physiol Perform*. 2019;14(2):270-273. doi:10.1123/ijsp.2018-0935
6. McLaren SJ, Macpherson TW, Coutts AJ, Hurst C, Spears IR, Weston M. The Relationships Between Internal and External Measures of Training Load and Intensity in Team Sports: A Meta-Analysis. *Sports Medicine*. 2018;48(3):641. doi:10.1007/s40279-017-0830-z
7. Birfer R, Sonne MW, Holmes MW. Manifestations of muscle fatigue in baseball pitchers: a systematic review. *PeerJ*. 2019;7:e7390. doi:10.7717/peerj.7390
8. Djaoui L, Haddad M, Chamari K, Dellal A. Monitoring training load and fatigue in soccer players with physiological markers. *Physiology & Behavior*. 2017;181:86-94. doi:10.1016/j.physbeh.2017.09.004

9. Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. *Sports Med.* 2014;44(S2):139-147. doi:10.1007/s40279-014-0253-z
10. Haddad M, Chaouachi A, Wong DP, et al. Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort. *Physiology & Behavior.* 2013;119:185-189. doi:10.1016/j.physbeh.2013.06.016
11. Grantham WJ, Byram IR, Meadows MC, Ahmad CS. The Impact of Fatigue on the Kinematics of Collegiate Baseball Pitchers. *Orthop J Sports Med.* 2014;2(6):2325967114537032. doi:10.1177/2325967114537032
12. Oeding JF, Boos AM, Kalk JR, et al. Pitch-Tracking Metrics as a Predictor of Future Shoulder and Elbow Injuries in Major League Baseball Pitchers: A Machine-Learning and Game-Theory Based Analysis. *Orthop J Sports Med.* 2024;12(8):23259671241264260. doi:10.1177/23259671241264260
13. McDonald S, Vieira R, Johnston DW. Analysing N-of-1 observational data in health psychology and behavioural medicine: a 10-step SPSS tutorial for beginners. *Health Psychol Behav Med.* 8(1):32-54. doi:10.1080/21642850.2019.1711096
14. Vieira R, McDonald S, Araújo-Soares V, Sniehotta FF, Henderson R. Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies. *Health Psychology Review.* 2017;11(3):222-234. doi:10.1080/17437199.2017.1343680
15. Kravitz RL, Duan N, Braslow J. Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages. *Milbank Quarterly.* 2004;82(4):661-687. doi:10.1111/j.0887-378x.2004.00327.x
16. Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. Published online May 14, 2015. doi:10.1136/bmj.h1738

17. Samuel JP, Wootton SH, Tyson JE. N-of-1 trials: The epitome of personalized medicine? *J Clin Transl Sci*. 7(1):e161. doi:10.1017/cts.2023.583
18. Selker HP, Cohen T, D'Agostino RB, et al. A Useful and Sustainable Role for N-of-1 Trials in the Healthcare Ecosystem. *Clin Pharma and Therapeutics*. 2022;112(2):224-232. doi:10.1002/cpt.2425
19. Daza EJ. Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Track Health Data. *arXiv*. Preprint posted online January 22, 2019:arXiv:1901.03423. doi:10.48550/arXiv.1901.03423
20. Rossiter A, Warrington GD, Comyns TM. Effects of Long-Haul Travel on Recovery and Performance in Elite Athletes: A Systematic Review. *The Journal of Strength & Conditioning Research*. 2022;36(11):3234. doi:10.1519/JSC.0000000000004021
21. Janse Van Rensburg DC, Jansen Van Rensburg A, Fowler P, et al. How to manage travel fatigue and jet lag in athletes? A systematic review of interventions. *Br J Sports Med*. 2020;54(16):960-968. doi:10.1136/bjsports-2019-101635
22. Araujo A, Julious S, Senn S. Understanding Variation in Sets of N-of-1 Trials. *PLoS One*. 2016;11(12):e0167167. doi:10.1371/journal.pone.0167167
23. Kwasnicka D, Naughton F. N-of-1 methods: A practical guide to exploring trajectories of behaviour change and designing precision behaviour change interventions. *Psychology of Sport and Exercise*. 2020;47:101570. doi:10.1016/j.psychsport.2019.101570
24. International Olympic Committee Injury and Illness Epidemiology Consensus Group, Bahr R, Clarsen B, et al. International Olympic Committee Consensus Statement: Methods for Recording and Reporting of Epidemiological Data on Injury and Illness in Sports 2020 (Including the STROBE Extension for Sports Injury and Illness Surveillance

- (STROBE-SIIS)). *Orthop J Sports Med.* 2020;8(2):2325967120902908. doi:10.1177/2325967120902908
25. Wilson DT, Hooper R, Brown J, Farrin AJ, Walwyn RE. Efficient and flexible simulation-based sample size determination for clinical trials with multiple design parameters. *Stat Methods Med Res.* 2021;30(3):799-815. doi:10.1177/0962280220975790
 26. Senn S. Sample size considerations for n-of-1 trials. *Stat Methods Med Res.* 2019;28(2):372-383. doi:10.1177/0962280217726801
 27. TrackMan Baseball - Data and Actionable insights. Accessed February 27, 2025. <https://www.trackman.com/baseball>
 28. Addleman JS, Lackey NS, DeBlauw JA, Hajduczuk AG. Heart Rate Variability Applications in Strength and Conditioning: A Narrative Review. *J Funct Morphol Kinesiol.* 2024;9(2):93. doi:10.3390/jfmk9020093
 29. DONG JG. The role of heart rate variability in sports physiology. *Exp Ther Med.* 2016;11(5):1531-1536. doi:10.3892/etm.2016.3104
 30. Nathan AM. The effect of spin on the flight of a baseball. *American Journal of Physics.* 2008;76(2):119-124. doi:10.1119/1.2805242
 31. Kimura A, Nakashima H, Kuroyanagi S, Ando Y, Liao P, Sakurai S. Estimating flight trajectories of breaking balls from four-seam fastballs. *Front Sports Act Living.* 2023;5:1092520. doi:10.3389/fspor.2023.1092520
 32. Whiteside D, McGinnis RS, Deneweth JM, Zernicke RF, Goulet GC. Ball flight kinematics, release variability and in-season performance in elite baseball pitching. *Scand J Med Sci Sports.* 2016;26(3):256-265. doi:10.1111/sms.12443

33. Nicholas Tierney, Dianne Cook. Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. *Journal of Statistical Software*. 2023;105(7):1-31. doi:10.18637/jss.v105.i07
34. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010;33:1-22. doi:10.18637/jss.v033.i01
35. Van Calster B, van Smeden M, De Cock B, Steyerberg EW. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. *Stat Methods Med Res*. 2020;29(11):3166-3178. doi:10.1177/0962280220921415
36. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Published online October 4, 2000:1.9-4. doi:10.32614/CRAN.package.mgcv
37. Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA. A linear exponent AR(1) family of correlation structures. *Stat Med*. 2010;29(17):1825-1838. doi:10.1002/sim.3928
38. Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*. 2019;11(3):e1460. doi:10.1002/wics.1460
39. Raykov T, DiStefano C. Evaluating Change in Adjusted R-Square and R-Square Indices: A Latent Variable Method Application. *Educ Psychol Meas*. Published online April 11, 2025:00131644251329178. doi:10.1177/00131644251329178
40. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014;7(3):1247-1250. doi:10.5194/gmd-7-1247-2014

CHAPTER III: N-of-1 Modeling of Rolling Fastball Pitch Count Associations with Pitching Biomechanics in Collegiate Baseball Pitchers

Garrett Fernandez^{1,2}, Garrett Bullock^{1,3}, Patrick Ward⁴, Joel Stitzel^{1,2}, Ryan McGinnis¹,
Tanner Filben⁵, Kristen Nicholson^{1,3}

¹Department of Biomedical Engineering, Wake Forest School of Medicine, Winston-Salem, NC, ²Virginia Tech-Wake Forest School of Biomedical Engineering and Sciences, Winston-Salem, NC, ³Department of Orthopedic Surgery and Rehabilitation, Wake Forest School of Medicine, Winston Salem, NC, ⁴Seattle Seahawks, Seattle, WA, USA, ⁵Elemance, LLC, Winston-Salem, NC

Abstract

Pitching biomechanics has become an important part of baseball research, monitoring, and interpretation, but how biomechanical responses change over time with accumulated pitch count remains unclear. Therefore, the purpose of this chapter was to examine individualized associations between rolling pitch count and pitching biomechanics in collegiate baseball pitchers using an N-of-1 modeling framework. Biomechanical outcomes reflected joint loading, segment positioning, and kinetic-chain organization across the 2024 and 2025 seasons, with 24 pitcher-seasons, 3,582 fastball observations, and 260 appearances included in the analytic sample. Final models were developed separately for each pitcher-season and biomechanical outcome using LASSO-based predictor screening followed by generalized additive mixed modeling, with rolling pitch count retained as the primary exposure and contextual covariates related to recovery, travel, and workload included when available. Across the 11 modeled outcomes, mean adjusted R^2 values were generally moderate, ranging from 0.30 to 0.60 in 2024 and from 0.31 to 0.61 in 2025. Maximum hip-shoulder separation demonstrated the strongest overall performance, with adjusted R^2 values of 0.60 in 2024 and 0.61 in 2025, whereas maximum shoulder rotational torque showed the lowest fit in 2024 (adjusted $R^2 = 0.30$) and lead knee flexion at ball release showed the lowest fit in 2025 (adjusted $R^2 = 0.31$). Maximum elbow varus torque also improved from an adjusted R^2 of 0.44 in 2024 to 0.57 in 2025, with RMSE decreasing from 7.52 to 5.35. Residual autocorrelation remained small overall, ranging from -0.079 to 0.077. Overall, rolling pitch count associations with biomechanics could be characterized at the pitcher-season level, but the strength and consistency of those associations differed across both athletes and outcomes, supporting individualized modeling for biomechanical workload-response analysis in collegiate baseball pitchers.

Introduction

Pitching biomechanics has become an increasingly common part of baseball performance analysis.^{1–5} Advances in motion-capture systems and pitch-tracking technology have made it possible to quantify specific biomechanical indicators during the pitching motion, including measures related to joint loading, segment positioning, and kinetic-chain coordination.^{6,7} In applied baseball settings, variables such as elbow varus torque, shoulder distraction force, shoulder abduction, elbow flexion, lead knee mechanics, and hip-shoulder separation are often used as key indicators of how a pitcher generates and distributes mechanical demands throughout the motion.^{1,2,4,5,7–9} These measures are of interest because they provide information beyond what can be learned from ball-flight outcomes alone. Whereas velocity, spin rate, and pitch movement describe characteristics of the pitch itself, biomechanics offer a more direct view of how that pitch was produced.

This distinction may be particularly important when evaluating pitcher response to workload.^{10–12} In baseball, accumulated throwing exposure is often monitored using pitch count because it is simple, accessible, and routinely available in both practice and competition settings.^{10–13} However, the physiological and mechanical consequences of workload are unlikely to appear uniformly across pitchers. Some athletes may show workload-related responses through visible changes in performance outcomes, whereas others may maintain those outcomes while exhibiting alterations in movement patterns or redistribution of mechanical demands across the kinetic chain.^{14,15} For that reason, an exclusive focus on ball-flight metrics may provide an incomplete picture of how pitchers respond to accumulated throwing load.

The broader concept of fatigue helps frame this problem but also highlights its complexity. Fatigue is frequently referred to in sport when athletes demonstrate performance decline, mechanical change, or altered recovery, yet it remains difficult to define and measure directly.^{14–16} In throwing athletes, fatigue-related responses are not expected to follow a single universal pattern. Instead, the expression of workload-related stress may differ according to the athlete, the competitive context, and the outcome being observed. From this perspective, biomechanics are not simply secondary to performance outcomes but represent another major domain in which workload responses may emerge.^{14–16}

Despite increasing use of biomechanical analysis in baseball, much of the existing workload literature has relied on group-based approaches that estimate average responses across athletes.^{2,4,7,11,14} These approaches are valuable for identifying broader patterns, but they may obscure important heterogeneity in biomechanical response.^{17–20} Pitchers exposed to similar rolling pitch counts may not respond in the same way, and even within the same pitcher, different biomechanical variables may show different levels of sensitivity to workload. One pitcher may demonstrate relatively stable mechanics despite increasing pitch count, whereas another may show meaningful changes in joint loading, timing, or segment organization.^{14,15,21} These differences are especially important in applied settings, where decisions are made for individual athletes rather than for an average member of a group.

An individualized analytic framework may therefore be better suited for studying these relationships. N-of-1 methods shift the unit of inference from the group to the individual, allowing repeated measures from a single athlete to be used to estimate within-pitcher associations over time.^{17–19,22–24} This approach is well aligned with pitching biomechanics data, where repeated observations can be collected longitudinally and

where relationships with workload may vary both across pitchers and across biomechanical outcomes within the same pitcher. By focusing on the individual pitcher, N-of-1 models provide an opportunity to examine whether rolling pitch count is associated with meaningful biomechanical changes within that athlete, rather than assuming that a single average response applies across the entire staff. This framing is also consistent with the broader dissertation rationale for using individualized methods when responses to workload are expected to be heterogeneous.^{17–19,23,24}

At the same time, rolling pitch count should not be interpreted in isolation. Biomechanical outputs are influenced by a broader workload environment that includes recent throwing history, recovery, travel-related demands, and internal load.^{10,12,25} As a result, models based only on pitch count may oversimplify the processes underlying mechanical change. Incorporating contextual variables provides a more complete framework for evaluating workload-related biomechanical patterns, while acknowledging that no single observed set of variables can fully explain all sources of variation in pitching mechanics.

Therefore, the purpose of this chapter was to examine individualized associations between rolling pitch count and pitching biomechanics in collegiate baseball pitchers using an N-of-1 modeling framework. Biomechanical outcomes reflecting joint loading, segment positioning, and kinetic-chain organization were evaluated to characterize how workload-related responses may be expressed mechanically within pitchers. By modeling these outcomes within pitchers and within a broader measured workload context, this chapter sought to characterize biomechanical patterns associated with rolling pitch count, determine the extent to which those patterns varied across pitchers, and assess whether responsiveness differed across biomechanical outcomes.

Methods

Study Design

This study used a longitudinal observational design to examine individualized associations between rolling pitch count and pitching biomechanics in collegiate baseball pitchers using an N-of-1 modeling framework.^{17,19,20,22,26,27} The analytic focus was on within-pitcher associations over time rather than population-average effects. Data were drawn from the 2024 and 2025 collegiate baseball seasons, and each pitcher-season was treated as a distinct longitudinal unit to preserve season-specific workload context and allow workload-related biomechanical associations to vary across both pitchers and seasons. This chapter focused on fastball pitching biomechanics outcomes reflecting joint loading, segment positioning, and kinetic-chain organization during the pitching motion.

Study Participants and Eligibility Criteria

This study was approved by the Wake Forest University School of Medicine Institutional Review Board. The study sample consisted of collegiate baseball pitchers with available pitching biomechanics data during the study period. Analyses were conducted at the pitcher-season level, such that each eligible pitcher within each season contributed an individualized observational series. This approach was selected to preserve within-season workload context and to allow workload-related biomechanical associations to vary across both pitchers and seasons.

Pitcher-seasons were eligible for inclusion if they contained pitch-level biomechanics data and corresponding rolling pitch count information sufficient for individualized longitudinal modeling. Because the analytic goal was to estimate within-pitcher associations, pitcher-seasons were excluded when the number of usable pitch-level observations was insufficient for model development. For each pitcher-season and

biomechanical outcome, initial LASSO-based predictor screening was performed only when at least 50 usable observations were available. Subsequent modeling procedures were then applied only to eligible pitcher-seasons meeting these data sufficiency requirements.

Sample Size and Data Sufficiency

Data sufficiency was evaluated using both the broader simulation-informed planning framework applied in this dissertation and empirical checks performed within the final individualized modeling pipeline. During study planning, simulation-based precision analyses were used to inform feasible model complexity for repeated-measures longitudinal modeling rather than to define a single fixed sample-size threshold for all final models.^{28,29} Based on that framework, final individualized models were restricted to a maximum of five predictor variables in addition to the primary pitch-count term, with allowance for up to one interaction term where supported by the data and model fit. This constraint was used to balance model flexibility with data sufficiency at the pitcher-season level.

For the final analyses, data sufficiency was evaluated separately for each pitcher-season and biomechanical outcome, consistent with the N-of-1 design. Initial predictor screening was performed only when at least 50 usable observations were available for a given pitcher-season and outcome. Final model adequacy was further examined through stabilization analyses, trial-reduction sensitivity analyses, and model-specification sensitivity analyses evaluating the effects of reduced data availability or altered model structure. This combined approach was used to assess data adequacy for individualized longitudinal modeling while accounting for variation in the amount of repeated biomechanics data available across pitcher-seasons and outcomes.

Data Sources and Variable Collection

Data were assembled from season-level pitching datasets spanning the 2024 and 2025 collegiate baseball seasons. Pitch-level biomechanics data were obtained using KinaTrax markerless motion capture system (Kinatrax, Boca Raton, FL), across both laboratory bullpen and on-field collection settings and were synchronized with rolling pitch count and related workload variables for individualized modeling. Only fastball observations were retained for analysis, consistent with the broader modeling framework used throughout this dissertation. Pitch count and additional baseball-related contextual variables were obtained from the Wake Forest Baseball Analytics team's records, while internal load data were obtained using WHOOP, a wrist-worn monitoring device for pitchers who elected to wear it, such that HRV data were available only for a subset of pitcher-seasons.^{30,31}

The biomechanics dataset included pitch-level variables reflecting joint loading, segment positioning, and kinetic-chain organization during the pitching motion. Pitcher identity, season, and appearance were retained as indexing variables for longitudinal analysis. Prior to model development, season files were synchronized to ensure consistent variable naming, coding structure, and data type formatting across years while preserving the original pitch-level structure of the observations. Because biomechanics data required specialized motion-capture infrastructure, the available dataset was more constrained than the ball-metric dataset used in Chapter 2, resulting in variation in the amount of repeated data available across pitchers, seasons, and outcomes.

Variable Definitions

Pitching Biomechanics

Pitching biomechanics variables were selected to represent key mechanical characteristics of the throwing motion related to joint loading, upper-extremity positioning, and kinetic-chain organization. In baseball, these variables are commonly interpreted as

biomechanical key performance indicators because they describe how force is generated, transferred, and managed throughout the pitching motion rather than only describing the resulting ball flight.^{2,5,32} Prior biomechanics literature has identified variables related to elbow loading, shoulder loading, lead-leg mechanics, and hip-shoulder separation as particularly relevant in throwing analysis due to their relationships with both performance and injury-related mechanical demands.^{4,33–35}

All biomechanical outcomes were analyzed at the pitch level and were restricted to fastball observations. Fastballs were selected to maintain consistency with the broader dissertation framework and to reduce variation attributable to differing pitch-type mechanics. Because these biomechanical variables capture different aspects of the pitching motion, including joint stress, joint positioning, and whole-body coordination, they were evaluated separately rather than combined into a single composite mechanical outcome. This allowed associations between rolling pitch count and biomechanics to vary both across pitchers and across specific biomechanical domains. The set of outcomes modeled in the final pipeline included elbow flexion angle at ball release, lead knee flexion angle at ball release, hip-shoulder separation at foot contact, maximum hip-shoulder separation, shoulder rotation angle at maximum external rotation, shoulder abduction angle at ball release, maximum resultant shoulder distraction force, elbow flexion angle at maximum external rotation, bodyweight-normalized maximum elbow varus torque, maximum shoulder rotational torque, and maximum elbow varus torque.

Elbow Variables

Elbow-related variables used included maximum elbow varus torque, bodyweight-normalized maximum elbow varus torque, elbow flexion angle at ball release, and elbow flexion angle at maximum external rotation. Maximum elbow varus torque was defined as the highest varus loading experienced at the elbow during the pitching motion.^{16,33,35} This

variable is one of the most commonly studied throwing biomechanics metrics and has been widely used as an indicator of medial elbow stress during high-demand phases of the throw. Prior literature has described a critical instant just before maximum external rotation at which peak elbow varus torque is often observed, reflecting the large valgus loading demands placed on the elbow during late cocking and early acceleration.^{16,33,35} Bodyweight-normalized maximum elbow varus torque was defined as this same peak elbow loading value scaled relative to body size to improve interpretability across pitchers of different builds. Elbow flexion angle at ball release was defined as the elbow joint angle at the instant the baseball left the hand.^{3,4,33} Elbow flexion angle at maximum external rotation was defined as the elbow joint angle at the time point when the shoulder reached maximum external rotation.^{4,5,33} These positional variables were included because elbow configuration during the throwing motion influences upper-extremity lever mechanics and may affect how rotational and translational loads are distributed across the arm.

Shoulder Variables

Shoulder-related variables used included maximum shoulder rotational torque, maximum resultant shoulder distraction force, shoulder abduction angle at ball release, and shoulder rotation angle at maximum external rotation. Maximum shoulder rotational torque was defined as the peak rotational loading about the shoulder generated during the pitching motion.^{5,32,35,36} This variable reflects the substantial rotational demands placed on the glenohumeral joint during late cocking and acceleration, when potential energy is transferred through the upper extremity to the ball.^{32,37,38} Maximum resultant shoulder distraction force was defined as the highest distractive force experienced at the shoulder during the throw.^{4,35,37} Prior throwing biomechanics literature describes this loading as especially relevant after ball release, when the arm must be decelerated rapidly and the shoulder experiences substantial distraction demands during follow-through and

deceleration.^{37,38} Shoulder abduction angle at ball release was defined as the degree of humeral abduction at the instant of ball release.^{4,35} Shoulder rotation angle at maximum external rotation was defined as the shoulder rotational position at the time point of maximum external rotation.^{4,5} Maximum external rotation is a commonly referenced event in the pitching motion because it represents a phase of high shoulder and elbow stress and is frequently used as a temporal landmark for biomechanical evaluation.^{5,32} Together, these shoulder variables were used to capture both loading magnitude and upper-extremity positional characteristics relevant to force production and force management during the throw.

Lower-Extremity and Kinetic-Chain Variables

Lower-extremity and kinetic-chain variables used included lead knee flexion angle at ball release, hip-shoulder separation at foot contact, and maximum hip-shoulder separation. Lead knee flexion angle at ball release was defined as the flexion angle of the lead knee at the instant of ball release.^{4,9,32} Lead-leg mechanics are relevant because the lead leg functions as a stabilizing base that helps dissipate and transfer energy up the kinetic chain during the pitch.^{4,34,39,40} Prior literature has described lead knee flexion as an important lower-extremity characteristic during the transition from stride through acceleration, with reduced ability to stabilize or extend the lead leg potentially altering energy transfer demands to the trunk and upper extremity.^{4,9,34,40} Hip-shoulder separation at foot contact was defined as the rotational difference between the pelvis and trunk at the instant of front-foot contact.^{1,4,37} Maximum hip-shoulder separation was defined as the greatest rotational difference between the pelvis and trunk observed during the pitching motion. Hip-shoulder separation is a commonly used measure of kinetic-chain function because it reflects how effectively lower-extremity and trunk rotation are sequenced before energy is transferred to the throwing arm.^{4,37,40,41} Prior work has described hip-shoulder

separation as a key indicator of torsional loading through the trunk and a meaningful contributor to energy transfer efficiency in baseball pitching.

Pitching Exposure

Pitching exposure, also referred to as pitch count, was defined as the total number of pitches thrown within a given appearance.^{12,13,42} For modeling purposes, rolling pitch count served as the primary exposure variable and was used to represent accumulated throwing volume over the course of an appearance. Pitching appearances were indexed longitudinally within pitcher-season series to preserve the sequential workload context of repeated observations. In addition to the primary rolling pitch count term, lagged pitch-count variables were constructed to represent recent pitch-count history from prior observations when supported by the available data structure. Previous appearance maximum pitch count was also considered as a marker of recent throwing burden carried into the current pitching context. This broader exposure structure was used to better reflect recent workload history rather than treating each pitch as independent of prior throwing volume.

Contextual Covariates

Additional workload-related covariates were included to provide broader context for interpreting associations between rolling pitch count and biomechanics. Internal load was assessed using heart rate variability derived from WHOOP, a wrist-worn monitoring device used by pitchers who elected to wear it. Because participation in WHOOP monitoring was not universal, HRV data were available only for a subset of pitcher-seasons.^{30,31} External and contextual covariates included rest, game context, and selected recent workload history variables.^{11,21} Rest was defined as recovery time between appearances. Game context distinguished competitive game settings from practice-based throwing settings when such information was available. Travel-related variables were

considered during data assembly and preprocessing, but due to the biomechanics dataset being limited to local appearances and home-game conditions, many travel variables had limited usable data and were not consistently retained across final pitcher-outcome models. Together, these internal and external covariates were used to place rolling pitch count within a broader workload environment rather than interpreting pitch count as a standalone measure of biomechanical response.

Missing Data

All data were assessed for missingness before statistical modeling. Missing data prevalence and availability were evaluated at both the overall dataset level and within pitcher-season series, because the N-of-1 framework required sufficient repeated observations within individual pitchers rather than only across the full sample. Missingness was examined for biomechanics data, rolling pitch count, and candidate workload-related covariates to determine whether variables were consistently available for individualized model development.

Missing data were handled using multiple imputations with chained equations, consistent with the broader analytical framework used in this dissertation.⁴³ This approach was used to preserve analytically useful information and reduce unnecessary data loss while accounting for incomplete covariate information. However, because the biomechanics analyses in this chapter were conducted separately by pitcher-season, missingness was also evaluated within each individualized series before model development. Variables with structurally limited availability, particularly certain contextual covariates, were interpreted cautiously within the modeling pipeline. This was especially relevant for variables such as HRV and some travel-related measures, which were unavailable for subsets of pitcher-seasons due to monitoring participation patterns or the constrained biomechanics sampling structure.

Biomechanics outcomes were collected as a common pitch-level set, such that when biomechanics data were missing for a given pitch, all modeled biomechanics variables were missing for that observation rather than only a single outcome. Accordingly, missingness in the biomechanics data primarily reflected pitch-level capture availability and broader differences in available repeated observations across pitchers and seasons, rather than substantial variation in missingness across individual biomechanics variables. Final analyses were therefore restricted to usable pitch-level observations with available biomechanics data and primary pitch-count exposure, while candidate predictors were evaluated for sufficient within-series availability before inclusion. Because the primary objective of this chapter was to characterize individualized biomechanical associations using observed longitudinal data, robustness to incomplete data structure was further evaluated through stabilization, model-specification sensitivity, and trial-reduction analyses.

Data Processing and Feature Construction

Prior to model development, biomechanics data from the 2024 and 2025 seasons were processed to create consistent pitcher-season longitudinal series for analysis. Dates and date-time fields were standardized, core identifiers were retained, and modeling variables were formatted to consistent data types across seasons. Continuous variables used in modeling were converted to numeric format when necessary, and categorical variables such as transportation were recoded into consistent factor levels.

Only fastball observations were retained for analysis. Appearances were defined at the pitcher-date level, such that each pitcher-session was treated as a distinct appearance for longitudinal modeling. Pitch-level observations were then ordered within appearance by pitch count to preserve the sequential structure of the throwing session.

This appearance-based structure was used throughout the individualized modeling pipeline.

Several derived features were then constructed for modeling. Appearance was generated as a pitcher-by-date identifier when not already available in the source data. Lagged pitch-count variables were created to represent recent throwing history, including up to three prior pitch-count terms when supported by the available data. Previous appearance workload information was also retained when available to reflect recent carryover workload. In addition, a within-appearance sequential time index was created to support autocorrelation modeling of repeated pitch-level observations.

Candidate predictor variables were also screened during preprocessing to support stable individualized model fitting. Numeric predictors were evaluated for within-series variation and availability, and predictors with no meaningful usable variation within a pitcher-season series were excluded from downstream consideration. Transportation was cleaned into consistent levels before modeling and was excluded when it lacked sufficient variation within a given series. Travel-related variables were considered during preprocessing, but because biomechanics data were limited to local appearances and home-game conditions for game observations, many travel-related variables had limited usable information and were not consistently retained across final models.

N-of-1 Model Development

Individualized model development was performed separately for each pitcher-season and for each fastball biomechanics outcome. Model construction followed a two-stage process consisting of predictor screening followed by final pitcher-season-specific model fitting. As in the broader dissertation framework, allowable model complexity was constrained to balance flexibility with the amount of repeated data available within each individualized series. Final models were therefore limited to a maximum of five predictors

in addition to the primary pitch-count term, with no more than one interaction term retained in a given final model.

During predictor screening, LASSO regression was used to identify a reduced set of candidate predictors for each pitcher-season and biomechanics outcome.^{44,45} Screening models were fit using cross-validated Gaussian glmnet procedures with standardized predictors. Rolling pitch count was forced into all screening models by assigning it a zero penalty, ensuring retention of the primary workload exposure regardless of selection of additional variables. Candidate predictors reflected recent workload history, recovery, HRV, game context, and available travel-related variables. When multiple predictors represented overlapping constructs, including lagged pitch-count terms and travel-related variables, only the strongest variable from a given group was retained for subsequent modeling. After screening, the retained predictor set was carried forward into individualized biomechanics model fitting.

Final individualized models were then fit using generalized additive mixed models (GAMMs) implemented with `mgcv::gamm`, with separate models estimated for each pitcher-season and biomechanics outcome. Rolling pitch count was modeled as a smooth term using `s(PitchCount, k = 4)` to allow nonlinear associations with biomechanical responses across the course of an appearance.⁴⁶ Predictors retained from the screening stage were entered as linear terms unless evaluated as part of a candidate interaction. Candidate interactions were tested one at a time, and at most one interaction was retained in any final model. For numeric variables, interactions with pitch count were modeled using tensor product smooths of the form `te(PitchCount, Z, k = c(4,4))`. For transportation, interactions were represented using `s(PitchCount, by = Transportation, k = 4) + Transportation`.⁴⁶ When an interaction term was included, the corresponding predictor was not simultaneously retained as an additional linear term.

For each pitcher-season and biomechanics outcome, multiple candidate models were evaluated, including a base model and interaction-augmented alternatives when supported by the available data. Each candidate model was then fit with and without an AR(1) residual correlation structure.⁴⁷ Because the biomechanics data consisted of sequential pitch-level observations clustered within appearances, the AR(1) structure was specified within appearance using the within-appearance time index, and a random intercept for appearance was included to account for repeated pitches within the same session. Final model selection was based primarily on Akaike information criterion, with secondary consideration given to model fit and interpretability. Final model objects and summary outputs were retained separately for each pitcher-season and biomechanics outcome.

Model Evaluation, Diagnostics, and Sensitivity Analyses

Final individualized model performance was evaluated separately for each pitcher-season and biomechanics outcome. For the purposes of chapter-level reporting, model fit was summarized primarily using root mean squared error (RMSE) and adjusted coefficient of determination (adjusted R^2).^{48,49} RMSE was used to quantify average prediction error in the original scale of each biomechanical outcome, whereas adjusted R^2 was used to describe within-series explanatory performance while accounting for model complexity. Model selection was guided primarily by Akaike Information Criterion (AIC), with Bayesian Information Criterion (BIC) examined as a secondary check to assess whether model selection was sensitive to differences in complexity penalties.⁵⁰ Mean squared error (MSE) and unadjusted R^2 were retained in underlying model outputs as supplemental evaluation metrics, but were not emphasized in the main results tables because RMSE and adjusted R^2 provided the clearest summary of comparative model performance across outcomes

and seasons. Final summary outputs were stored separately for each pitcher-season and biomechanics outcome.

Residual diagnostics were used to assess the adequacy of final individualized model fit. For each retained model, diagnostic plots included residuals versus fitted values, quantile-quantile plots of residuals, residuals ordered by pitch index, and autocorrelation function (ACF) plots of residuals. These diagnostics were used to evaluate homoscedasticity, departures from normality, temporal patterning in residuals, and remaining serial dependence after model fitting. AR(1) correlation structures were evaluated during final model fitting and were retained when they improved model fit based on AIC. Because the biomechanics data consisted of sequential pitch-level observations clustered within appearances, assessment of residual dependence was an important component of final model adequacy.

Sensitivity analyses were performed to examine the robustness of individualized biomechanics models to data availability and modeling decisions. Stabilization analyses were used to assess how model behavior changed as more observations accumulated within a pitcher-season series. Model-specification sensitivity analyses were then performed by comparing final model performance to alternative model formulations, with changes in RMSE used as the primary summary of specification sensitivity. Trial-reduction sensitivity analyses were also conducted to evaluate the impact of reduced sample density on model performance by removing a proportion of pitch-level observations and summarizing the resulting change in RMSE. Together, these procedures were used to determine whether final individualized biomechanics models were interpretable, stable, and adequately fit to the repeated pitch-level data.

Statistical Analysis

Statistical analyses were performed using an individualized N-of-1 framework to evaluate within-pitcher associations between rolling pitch count and fastball biomechanics outcomes. Analyses were conducted separately by pitcher-season and by biomechanics outcome. Descriptive statistics, individualized model fitting, residual diagnostics, and sensitivity analyses were used to evaluate model performance, stability, and heterogeneity of associations across pitchers.

All data preprocessing, model development, model evaluation, sensitivity analyses, and visualization were performed in R version 4.5.1 (R Core Team, 2025). Key packages used in the Aim 2 biomechanics pipeline included dplyr 1.1.4, ggplot2 3.5.2, readxl 1.4.5, readr 2.1.5, tidyr 1.3.1, lubridate 1.9.4, car 3.1.3, cli 3.6.5, glmnet 4.1.9, mgcv 1.9.3, nlme 3.1.168, purrr 1.1.0, stringr 1.5.1, and tibble 3.3.0. Sample code illustrating the model development pipeline and overall analytic workflow is provided in Appendix D.

Results

Descriptive Statistics

Across the 2024 and 2025 seasons, the analytic biomechanics sample included 24 pitcher-seasons, 3,582 fastball observations, and 260 appearances from collegiate pitchers aged 18 to 24 years. In 2024, the biomechanics analyses included 9 pitcher-seasons contributing 1,332 pitches across 85 appearances, whereas in 2025 they included 15 pitcher-seasons contributing 2,250 pitches across 175 appearances. This increase in recorded pitches and appearances from 2024 to 2025 primarily reflected expanded team use of motion-capture systems rather than a substantive change in the analytic framework. Relative to the broader ball-metric dataset, the biomechanics sample remained more constrained because biomechanical data collection depended on motion-capture availability, resulting in fewer total pitches and appearances available for

individualized modeling. Extended descriptive statistics for each biomechanics outcome variable including mean and standard deviation are provided in the Appendix B.

Relative variability across appearances within pitchers is shown in Figure 9. Median relative variability differed meaningfully across biomechanics signals. Maximum shoulder rotational torque demonstrated the greatest between-appearance variability, followed by lead knee flexion at ball release and elbow flexion at ball release. In contrast, shoulder rotation at maximum external rotation and shoulder abduction at ball release demonstrated the lowest relative variability. The remaining variables showed intermediate levels of fluctuation across appearances. Relative variability therefore differed across biomechanics signals, with some variables demonstrating greater between-appearance dispersion than others.

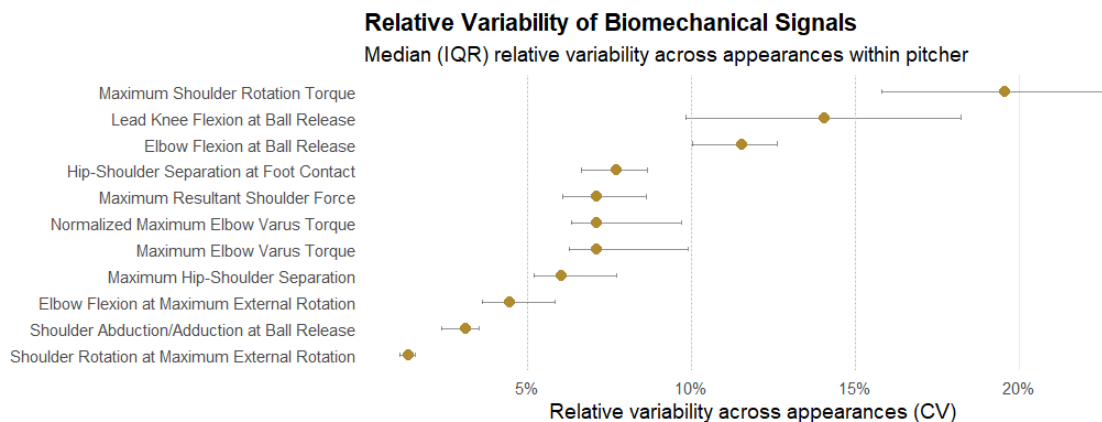


Figure 9: Relative variability of fastball biomechanics outcomes across appearances. Median and interquartile range of coefficient of variation values for each modeled biomechanics outcome across appearances within pitchers. Outcomes are ordered from greatest to least relative variability.

Biomechanics Model Performance Overview

Final individualized biomechanics model performance varied across outcomes and between seasons. Mean outcome-level performance summaries across pitcher-specific models are presented in Table VI, and the overall distribution of adjusted R^2 values across

all biomechanics models is shown in Figure 10. Overall, adjusted R^2 values were moderate, with mean values for most outcomes falling in the approximate range of 0.30 to 0.60 across seasons. Across the 11 modeled outcomes, 2025 generally demonstrated similar or improved mean model performance relative to 2024, particularly for several loading-related elbow and shoulder variables.

Table VI: Mean performance of individualized biomechanics models by outcome and season

Season	Elbow Flexion at BR	Elbow Flexion at MER	Hip Shoulder Separation at FC	Lead Knee Flexion at BR	Max Elbow Varus Torque	Max Hip Shoulder Separation
2024 Adj. R^2	0.43	0.32	0.59	0.38	0.44	0.60
2025 Adj. R^2	0.48	0.37	0.57	0.31	0.57	0.61
2024 RMSE	1.84	2.93	2.87	5.23	7.52	2.25
2025 RMSE	1.82	2.94	2.45	4.87	5.35	2.13

Note. Values represent the mean adjusted R^2 and mean RMSE across all final pitcher-specific models within each season and outcome. BR = ball release; MER = maximum external rotation; FC = foot contact; Adj. R^2 = adjusted coefficient of determination; RMSE = root mean squared error.

Table VI (continued): Mean performance of individualized biomechanics models by outcome and season

Season	Max Resultant Shoulder Force	Max Shoulder Rotational Torque	Normalized Max Elbow Varus Torque	Shoulder Abduction/Adduction at BR	Shoulder Rotational at MER
2024 Adj. R²	0.34	0.30	0.43	0.47	0.50
2025 Adj. R²	0.49	0.52	0.54	0.51	0.52
2024 RMSE	80.59	17.10	0.01	1.81	2.03
2025 RMSE	63.33	14.53	0.01	1.82	1.52

Note. Values represent the mean adjusted R² and mean RMSE across all final pitcher-specific models within each season and outcome. BR = ball release; MER = maximum external rotation; Adj. R² = adjusted coefficient of determination; RMSE = root mean squared error.

Among the elbow and lower-extremity or kinetic-chain outcomes, hip-shoulder separation at foot contact and maximum hip-shoulder separation demonstrated the highest mean adjusted R² values in both seasons. Maximum elbow varus torque also showed stronger mean model performance in 2025 than in 2024, increasing from an adjusted R² of 0.44 to 0.57, with a corresponding reduction in RMSE from 7.52 to 5.35. In contrast, elbow flexion at ball release and elbow flexion at maximum external rotation showed comparatively lower but still moderate mean adjusted R² values, ranging from 0.32 to 0.48 across seasons. Lead knee flexion at ball release demonstrated moderate

mean fit overall, although adjusted R^2 was lower in 2025 than in 2024 despite a modest reduction in RMSE.

For the shoulder-related outcomes, shoulder rotation at maximum external rotation and shoulder abduction/adduction at ball release demonstrated moderate and relatively consistent mean adjusted R^2 values across seasons. Maximum resultant shoulder force and maximum shoulder rotational torque showed clearer improvement in 2025, with adjusted R^2 increasing from 0.34 to 0.49 and from 0.30 to 0.52, respectively. These changes were accompanied by lower RMSE values in 2025, indicating improved mean model fit for these outcomes in the later season. Bodyweight-normalized maximum elbow varus torque similarly demonstrated stronger mean performance in 2025 than in 2024, with adjusted R^2 increasing from 0.43 to 0.54, while RMSE remained low and unchanged at 0.01 due to the smaller scale of that outcome.

The distributional summaries shown in Figure 10 demonstrate that overall biomechanics model performance differed between seasons. Adjusted R^2 values were generally higher in 2025 than in 2024, with the 2025 distribution showing a higher median and a broader upper range. At the same time, substantial variability in adjusted R^2 values remained within both seasons.

Overall biomechanics model performance

Adjusted R^2 distribution across all biomechanics models

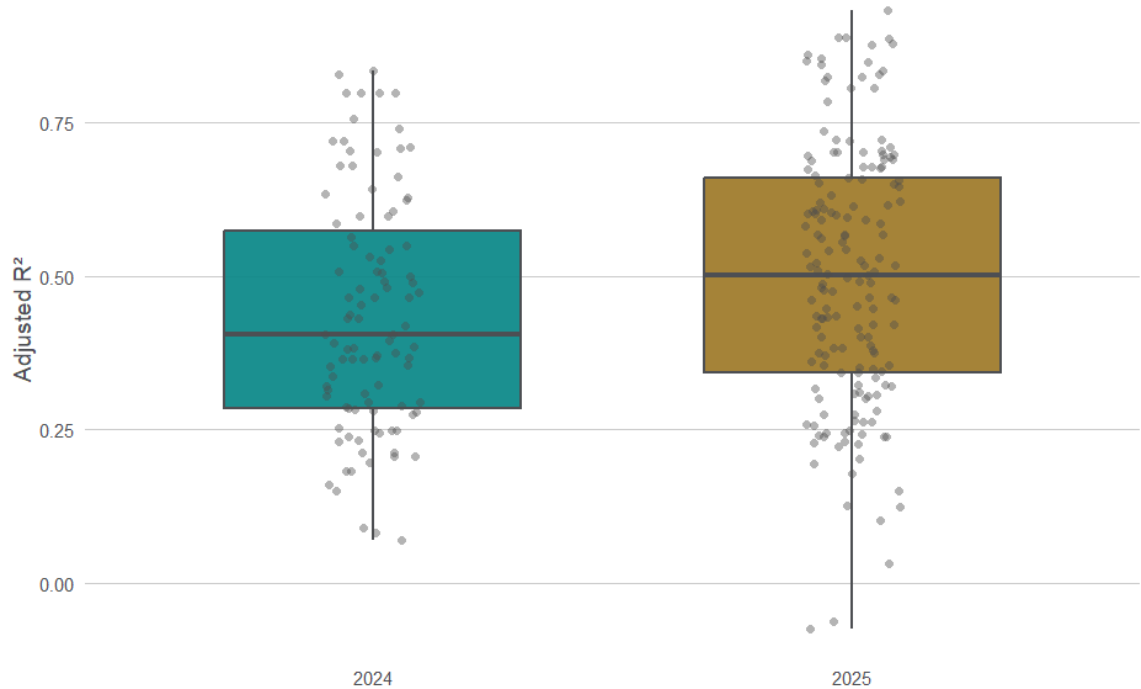


Figure 10: Overall performance of individualized biomechanics models by season. Distribution of adjusted R^2 values across all final individualized biomechanics models in 2024 and 2025. Boxes represent interquartile ranges, center lines indicate medians, and points represent individual pitcher-outcome models.

Outcome-Specific Biomechanics Model Performance

Outcome-specific distributions of individualized model performance are shown in Figure 11. In contrast to the mean summaries presented in the previous section, these distributions illustrate the spread of adjusted R^2 values across individual pitcher-season models within each biomechanics outcome. The distribution of adjusted R^2 values varied across the 11 outcomes, with several variables showing relatively higher and more concentrated adjusted R^2 distributions, whereas others demonstrated broader spread and greater variability across individualized models.

For the elbow and lower-extremity and kinetic-chain outcomes, hip-shoulder separation at foot contact and maximum hip-shoulder separation showed some of the

strongest overall distributions, with many models clustering in the upper portion of the observed adjusted R^2 range. Maximum elbow varus torque also demonstrated comparatively favorable model performance, particularly in 2025, where the distribution appeared shifted upward relative to 2024. In contrast, elbow flexion at ball release, elbow flexion at maximum external rotation, and lead knee flexion at ball release showed more moderate distributions overall, with greater overlap between lower- and higher-performing individualized models.

Among the shoulder-related outcomes, shoulder rotation at maximum external rotation and shoulder abduction/adduction at ball release demonstrated moderate and relatively consistent adjusted R^2 distributions across seasons. Maximum resultant shoulder force and maximum shoulder rotational torque showed greater spread, but also clearer upward shifts in 2025, indicating that stronger-performing individualized models were more common in the later season. Bodyweight-normalized maximum elbow varus torque also demonstrated a generally improved distribution in 2025, with more models concentrated in the moderate-to-higher adjusted R^2 range.

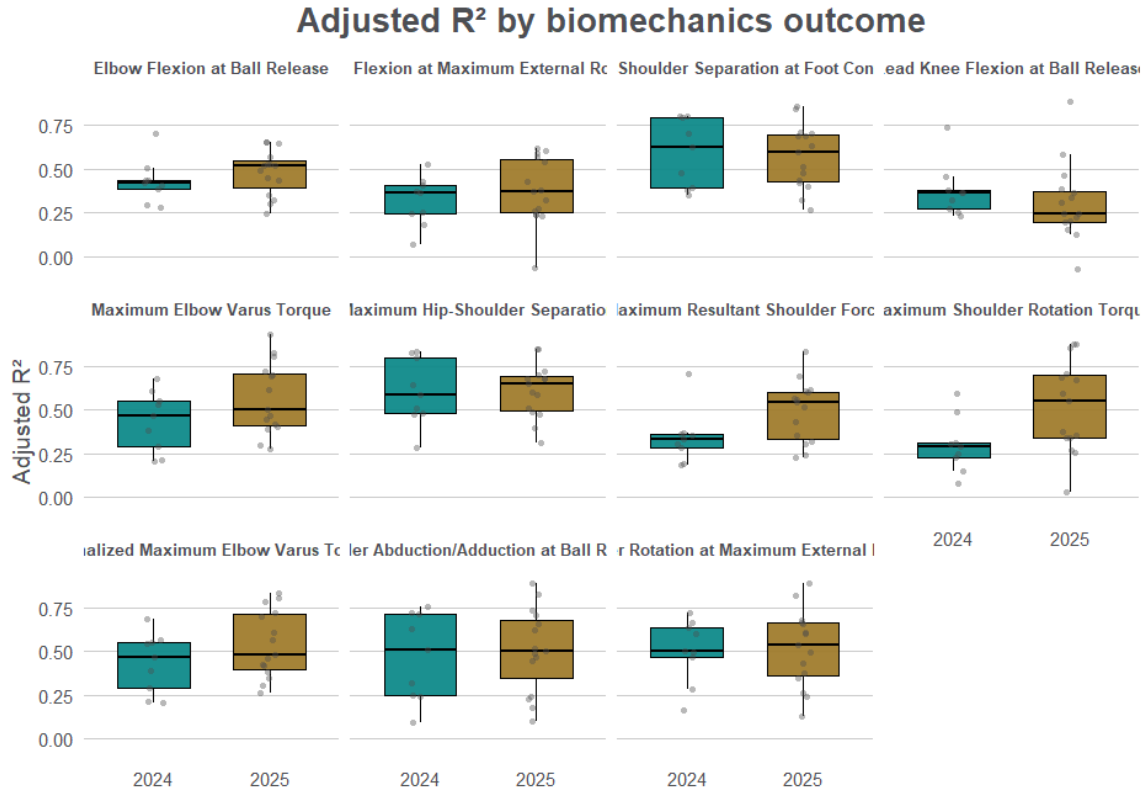


Figure 11: Distribution of adjusted R² values for final individualized biomechanics models across outcomes and seasons. Boxes represent interquartile ranges, center lines indicate medians, and points represent individual pitcher-season models.

Pitcher-Specific Heterogeneity

Pitcher-specific heterogeneity in biomechanics model fit is shown in Figure 12. No single outcome demonstrated uniformly high or low fit across all pitcher-season entities. Instead, the heat map shows distinct pitcher-specific fit profiles across the 11 modeled biomechanics variables. Several signals, including maximum hip-shoulder separation, hip-shoulder separation at foot contact, and maximum elbow varus torque, demonstrated moderate-to-higher adjusted R² values for many pitcher-season entities, although the observed fit remained uneven across the cohort. Other variables, including lead knee flexion at ball release and elbow flexion at maximum external rotation, showed broader

dispersion, with some pitcher-season entities demonstrating stronger fit and others remaining comparatively lower. Overall, the heat map shows that the pattern of adjusted R^2 values differed across both pitchers and biomechanics outcomes.

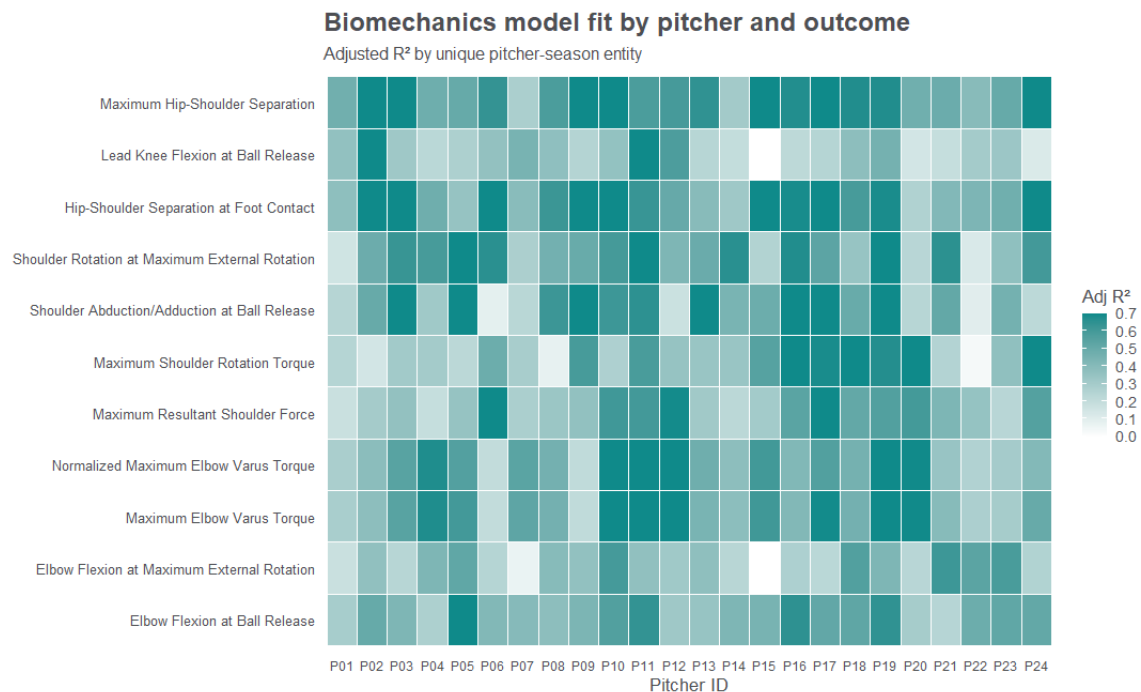


Figure 12: Heat map of adjusted R^2 values for each biomechanics outcome across individualized pitcher-season models. Darker shading indicates higher adjusted R^2 values, reflecting stronger model fit for a given biomechanics signal within a pitcher-season entity.

Residual Autocorrelation, Stability, and Sensitivity Analyses

Residual autocorrelation summaries are presented in Table VII. Across outcomes, mean $AR(1)$ values were generally small in magnitude and often centered near zero. Several biomechanics variables demonstrated predominantly negative mean $AR(1)$ values across pitcher-season entities, including elbow flexion at ball release, elbow flexion at maximum external rotation, lead knee flexion at ball release, maximum resultant shoulder force, and maximum shoulder rotational torque. In contrast, hip-shoulder separation at foot contact and maximum hip-shoulder separation more often demonstrated

positive mean AR(1) values. Maximum hip-shoulder separation showed the highest average mean AR(1) value among the modeled outcomes (0.077), whereas lead knee flexion at ball release showed the lowest (-0.079). Overall, these summaries showed that residual autocorrelation was present for some individualized biomechanics models, although the direction and magnitude differed across outcomes.

Table VII: Residual autocorrelation summaries for final individualized biomechanics models by outcome

Outcomes	Average Mean AR(1)	Median AR(1)	Range of Mean AR(1)	% AR(1) Positive/ Negative
Elbow Flexion at BR	-0.038	-0.003	-0.245 to 0.023	22.2% / 77.8%
Elbow Flexion at MER	-0.064	-0.09	-0.245 to 0.005	11.1% / 88.9%
Hip-Shoulder Separation at FC	0.016	0.027	-0.142 to 0.133	66.7% / 33.3%
Lead Knee Flexion at BR	-0.079	-0.135	-0.158 to 0.059	11.1% / 88.9%
Max Elbow Varus Torque	0.022	-0.002	-0.156 to 0.286	44.4% / 55.6%
Max Hip- Shoulder Separation	0.077	0.127	-0.122 to 0.337	77.8% / 22.2%

Max Resultant Shoulder Force	-0.042	-0.04	-0.119 to 0.069	11.1% / 88.9%
Max Shoulder Rotational Torque	-0.045	-0.057	-0.216 to 0.135	33.3% / 66.7%
Normalized Maximum Elbow Varus Torque	0.022	-0.001	-0.156 to 0.286	44.4% / 55.6%
Shoulder Abduction/ Adduction at BR	-0.019	0.037	-0.199 to 0.092	55.6% / 44.4%
Shoulder Rotation at MER	-0.04	-0.07	-0.14 to 0.047	33.3% / 66.7%

Stability and sensitivity summaries are presented in Table VIII. Curve stabilization analyses showed that the magnitude of prefix-versus-full RMSE differences varied across outcomes. Loading-related outcomes generally demonstrated larger stabilization deltas, with maximum resultant shoulder force (41.8 [20.33, 73.55]), maximum shoulder rotational torque (7.53 [4.02, 14.49]), and maximum elbow varus torque (5.04 [2.62, 7.10]) showing the largest values among the outcomes reported. In contrast, elbow flexion at ball release, elbow flexion at maximum external rotation, hip-shoulder separation at foot contact, and maximum hip-shoulder separation showed smaller stabilization differences, although nonzero changes were still observed across outcomes.

Alternate specification sensitivity analyses demonstrated generally smaller RMSE changes than the stabilization analyses, although the magnitude again differed by

outcome. Maximum resultant shoulder force, maximum shoulder rotational torque, and maximum elbow varus torque showed the largest alternate-specification deltas among the outcomes reported, whereas elbow flexion and lead knee flexion measures showed smaller changes overall. Trial-reduction sensitivity analyses using 30% observation removal also demonstrated outcome-specific differences. Maximum resultant shoulder force, maximum shoulder rotational torque, and maximum elbow varus torque again showed the largest RMSE increases, whereas elbow flexion and hip-shoulder separation measures showed smaller trial-reduction effects. Across the reported outcomes, the magnitude of RMSE change under reduced-data and alternate-specification conditions differed by biomechanics signal, with several loading-related variables showing larger deltas than several positional variable.

Table VIII: Stability and sensitivity summaries for final individualized biomechanics models by outcome

Outcome	Curve stabilization ▲ RMSE (prefix vs full)	Alternate Specification vs Baseline ▲ RMSE	Trial Sensitivity (30% removal) ▲ RMSE
Elbow Flexion at BR	0.99 [0.52, 1.59]	0.19 [0.13, 0.28]	0.21 [0.13, 0.34]
Elbow Flexion at MER	1.52 [0.61, 2.56]	0.23 [0.1, 0.41]	0.3 [0.19, 0.44]
Hip-Shoulder Separation at FC	1.59 [0.78, 3.22]	0.44 [0.14, 0.87]	0.33 [0.22, 0.48]

Lead Knee Flexion at BR	2.29 [1.49, 3.63]	0.24 [0.06, 0.56]	0.51 [0.34, 0.84]
Max Elbow Varus Torque	5.04 [2.62, 7.1]	0.89 [0.42, 2.51]	0.75 [0.49, 1.18]
Max Hip-Shoulder Separation	1.36 [0.7, 3.15]	0.51 [0.22, 0.86]	0.29 [0.2, 0.41]
Max Resultant Shoulder Force	41.8 [20.33, 73.55]	3.81 [2.92, 10.25]	7.41 [4.65, 11.05]
Max Shoulder Rotational Torque	7.53 [4.02, 14.49]	1.54 [0.51, 3.09]	1.61 [1.03, 2.79]
Normalized Maximum Elbow Varus Torque Shoulder Abduction/ Adduction at BR	0.01 [0, 0.01]	0 [0, 0]	0 [0, 0]
Shoulder Rotation at MER	1.13 [0.51, 1.88]	0.23 [0.12, 0.42]	0.21 [0.13, 0.31]

Bootstrap Uncertainty Summaries

Bootstrap uncertainty summaries across pitcher-specific N-of-1 biomechanics models are presented in Table IX. Confidence interval width differed substantially across outcomes. Loading-related variables demonstrated some of the widest bootstrap intervals, particularly maximum resultant shoulder force, maximum shoulder rotational torque, and maximum elbow varus torque, whereas bodyweight-normalized maximum elbow varus

torque demonstrated the narrowest intervals. Positional and separation-related measures showed intermediate uncertainty, although lead knee flexion at ball release demonstrated comparatively broader intervals than several of the other kinematic variables. Together, these summaries show that bootstrap-derived prediction uncertainty was outcome dependent across pitcher-specific N-of-1 biomechanics models.

Table IX: Bootstrap uncertainty summaries across pitcher-specific N-of-1 biomechanics models by outcome.

Outcome	Mean CI width	Median CI width [IQR]
Elbow Flexion at BR	3.92	2.96 [1.83]
Elbow Flexion at MER	4.82	4.42 [2.73]
Hip-Shoulder Separation at FC	6.36	5.65 [3.22]
Lead Knee Flexion at BR	10.86	7.21 [5.65]
Max Elbow Varus Torque	14.73	10.70 [9.40]
Max Hip-Shoulder Separation	5.91	5.56 [2.62]
Max Resultant Shoulder Force	143.40	113.48 [87.74]
Max Shoulder Rotational Torque	30.72	20.34 [14.48]

Normalized Maximum Elbow Varus Torque	0.02	0.01 [0.01]
Shoulder Abduction/ Adduction at BR	4.05	3.19 [2.27]
Shoulder Rotation at MER	3.91	3.11 [2.00]

Discussion

The purpose of this chapter was to examine individualized associations between rolling pitch count and fastball biomechanics in collegiate baseball pitchers using an N-of-1 modeling framework. More specifically, this chapter sought to determine whether workload-related responses could be identified through biomechanical outcomes reflecting joint loading, segment positioning, and kinetic-chain organization, and whether those associations differed across pitchers and across biomechanical signals. Overall, the findings support that rolling pitch count was associated with fastball biomechanics in an individualized manner, with model performance varying across both outcomes and pitcher-season entities. This framing is consistent with the intended Aim 2 emphasis on individualized biomechanical associations and the idea that biomechanics may capture workload responses not fully reflected in ball-flight outcomes alone.

Several broad findings emerged from these analyses. First, individualized model fit was generally moderate, with many outcomes demonstrating adjusted R^2 values in a range that suggested meaningful, but not uniform, within-pitcher associations. Second, model performance differed across biomechanics outcomes. Signals such as hip-shoulder separation at foot contact, maximum hip-shoulder separation, and maximum elbow varus torque generally demonstrated stronger overall model performance, whereas other variables showed more modest fit or broader spread across pitcher-season models. Third,

model fit was not consistent across pitchers. Instead, the results demonstrated distinct pitcher-specific fit profiles across the 11 modeled biomechanics variables, with some pitcher-season entities showing stronger fit in trunk or separation-related measures and others showing stronger fit in elbow or shoulder-related signals.

These findings are important because they support the broader concept that workload-related responses in pitchers may emerge in more than one way. In some athletes, those responses may be reflected more clearly through changes in ball-flight or performance outcomes, whereas in others they may be reflected more strongly through biomechanics. In that sense, this chapter suggests that biomechanics represent another major domain in which pitcher response to workload may be expressed. This is particularly relevant in baseball, where an athlete may preserve visible performance while altering how force is generated, transferred, or distributed throughout the pitching motion.

The results also reinforce the rationale for using an individualized modeling framework in this setting. No single biomechanics variable consistently demonstrated the strongest fit across all pitcher-season entities, and no single pattern characterized the cohort as a whole. In addition, the number of biomechanics outcomes showing moderate-to-strong model fit differed across pitcher-season entities. Some pitchers demonstrated relatively stronger fit across several signals, whereas others showed stronger fit for only a limited number of outcomes. Together, these results indicate that the observed associations differed not only in which signals appeared most responsive, but also in how broadly those associations were expressed within a given pitcher-season. This further supports that workload-related biomechanical responses were expressed in a pitcher-specific manner rather than through a single common pattern.

Some season-to-season differences in model fit were also observed and should be interpreted cautiously. These differences may reflect variation in sample size, pitcher

usage patterns, and data-collection context across years rather than a single underlying change in model behavior. In addition, because biomechanical variables differ in their inherent variability and measurement characteristics, some year-to-year differences in RMSE or explained variance may fall within expected ranges of signal and measurement fluctuation rather than indicating a practically meaningful shift. Pitcher-seasons were modeled independently, and only a small number of pitchers contributed data in both seasons; prior-season data were not used to inform model fitting in a subsequent season. As a result, repeated pitchers across years contributed separate season-specific models rather than a pooled longitudinal model across seasons.

The robustness analyses further clarified that not all biomechanics outcomes behaved similarly from a modeling standpoint. Residual autocorrelation was generally small in magnitude, although the direction and magnitude differed across outcomes. Stability, alternate-specification sensitivity, and trial-reduction analyses showed that some variables were more sensitive to reduced data availability or altered model structure than others. In particular, several loading-related variables demonstrated larger RMSE deltas under these conditions than several positional variables. Bootstrap uncertainty summaries also showed meaningful variation across outcomes, with some signals demonstrating relatively narrow confidence interval widths and others showing broader uncertainty ranges. Taken together, these findings suggest that individualized biomechanics modeling is feasible and informative, but that both the strength and reliability of model performance depend in part on the specific biomechanical signal being studied.

Limitations

Several limitations should be considered. First, biomechanics data collection depended on specialized motion-capture infrastructure, resulting in a more constrained dataset than the broader ball-metric analyses. In particular, game biomechanics data were

limited to home-game or local capture contexts, reducing the number of available pitches and appearances and limiting the utility of several travel-related variables. Second, not all candidate covariates were available equally across pitcher-seasons. HRV data were available only for pitchers who elected to wear WHOOP devices, and several travel-related variables had structurally limited availability. Third, repeated data density varied across individualized series, which may have influenced model stability and uncertainty for some pitcher-outcome combinations. Finally, the study was conducted within a single collegiate baseball program and used an observational design, which limits both generalizability and causal interpretation.

Applied Implications and Future Directions

These findings suggest that biomechanics may provide pitcher-specific information not fully captured by ball metrics alone. For some pitchers, loading or movement-pattern variables may show clearer workload-related associations than outward performance outcomes, supporting a more individualized monitoring approach rather than reliance on a single universal marker. The results also suggest that some biomechanics signals may be more practically informative than others, which may help guide future prioritization of key performance indicators in applied baseball settings.

Future work should evaluate these findings in larger and more diverse baseball populations, expand biomechanics capture across broader competitive settings, and examine how individualized biomechanics models can be integrated with ball-metric models and practical monitoring tools. Such an approach may be particularly valuable for pitchers whose workload response is expressed more clearly through biomechanics than through visible performance change.

Conclusion

In conclusion, rolling pitch count demonstrated individualized associations with fastball biomechanics in collegiate baseball pitchers, with the strength and distribution of those associations varying across both biomechanic outcomes and pitcher-season entities. Several signals, particularly hip-shoulder separation and elbow loading variables, demonstrated comparatively stronger model performance overall, although no single biomechanical variable or pitcher-specific pattern was uniform across the cohort. Some pitcher-season entities showed broader sets of moderate-to-strong models across multiple biomechanics signals, whereas others demonstrated stronger fit for only a smaller number of outcomes, further emphasizing the individualized nature of workload-related biomechanical response. Taken together, these findings suggest that biomechanics represent an important domain in which pitcher response to workload may be expressed and that these responses may not emerge uniformly across athletes or signals. As such, individualized biomechanical modeling may provide a useful complement to ball-metric monitoring within a broader workload assessment framework, particularly when the goal is to better characterize pitcher-specific response patterns over time. However, further work is needed to determine what magnitude or pattern of biomechanical change is practically meaningful for informing training load and recovery decisions.

References

1. Agresta C, Freehill MT, Nakamura B, Guadagnino S, Cain SM. Using Sensors for Player Development: Assessing Biomechanical Factors Related to Pitch Command and Velocity. *Sensors (Basel)*. 2022;22(21):8488. doi:10.3390/s22218488
2. Bullock GS, Menon G, Nicholson K, Butler RJ, Arden NK, Filbay SR. Baseball pitching biomechanics in relation to pain, injury, and surgery: A systematic review. *J Sci Med Sport*. 2021;24(1):13-20. doi:10.1016/j.jsams.2020.06.015
3. Dong M, Li M, Qu Q, Kim Y, Kim S. Arm slot angles affect elbow and shoulder joint torque in elite college pitchers. *Sports Biomechanics*. Published online January 2, 2025:1-18. doi:10.1080/14763141.2024.2431927
4. Nicholson KF, Mylott JA, Waterman BR, Bullock GS. Pitching Biomechanics Normative Values and Kinetic Differences by Competition Level. *J Surg Orthop Adv*. 2022;31(3):177-180.
5. Trasolini NA, Nicholson KF, Mylott J, Bullock GS, Hulburt TC, Waterman BR. Biomechanical Analysis of the Throwing Athlete and Its Impact on Return to Sport. *Arthrosc Sports Med Rehabil*. 2022;4(1):e83-e91. doi:10.1016/j.asmr.2021.09.027
6. Cardinale T, Salle TL, Aguinaldo A. CONCURRENT VALIDITY AND RELIABILITY OF IN-LAB MARKERLESS MOTION CAPTURE IN ESTIMATING JOINT KINEMATICS IN BASEBALL PITCHING. *ISBS Proceedings Archive*. 2024;42(1):144.
7. Giordano KA, Schmitt A, Nebel A, Yanagita Y, Oliver GD. Normative In-Game Data for Collegiate Baseball Pitchers Using Markerless Tracking Technology. *Orthop J Sports Med*. 2024;12(10):23259671241274137. doi:10.1177/23259671241274137
8. Nicholson KF, Collins GS, Waterman BR, Bullock GS. Machine Learning and Statistical Prediction of Pitching Arm Kinetics. *Am J Sports Med*. 2022;50(1):238-247. doi:10.1177/03635465211054506

9. Dowling B, Manzi JE, Raab G, Coladonato C, Dines JS, Fleisig GS. The relationship among lead knee extension, fastball velocity and elbow torque in professional baseball pitchers. *Sports Biomech.* 2024;23(12):2664-2674. doi:10.1080/14763141.2022.2050801
10. Dowling B, McNally MP, Chaudhari AMW, Oñate JA. A Review of Workload-Monitoring Considerations for Baseball Pitchers. *J Athl Train.* 2020;55(9):911-917. doi:10.4085/1062-6050-0511-19
11. Bakshi NK, Inclan PM, Kirsch JM, Bedi A, Agresta C, Freehill MT. Current Workload Recommendations in Baseball Pitchers: A Systematic Review. *Am J Sports Med.* 2020;48(1):229-241. doi:10.1177/0363546519831010
12. Reinold MM, Dowling B, Fleisig GS, et al. An Interval Throwing Program for Baseball Pitchers Based upon Workload Data. *IJSPT.* 2024;19(3):326-336. doi:10.26603/001c.94146
13. Shanley E, Bailey L, Sandago MP, Pinkerton A, Singleton SB, Thigpen CA. The use of a pitch count estimator to calculate exposure in collegiate baseball pitchers. *Phys Ther Sport.* 2015;16(4):344-348. doi:10.1016/j.ptsp.2015.02.002
14. Birfer R, Sonne MW, Holmes MW. Manifestations of muscle fatigue in baseball pitchers: a systematic review. *PeerJ.* 2019;7:e7390. doi:10.7717/peerj.7390
15. Grantham WJ, Byram IR, Meadows MC, Ahmad CS. The Impact of Fatigue on the Kinematics of Collegiate Baseball Pitchers. *Orthop J Sports Med.* 2014;2(6):2325967114537032. doi:10.1177/2325967114537032
16. Okoroha KR, Meldau JE, Lizzio VA, et al. Effect of Fatigue on Medial Elbow Torque in Baseball Pitchers: A Simulated Game Analysis. *Am J Sports Med.* 2018;46(10):2509-2513. doi:10.1177/0363546518782451

17. Vieira R, McDonald S, Araújo-Soares V, Sniehotta FF, Henderson R. Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies. *Health Psychology Review*. 2017;11(3):222-234. doi:10.1080/17437199.2017.1343680
18. Daza EJ. Causal Analysis of Self-tracked Time Series Data Using a Counterfactual Framework for N-of-1 Trials*. *Methods Inf Med*. 2018;57(Suppl 1):e10-e21. doi:10.3414/ME16-02-0044
19. Kwasnicka D, Naughton F. N-of-1 methods: A practical guide to exploring trajectories of behaviour change and designing precision behaviour change interventions. *Psychology of Sport and Exercise*. 2020;47:101570. doi:10.1016/j.psychsport.2019.101570
20. Daza EJ. Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Tracked Health Data. *arXiv*. Preprint posted online January 22, 2019:arXiv:1901.03423. doi:10.48550/arXiv.1901.03423
21. Jones CM, Griffiths PC, Mellalieu SD. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Med*. 2017;47(5):943-974. doi:10.1007/s40279-016-0619-5
22. McDonald S, Vieira R, Johnston DW. Analysing N-of-1 observational data in health psychology and behavioural medicine: a 10-step SPSS tutorial for beginners. *Health Psychol Behav Med*. 8(1):32-54. doi:10.1080/21642850.2019.1711096
23. Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. Published online May 14, 2015. doi:10.1136/bmj.h1738
24. Yang L, Armijo-Olivo S, Gross DP. Single-Case Experimental Design in Rehabilitation: Basic Concepts, Advantages, and Challenges. *Am J Phys Med Rehabil*. 2023;102(5):454-460. doi:10.1097/PHM.0000000000002215

25. Bailey C. Longitudinal Monitoring of Athletes: Statistical Issues and Best Practices. *Journal of Science in Sport and Exercise*. 2019;1(3):217-227. doi:10.1007/s42978-019-00042-4
26. Araujo A, Julious S, Senn S. Understanding Variation in Sets of N-of-1 Trials. *PLoS One*. 2016;11(12):e0167167. doi:10.1371/journal.pone.0167167
27. International Olympic Committee Injury and Illness Epidemiology Consensus Group, Bahr R, Clarsen B, et al. International Olympic Committee Consensus Statement: Methods for Recording and Reporting of Epidemiological Data on Injury and Illness in Sports 2020 (Including the STROBE Extension for Sports Injury and Illness Surveillance (STROBE-SIIS)). *Orthop J Sports Med*. 2020;8(2):2325967120902908. doi:10.1177/2325967120902908
28. Wilson DT, Hooper R, Brown J, Farrin AJ, Walwyn RE. Efficient and flexible simulation-based sample size determination for clinical trials with multiple design parameters. *Stat Methods Med Res*. 2021;30(3):799-815. doi:10.1177/0962280220975790
29. Senn S. Sample size considerations for n-of-1 trials. *Stat Methods Med Res*. 2019;28(2):372-383. doi:10.1177/0962280217726801
30. Addleman JS, Lackey NS, DeBlauw JA, Hajduczuk AG. Heart Rate Variability Applications in Strength and Conditioning: A Narrative Review. *J Funct Morphol Kinesiol*. 2024;9(2):93. doi:10.3390/jfmk9020093
31. DONG JG. The role of heart rate variability in sports physiology. *Exp Ther Med*. 2016;11(5):1531-1536. doi:10.3892/etm.2016.3104
32. Diffendaffer AZ, Bagwell MS, Fleisig GS, et al. The Clinician's Guide to Baseball Pitching Biomechanics. *Sports Health: A Multidisciplinary Approach*. 2023;15(2):274-281. doi:10.1177/19417381221078537

33. Aguinaldo AL, Chambers H. Correlation of throwing mechanics with elbow valgus load in adult baseball pitchers. *Am J Sports Med.* 2009;37(10):2043-2048. doi:10.1177/0363546509336721
34. Matsuda R, Hirano Y, Umakoshi J, Kimura A. Energy flows with intentional changes in leg movements during baseball pitching. *Front Sports Act Living.* 2025;7:1534596. doi:10.3389/fspor.2025.1534596
35. Matsuo T, Fleisig GS. Influence of shoulder abduction and lateral trunk tilt on peak elbow varus torque for college baseball pitchers during simulated pitching. *J Appl Biomech.* 2006;22(2):93-102. doi:10.1123/jab.22.2.93
36. Manzi JE, Estrada JA, Quan T, Ruzbarsky JJ, Ciccotti MC, Dines JS. The influence of excessive ipsilateral trunk tilt on upper-extremity throwing mechanics: a newly characterized parameter for biomechanical evaluation in high school and professional pitchers. *Journal of Shoulder and Elbow Surgery.* 2022;31(9):1909-1921. doi:10.1016/j.jse.2022.01.153
37. Cross JA, Higgins AW, Dziuk CC, Harris GF, Raasch WG. Relationships Among Shoulder Rotational Strength, Range of Motion, Pitching Kinetics, and Pitch Velocity in Collegiate Baseball Pitchers. *Journal of Strength and Conditioning Research.* Published online March 23, 2022. doi:10.1519/JSC.0000000000004246
38. Erickson BJ, Chalmers PN, D'Angelo J, Ma K, Romeo AA. Performance and return to sport following rotator cuff surgery in professional baseball players. *J Shoulder Elbow Surg.* 2019;28(12):2326-2333. doi:10.1016/j.jse.2019.01.029
39. Downs Talmage J, Gilliam J, Chardhari A, Oliver GD. Differences in Lower Extremity Kinematics Between Collegiate and Youth Softball Pitchers. *Orthopaedic Journal of Sports Medicine.* 2021;9(11):23259671211052022. doi:10.1177/23259671211052022

40. Dowling B, Knapik DM, Luera MJ, Garrigues GE, Nicholson GP, Verma NN. Influence of Pelvic Rotation on Lower Extremity Kinematics, Elbow Varus Torque, and Ball Velocity in Professional Baseball Pitchers. *Orthop J Sports Med.* 2022;10(11):23259671221130340. doi:10.1177/23259671221130340
41. Luera MJ, Dowling B, Muddle TWD, Jenkins NDM. Differences in Rotational Kinetics and Kinematics for Professional Baseball Pitchers With Higher Versus Lower Pitch Velocities. *J Appl Biomech.* 2020;36(2):68-75. doi:10.1123/jab.2019-0235
42. Zaremski JL, Pazik M, Vasilopoulos T, Horodyski M. Workload Risk Factors for Pitching-Related Injuries in High School Baseball Pitchers. *Am J Sports Med.* 2024;52(7):1685-1691. doi:10.1177/03635465241246559
43. Nicholas Tierney, Dianne Cook. Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. *Journal of Statistical Software.* 2023;105(7):1-31. doi:10.18637/jss.v105.i07
44. Van Calster B, van Smeden M, De Cock B, Steyerberg EW. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. *Stat Methods Med Res.* 2020;29(11):3166-3178. doi:10.1177/0962280220921415
45. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software.* 2010;33:1-22. doi:10.18637/jss.v033.i01
46. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Published online October 4, 2000:1.9-4. doi:10.32614/CRAN.package.mgcv
47. Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA. A linear exponent AR(1) family of correlation structures. *Stat Med.* 2010;29(17):1825-1838. doi:10.1002/sim.3928

48. Raykov T, DiStefano C. Evaluating Change in Adjusted R-Square and R-Square Indices: A Latent Variable Method Application. *Educ Psychol Meas*. Published online April 11, 2025:00131644251329178. doi:10.1177/00131644251329178
49. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014;7(3):1247-1250. doi:10.5194/gmd-7-1247-2014
50. Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*. 2019;11(3):e1460. doi:10.1002/wics.1460

CHAPTER IV: Comparative Evaluation of Group-Based and Individualized N-of-1 Models for Pitch Count Associations in Baseball Pitchers

Garrett Fernandez^{1,2}, Garrett Bullock^{1,3}, Patrick Ward⁴, Joel Stitzel^{1,2}, Ryan McGinnis¹,
Tanner Filben⁵, Kristen Nicholson^{1,3}

¹Department of Biomedical Engineering, Wake Forest School of Medicine, Winston-Salem, NC, ²Virginia Tech-Wake Forest School of Biomedical Engineering and Sciences, Winston-Salem, NC, ³Department of Orthopedic Surgery and Rehabilitation, Wake Forest School of Medicine, Winston Salem, NC, ⁴Seattle Seahawks, Seattle, WA, USA, ⁵Elemance, LLC, Winston-Salem, NC

Abstract

The purpose of this chapter was to compare group-based and individualized N-of-1 models for quantifying rolling pitch count associations in collegiate baseball pitchers. Comparative evaluation focused on fastball velocity as the primary performance outcome and four representative biomechanics outcomes as the secondary comparison domain: maximum elbow varus torque, shoulder abduction angle at ball release, lead knee flexion angle at ball release, and maximum hip-shoulder separation. The velocity comparison set included 35 pitcher-seasons across the 2023, 2024, and 2025 seasons, and the primary biomechanics comparison set included 96 pitcher-outcome comparisons across the 2024 and 2025 seasons. Model performance was evaluated using root mean square error and adjusted R^2 , with emphasis placed on within-pitcher explanatory performance rather than aggregate fit alone. For velocity, individualized N-of-1 models consistently outperformed the group-based framework across all seasons, with overall RMSE of 0.87 versus 2.43 and adjusted R^2 of 0.53 versus -5.84. The same pattern was observed by season, with N-of-1 velocity RMSE ranging from 0.80 to 0.96 compared with 2.10 to 2.86 for the group model, and adjusted R^2 ranging from 0.51 to 0.57 compared with -2.23 to -8.74. Across biomechanics, individualized models also demonstrated lower RMSE and higher adjusted R^2 than the group-based framework for all four selected outcomes: maximum elbow varus torque (6.16 vs 18.94; 0.52 vs -3.93), shoulder abduction angle at ball release (1.82 vs 6.21; 0.51 vs -6.92), lead knee flexion angle at ball release (5.00 vs 14.53; 0.33 vs -7.18), and maximum hip-shoulder separation (2.17 vs 7.72; 0.61 vs -4.94). Overall, individualized N-of-1 models outperformed group-based models across both performance and biomechanics domains, indicating that pitcher-specific workload-response patterns were more successful at capturing individualized variance and modeling than by population-average approaches.

Introduction

Chapters 2 and 3 of this dissertation demonstrated that rolling pitch count associations could be modeled at the pitcher-season level using an individualized N-of-1 framework. In Chapter 2, individualized models were used to characterize associations between rolling pitch count and fastball performance metrics, while Chapter 3 extended this framework to biomechanics outcomes reflecting joint loading, segment positioning, and kinetic-chain organization. Across both chapters, the results supported that workload-related patterns were not uniform across pitchers and were not always expressed in the same outcome domain. Instead, the observed associations varied across pitchers, across seasons, and across performance and biomechanics signals, reinforcing the broader dissertation premise that workload response is heterogeneous rather than interchangeable across athletes.

Although those findings support the feasibility and value of individualized modeling, they also raise an important follow-up question. Specifically, if rolling pitch count associations can be characterized within pitchers, to what extent does that individualized framework provide information that differs from, or improves upon, more traditional group-based approaches? This is a particularly relevant question in sports science, where mixed-effects and other population-level models remain common because they provide a practical way to estimate average relationships across repeated observations. Such models are useful for describing broad trends across athletes, but their primary inferential emphasis is often placed on group-level effects, even though individual deviations can be estimated through random effects. However, the extent to which individual response patterns are explicitly characterized and interpreted is often limited. In settings where meaningful heterogeneity is present, that distinction may have important implications for how workload-response patterns are interpreted.

This issue is especially relevant in baseball because the applied decision-making target is not the average pitcher, but the specific athlete. Workload management, recovery planning, and performance monitoring are all carried out at the individual level. A model that summarizes how pitch count tends to relate to outcomes across a roster may therefore be useful for general inference yet can still fail to adequately represent the response pattern of a given pitcher. This concern applies not only to fastball performance metrics such as velocity, spin rate, and vertical break, but also to biomechanical outcomes, where prior chapter results suggested that workload-related responses may emerge through multiple mechanical pathways and may not be reflected uniformly across athletes.

From this standpoint, group-based and individualized models should not be viewed as interchangeable versions of the same analytic objective. Rather, they represent different inferential perspectives. Group-based models are designed to estimate population-level associations across pitchers while accounting for clustered observations, whereas individualized N-of-1 models are designed to estimate within-pitcher associations using repeated measures from a single athlete. A key distinction between these approaches lies in the degree of pooling, with mixed-effects models relying on partial pooling across athletes and N-of-1 models treating each athlete as a fully independent analytic unit. The practical significance of that distinction depends on whether the two approaches yield similar conclusions or whether individualized models capture pitcher-specific structure that is attenuated or obscured in the group average.

Accordingly, after establishing individualized workload-related associations in the prior chapters, the next logical step is to directly compare those individualized models with traditional population-level alternatives. This comparison is necessary not only to evaluate relative explanatory performance, but also to clarify what may be gained or lost when pitcher-specific workload-response patterns are summarized through a group framework.

If both approaches perform similarly, then the practical advantage of individualized modeling may be limited, particularly given the need to fit and interpret separate models for each athlete rather than a single population-level model. If they differ meaningfully, however, that difference would support the use of individualized approaches when the goal is to understand or monitor workload response within a pitcher rather than across a population.

Therefore, the purpose of this chapter was to compare group-based mixed-effects models and individualized N-of-1 generalized additive mixed models for quantifying rolling pitch count associations in collegiate baseball pitchers. Comparisons focused on fastball velocity as the primary performance outcome and on pitching biomechanics as a broader secondary outcome domain. By evaluating model performance across these two domains, this chapter sought to determine whether individualized models provided superior within-pitcher explanatory value relative to traditional group-based approaches and to clarify the extent to which population-average models may obscure meaningful pitcher-specific workload-response patterns.

Methods

Study Design

The velocity comparison set included 35 pitcher-seasons across the 2023, 2024, and 2025 seasons. The full biomechanics comparison outputs included 264 pitcher-outcome models across 2024 and 2025. Within the primary biomechanics comparison set emphasized in this chapter, 96 pitcher-outcome comparisons were included across maximum elbow varus torque, shoulder abduction angle at ball release, lead knee flexion angle at ball release, and maximum hip-shoulder separation. This consisted of 9 pitcher comparisons per outcome in 2024 and 15 per outcome in 2025.

Data Sources and Outcome Selection

Performance outcomes for this chapter were drawn from the Chapter 2 analytic framework and included fastball data from the 2023, 2024, and 2025 collegiate baseball seasons. Velocity was retained as the primary performance outcome for direct comparison because it provided the clearest applied basis for evaluating differences between group-based and individualized models. Spin rate and vertical break were also modeled in Chapter 2 but were not emphasized in the present chapter in order to maintain focus on the clearest performance comparison and avoid redundancy with prior results.

Biomechanics outcomes for this chapter were drawn from the Chapter 3 analytic framework and included fastball biomechanics data from the 2024 and 2025 collegiate baseball seasons. For the primary biomechanics comparison, four outcomes were retained: maximum elbow varus torque, shoulder abduction angle at ball release, lead knee flexion angle at ball release, and maximum hip-shoulder separation. These outcomes were selected to reflect diversity across major mechanical domains of the pitching motion, including upper-extremity loading, upper-extremity positioning, lower-extremity mechanics, and kinetic-chain organization, rather than being chosen solely on the basis of the strongest comparative model results. This approach allowed model comparison across distinct biomechanical domains while avoiding unnecessary redundancy with the broader outcome coverage presented in Chapter 3.

Because dataset construction, variable definitions, and preprocessing procedures were described previously, only details relevant to model comparison are presented in this chapter.

Analytic Unit and Data Structure

For individualized analyses, the primary analytic unit was the pitcher-season. Within each pitcher-season, repeated pitch-level observations were used to estimate within-pitcher associations between rolling pitch count and the outcome of interest over

time. These observations were longitudinally structured, with pitches ordered within appearances to preserve the sequential nature of each throwing session. This framework was retained from Chapters 2 and 3 in order to preserve season-specific workload context and allow associations to vary across both pitchers and seasons.

For group-based analyses, the same pitch-level observations were modeled across pitchers within season and, where applicable, across pooled data. In this framework, repeated observations remained nested within pitchers, but the inferential target shifted from the individual pitcher to the broader population-level average. Accordingly, the individualized and group-based models were fit to the same general longitudinal data structure but differed in the level at which workload-response associations were estimated and interpreted. Because both approaches used the same underlying pitch-level series, model comparisons reflected differences in analytic framework rather than differences in data source.

Model Development

Individualized N-of-1 Models

Individualized models used in this chapter were the final pitcher-season specific models developed in Chapters 2 and 3 and were carried forward without redevelopment for the present comparison. Briefly, model development in the prior chapters followed a two-stage process consisting of predictor screening and final model fitting. LASSO regression with cross-validated Gaussian glmnet models was used to screen candidate predictors for each pitcher-season and outcome, while forcing rolling pitch count into all models by assigning it a zero penalty. Final pitcher-season specific models were then fit using generalized additive mixed models implemented with `mgcv::gamm`, with rolling pitch count modeled as a smooth term using `s(PitchCount, k = 4)`. Predictors retained from screening were incorporated as linear terms unless evaluated as part of a candidate

interaction, and no more than one interaction term was retained in a final model. Candidate models were evaluated with and without an AR(1) residual correlation structure specified within appearance, and a random intercept for appearance was included to account for clustering of pitches within appearances. Final individualized model selection was based primarily on Akaike information criterion, with consideration of model fit and interpretability.

Group-Based Models

Group-based models were fit across pitchers using the same longitudinal pitch-level outcome domains examined in the individualized analyses. In contrast to the N-of-1 framework, which estimated separate pitcher-season specific associations, the group framework estimated workload-response patterns using combined pitch-level observations within the relevant comparison dataset. For velocity, group-model outputs from the established performance comparison pipeline were used, including season-specific summaries and pooled summaries that were retained for direct comparison with the individualized velocity models.

For the biomechanics group models, two candidate structures were evaluated for each selected outcome. The first was an intercept-only model fit using `mgcv::gam` with the form $y \sim s(\text{PitchCount}, k = 4)$, representing a single shared smooth association between pitch count and the outcome across pitchers. The second was a hierarchical model fit using `mgcv::gamm` with the same smooth pitch-count term and pitcher-specific random intercepts and random slopes for pitch count specified as `random = list(Pitcher = ~ 1 + PitchCount)`. This hierarchical specification allowed pitchers to vary in baseline level and linear pitch-count trend while retaining a common higher-level modeling structure. Both candidate models were fit using restricted maximum likelihood. For each selected outcome, two group-based structures were evaluated: an intercept-only model

representing a shared smooth association across pitchers and a hierarchical model allowing pitcher-specific random intercepts and random slopes for pitch count.

Across both modeling frameworks, the same general predictor sets were used within each outcome domain. Rolling pitch count was retained as the primary exposure in all models, and candidate covariates were drawn from the same broader workload context previously described in Chapters 2 and 3, including recent workload-history variables, recovery-related measures, game context, internal load when available, and selected travel-related variables. Accordingly, comparisons between individualized and group-based models were intended to reflect differences in analytic framework rather than differences in the underlying predictor domains considered for model development.

These models were used to characterize workload-response patterns at the group level under both simpler shared-curve and hierarchical random-effects structures. Final group-model summaries and candidate-model outputs were then carried forward for direct comparison with the finalized individualized models in the present chapter.

Statistical Analysis for Model Comparison

Comparative evaluation focused on differences in explanatory performance between finalized individualized N-of-1 models and corresponding group-based models within the same outcome domains. For both velocity and the selected biomechanics outcomes, comparisons were based on the final individualized models and the corresponding group-model outputs derived from the established comparison pipeline.

Model performance was evaluated primarily using root mean square error (RMSE) and adjusted coefficient of determination (adjusted R^2). RMSE was used to quantify prediction error in the original scale of each outcome, whereas adjusted R^2 was used to summarize explanatory performance while accounting for model complexity. Because the primary objective of this chapter was to compare how well each framework represented

workload-response patterns within pitchers, emphasis was placed on pitcher-level model performance, including within-pitcher adjusted R^2 , rather than only on aggregate fit across the full sample. Supplemental metrics such as mean squared error (MSE) and unadjusted R^2 were retained in underlying comparison outputs but were not emphasized in the main chapter summaries.

Comparative summaries were then examined across pitcher-level, season-level, and overall outputs as appropriate for the available comparison files. These summaries were used to determine whether individualized models provided improved within-pitcher explanatory performance relative to group-based approaches across both performance and biomechanics outcomes.

All data preprocessing, model development, comparative model evaluation, sensitivity analyses, and visualization were performed in R version 4.5.1 (R Core Team, 2025). Key packages used in the Aim 3 comparison pipeline included dplyr 1.1.4, readxl 1.4.5, readr 2.1.5, ggplot2 3.5.2, glmnet 4.1.9, mgcv 1.9.3, nlme 3.1.168, purrr 1.1.0, stringr 1.5.1, tidyr 1.3.1, tibble 3.3.0, patchwork 1.3.1, and scales 1.4.0.

Results

Comparison Set Overview

The velocity comparison set included 35 pitcher-seasons across the 2023, 2024, and 2025 seasons. The full biomechanics comparison outputs included 264 pitcher-outcome models across 2024 and 2025. Within the primary biomechanics comparison set emphasized in this chapter, 96 pitcher-outcome comparisons were included across maximum elbow varus torque, shoulder abduction angle at ball release, lead knee flexion angle at ball release, and maximum hip-shoulder separation. This consisted of 9 pitcher comparisons per outcome in 2024 and 15 per outcome in 2025.

Velocity Model Comparison

Velocity model performance differed consistently between the individualized N-of-1 and group-based frameworks across all seasons. As shown in Table X, N-of-1 models had lower RMSE and higher within-pitcher adjusted R^2 than the group-based model in 2023, 2024, and 2025. The same pattern was present in the overall summary, with the group-based model showing negative adjusted R^2 values in each season. Negative adjusted R^2 values for the group-based model indicate that, when evaluated at the pitcher level, the population-level model explained within-pitcher variation poorly relative to a baseline model. This reflects a mismatch between the group-level modeling target and pitcher-specific response patterns rather than a model-fitting error.

Table X: Comparison of velocity model performance for individualized N-of-1 and group-based models.

Outcome	Season	N-of-1 RMSE	Group Model RMSE	N-of-1 Adj. R^2	Group Model Adj. R^2
Velocity	All	0.87	2.43	0.53	-5.84
	2023	0.80	2.10	0.57	-2.23
	2024	0.96	2.86	0.51	-6.55
	2025	0.85	2.35	0.55	-8.74

This pattern was also evident at the pitcher level shown in Figure 13. Across seasons, the distributions of N-of-1 adjusted R^2 values were centered above the corresponding group-based values, and RMSE distributions were lower for the N-of-1 models than for the group-based model. These results showed a consistent separation in

velocity model performance between the two frameworks across both season-level summaries and pitcher-level comparisons.

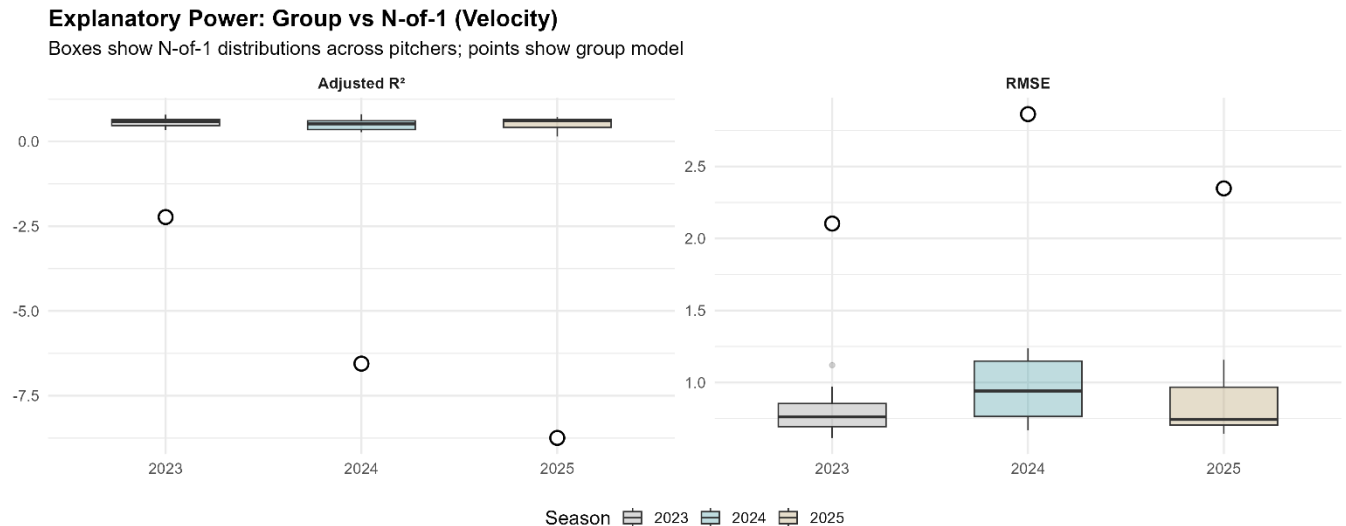


Figure 13: Pitcher-level velocity model performance comparing individualized N-of-1 and group-based approaches across the 2023, 2024, and 2025 seasons. Boxplots represent the distribution of N-of-1 within pitcher adjusted R^2 and RMSE values across pitchers within each season, whereas points represent the corresponding group-based model values. Negative adjusted R^2 values for the group-based model indicate poor fit when evaluated at the pitcher level relative to a baseline model

Biomechanics Model Comparison

Biomechanics model performance also differed consistently between the individualized N-of-1 and group-based frameworks across the four selected outcomes. As shown in Table XI, N-of-1 models had lower RMSE and higher within-pitcher adjusted R^2 than the group-based model for maximum elbow varus torque, shoulder abduction angle at ball release, lead knee flexion angle at ball release, and maximum hip-shoulder separation. This pattern was present in both 2024 and 2025 and was retained in the overall summaries for all four outcomes.

Table XI: Comparison of biomechanics model performance for individualized N-of-1 and group-based models across selected representative outcomes

Outcome	Season	N-of-1 RMSE	Group Model RMSE	N-of-1 Adj. R²	Group Model Adj. R²
Max Elbow	All	6.16	18.94	0.52	-3.93
Varus	2024	7.52	20.92	0.44	-5.51
Torque	2025	5.35	16.06	0.57	-2.76
Shoulder	All	1.82	6.21	0.51	-6.92
Abduction	2024	1.81	6.88	0.47	-9.15
Angle at BR	2025	1.82	5.50	0.51	-4.80
Lead Knee	All	5.00	14.53	0.33	-7.18
Flexion	2024	5.23	15.91	0.38	-10.58
Angle at BR	2025	4.87	12.92	0.31	-4.95
Max Hip	All	2.17	7.72	0.61	-4.94
Shoulder	2024	2.25	7.38	0.60	-4.06
Separation	2025	2.12	7.71	0.61	-5.38

The magnitude of the separation varied across outcomes. Maximum hip-shoulder separation and maximum elbow varus torque showed the highest N-of-1 adjusted R² values overall, whereas lead knee flexion angle at ball release showed comparatively lower adjusted R² despite still outperforming the group-based model. A similar pattern was observed for RMSE, with individualized models showing the lowest error across all outcomes and seasons relative to the corresponding group-based model values.

This pattern was also evident at the pitcher level in Figures 14 and 15. Across both seasons, the N-of-1 distributions were centered above the group-based values for adjusted R^2 and below the group-based values for RMSE for each selected biomechanics outcome. Together, these results show a consistent separation in model performance between individualized and group-based frameworks across multiple biomechanical domains.

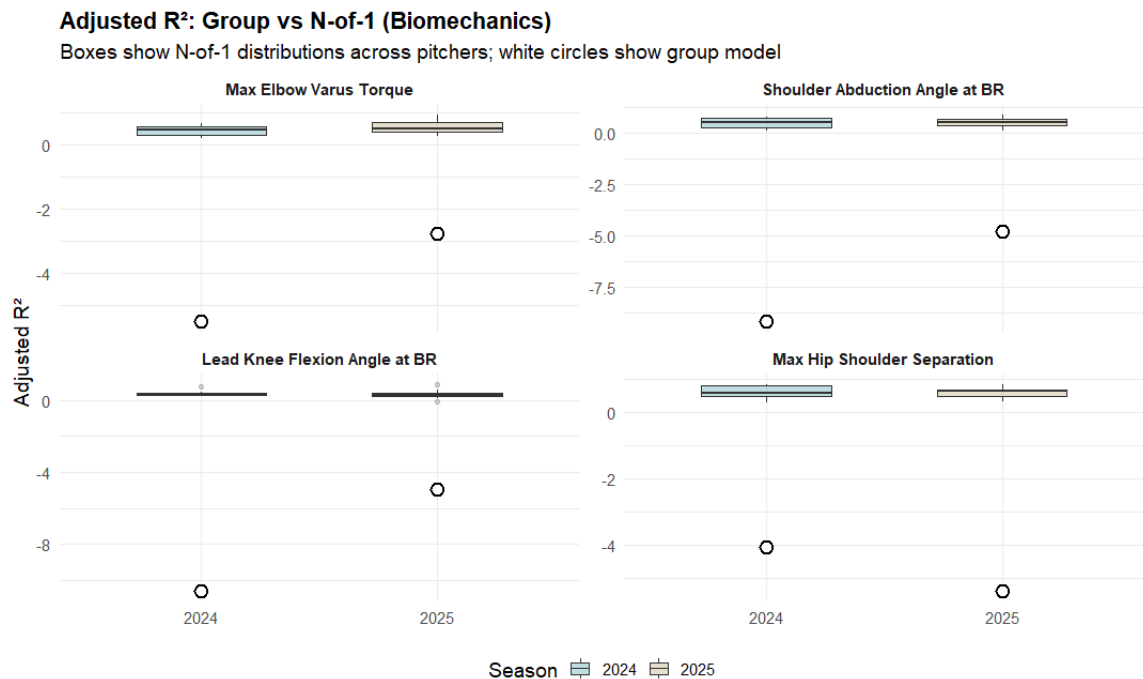


Figure 14: Pitcher-level adjusted R^2 for selected biomechanics outcomes comparing individualized N-of-1 and group-based approaches across the 2024 and 2025 seasons. Boxplots represent the distribution of N-of-1 within-pitcher adjusted R^2 values across pitchers within each season, whereas white circles represent the corresponding group-based model values.

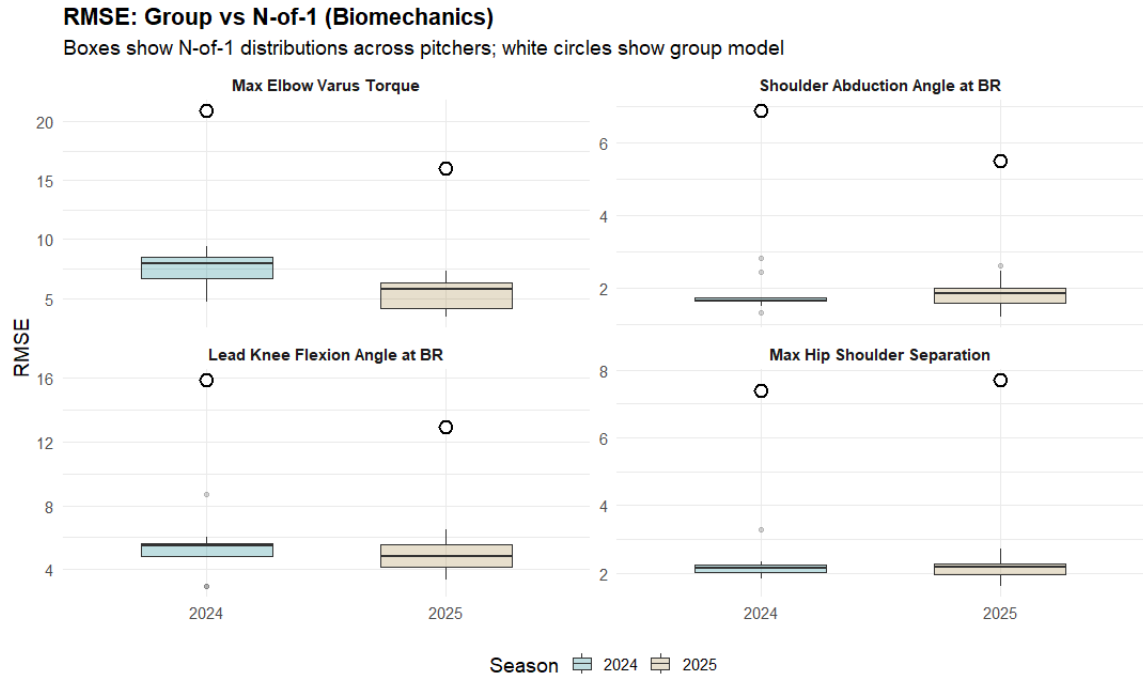


Figure 15: Pitcher-level RMSE for selected biomechanics outcomes comparing individualized N-of-1 and group-based approaches across the 2024 and 2025 seasons. Boxplots represent the distribution of N-of-1 RMSE values across pitchers within each season, whereas white circles represent the corresponding group-based model values.

Residual Autocorrelation Comparison

Residual autocorrelation patterns also differed across modeling stages. For velocity, the baseline models demonstrated generally positive within-appearance lag-1 autocorrelation, whereas the final N-of-1 models showed lower residual autocorrelation overall. This reduction was most apparent among N-of-1 models that retained AR(1) structures, which tended to show residual ACF1 values below those observed in the baseline models and below those from N-of-1 models fit without AR correction (Figure 16).

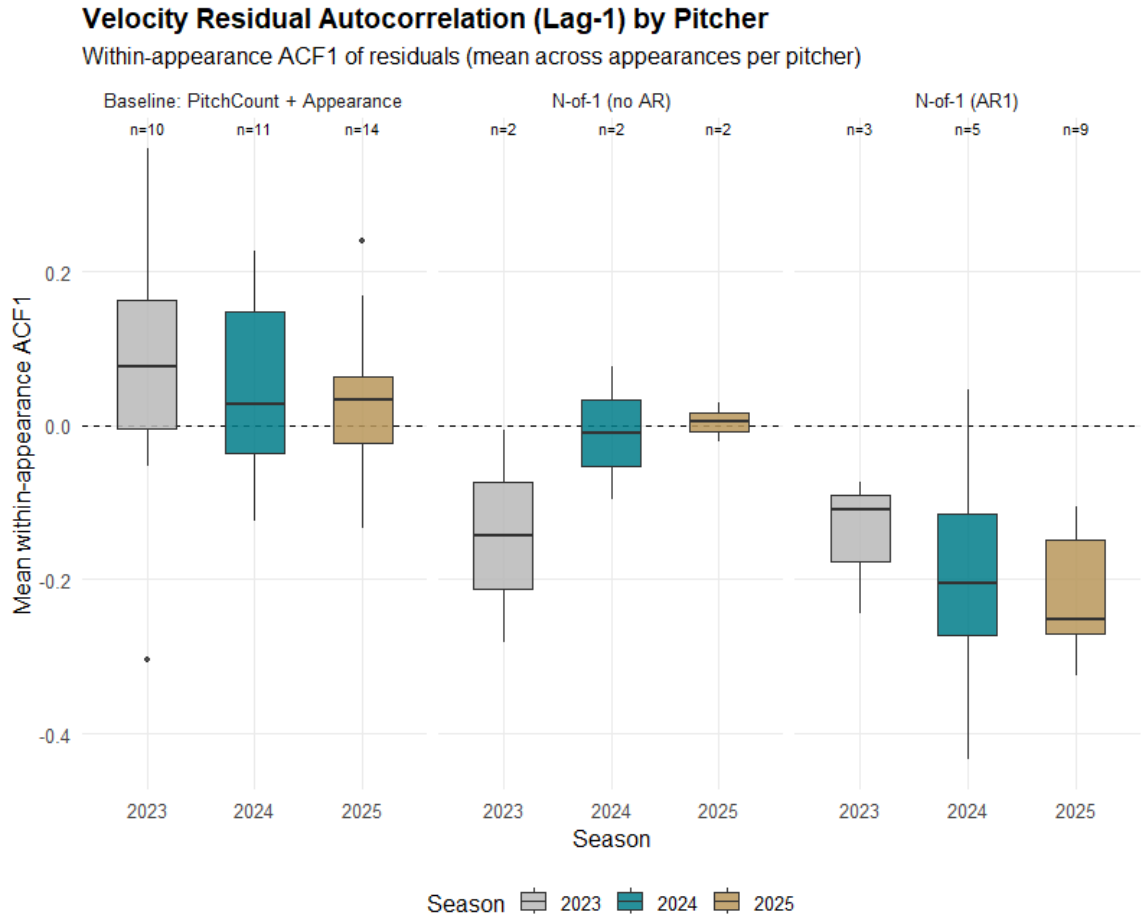


Figure 16: Pitcher-level residual autocorrelation for velocity models across modeling stages and seasons. Boxplots represent mean within-appearance lag-1 residual autocorrelation (ACF1) across pitchers for baseline models, final N-of-1 models without AR correction, and final N-of-1 models retaining AR(1) structures.

A similar general pattern was observed for the selected biomechanics outcomes. Baseline models often showed within-appearance residual autocorrelation at or above zero, whereas final N-of-1 models, particularly those retaining AR(1) structures, tended to show lower lag-1 autocorrelation. This was evident for representative outcomes such as maximum elbow varus torque, where residual dependence was reduced after individualized model fitting (Figure 17). However, the number of biomechanics models retaining AR(1) structures varied across outcomes and seasons, and several AR(1)-retained subgroups were small. Accordingly, these figures are best interpreted as

descriptive evidence that residual dependence was reduced in many cases after individualized modeling rather than as a balanced comparison of equally sized model groups.

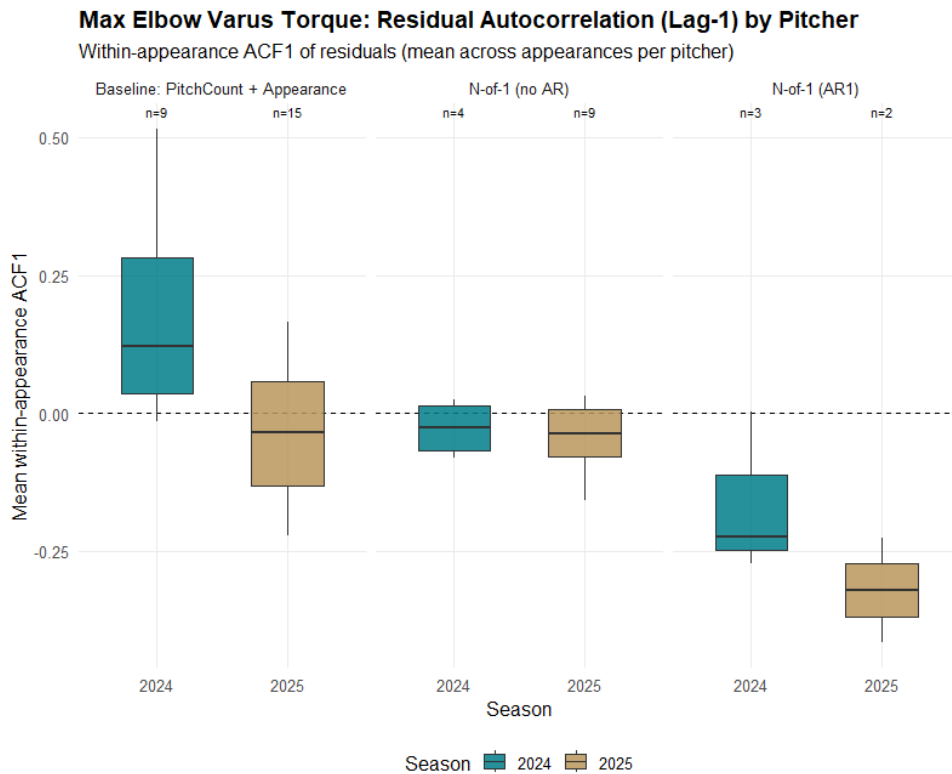


Figure 17: Pitcher-level residual autocorrelation for maximum elbow varus torque models across modeling stages and seasons. Boxplots represent mean within-appearance lag-1 residual autocorrelation (ACF1) across pitchers for baseline models, final N-of-1 models without AR correction, and final N-of-1 models retaining AR(1) structures.

Discussion

The purpose of this chapter was to compare group-based and individualized N-of-1 models for quantifying rolling pitch count associations in collegiate baseball pitchers. The primary finding was that individualized models demonstrated stronger performance than group-based models across both outcome domains examined. For velocity, N-of-1 models consistently showed lower RMSE and higher within-pitcher adjusted R^2 than the group-based model across all seasons. The same overall pattern was observed for the

selected biomechanics outcomes, where individualized models again demonstrated lower prediction error and stronger explanatory performance than the corresponding group-based framework. Residual autocorrelation findings also supported this comparison, as baseline models often showed positive within-appearance lag-1 autocorrelation, whereas final N-of-1 models, particularly those retaining AR(1) structures, generally demonstrated lower residual dependence. Together, these findings suggest that the individualized framework was not only feasible in the prior chapters, but also more effective than a population-level approach for representing pitcher-specific workload-response patterns. This interpretation is further supported by the consistently negative adjusted R^2 values observed for group-based models when evaluated at the pitcher level, indicating poor representation of within-pitcher variation relative to a baseline model.

The velocity comparison provided the clearest example of this difference. Across the 35 pitcher-seasons included in this chapter, individualized models consistently outperformed the group-based model in both season-level summaries and pitcher-level distributions. This is important because velocity served as the primary performance outcome in the chapter and is also one of the most interpretable outcomes in applied baseball settings. In practice, velocity is often the first metric used to judge whether a pitcher is maintaining or losing effectiveness over the course of an appearance. The present findings indicate that the magnitude and shape of workload-response patterns in velocity varied meaningfully across pitchers, rather than following a single shared trend. Although the group model was able to estimate an average workload-response relationship, that average pattern masked meaningful heterogeneity in velocity responses across pitchers.

The same general conclusion extended to biomechanics, although the magnitude of the difference varied across outcomes. Individualized models outperformed the group-

based framework for maximum elbow varus torque, shoulder abduction angle at ball release, lead knee flexion angle at ball release, and maximum hip-shoulder separation in both 2024 and 2025. Because these outcomes reflect different mechanical domains of the pitching motion, including upper-extremity loading, upper-extremity positioning, lower-extremity mechanics, and kinetic-chain organization, the consistency of this pattern across outcomes suggests that workload-response heterogeneity extends across multiple biomechanical domains rather than being confined to a single aspect of the pitching motion. More importantly, the variation in adjusted R^2 across outcomes indicates that workload response was not expressed uniformly across biomechanical domains, with some measures demonstrating stronger and more consistent associations than others. Some outcomes, such as maximum hip-shoulder separation and maximum elbow varus torque, showed comparatively stronger individualized explanatory performance, whereas others, such as lead knee flexion angle at ball release, showed lower adjusted R^2 overall despite still favoring the N-of-1 framework. This pattern is consistent with the broader findings of Chapter 3, where workload-related responses differed not only across pitchers, but also across the specific performance and mechanical domains being measured.

The residual autocorrelation results add an important layer to this comparison. One possible counterargument would be that the explanatory performance advantage of the individualized models simply reflected overfitting or greater flexibility, rather than better representation of the underlying data structure. However, the residual autocorrelation figures suggest that the individualized models were also better aligned with the serial nature of the pitch-level observations. For velocity, baseline models often showed positive within-appearance lag-1 autocorrelation, whereas final N-of-1 models showed lower residual dependence overall, particularly when AR(1) structures were retained. A similar general pattern was observed for the selected biomechanics outcomes, although the

number of biomechanics models retaining AR(1) structures varied across outcomes and seasons. These findings do not mean that all serial dependence was eliminated in every model, but they do support that individualized model development was better able to account for within-appearance dependence than simpler baseline approaches. That is especially relevant in this setting because pitch-level data are sequential by nature, and models that fail to account for that structure may produce misleadingly simplified summaries of workload-response patterns.

These results help clarify what group-based and individualized models each contributed in this dissertation. Group-based models remain useful for describing average workload-response patterns across pitchers and for providing broader league-level or population-level summaries. They are also better positioned to incorporate stable between-athlete characteristics, such as age, height, or baseline physical measures, that are commonly used in group frameworks but are less informative in N-of-1 models when they do not vary meaningfully within a pitcher-season. In contrast, the present findings show that individualized models were better suited for representing pitcher-specific workload-response patterns within this dataset. From that standpoint, the two approaches should not be viewed as complete substitutes. Rather, they are complementary tools that address different inferential targets and can help offset one another's limitations. Group models can provide broader context and general benchmarks, whereas N-of-1 models can better represent the specific athlete when the goal is pitcher-level monitoring or interpretation.

This chapter therefore extends the findings of Chapters 2 and 3 in an important way. Chapter 2 showed that rolling pitch count associations with fastball performance metrics were individualized, and Chapter 3 showed that workload-related biomechanical patterns were likewise individualized and outcome-specific. The present chapter builds on

those findings by showing that this individualized structure was not merely descriptive, but comparatively meaningful when evaluated against a more traditional group-based alternative. In other words, the N-of-1 framework did not simply produce interpretable pitcher-specific models; it also provided stronger representation of within-pitcher workload-response patterns than the population-level comparison models used in this chapter. Taken together, these three chapters support the broader dissertation premise that when workload-related responses are heterogeneous, individualized models may be better aligned with both the scientific question and the applied decision target than models centered on the group average.

Limitations

Several limitations should be considered when interpreting these findings. First, all comparisons were conducted within a single collegiate baseball population, which limits generalizability to other settings such as professional, high school, or youth baseball. Second, the biomechanics comparison set was more limited than the velocity comparison set because biomechanics data were available in fewer collection settings and across fewer seasons, resulting in fewer eligible model series and repeated observations. Third, although broader biomechanics comparison outputs were available, this chapter emphasized four representative outcomes, so not all modeled biomechanics variables were presented in equal detail. Fourth, the number of models retaining AR(1) structures varied across outcomes and seasons, particularly in the biomechanics analyses, which makes those residual autocorrelation comparisons best interpreted descriptively. Fifth, athlete-level variables commonly incorporated into group-based models, such as age, height, and single-timepoint physical characteristics, were less compatible with the individualized framework because they did not vary meaningfully within pitcher-season series. As a result, the present comparison emphasized repeatedly measured workload-

related and contextual variables rather than broader between-athlete characteristics. Stronger comparative model performance does not establish causality. These findings support that individualized models better represented pitcher-specific workload-response structure in this dataset, but they do not by themselves establish that rolling pitch count caused the observed performance or biomechanical changes. Because individualized models required sufficient within-pitcher observations for estimation, comparisons were limited to pitcher-seasons meeting data sufficiency criteria, and it remains unclear how these approaches would perform relative to one another in lower-data contexts. In such settings, population-level models may provide useful baseline estimates until sufficient observations are available to support individualized modeling.

Applied Implications and Future Directions

These findings suggest that the most useful applied approach is not choosing between group-based and individualized models as complete substitutes, but using them together for different purposes. Group-based models remain valuable for broader staff-level summaries and for incorporating stable between-athlete characteristics that are less informative in an N-of-1 framework. In contrast, individualized models appear better suited for representing pitcher-specific workload-response patterns and may therefore be more useful when the goal is player-level monitoring, recovery planning, or interpretation of changes over time.

The present results also strengthen the rationale for using individualized models as the foundation for later monitoring applications in this dissertation. Because the N-of-1 framework more consistently represented within-pitcher structure than the group-based alternatives examined here, it provides a stronger basis for downstream pitcher-specific monitoring approaches. At the same time, the complementary value of group-based

models supports their continued use for broader contextual interpretation alongside individualized monitoring.

Further work is still needed to extend these comparisons to other baseball populations, additional pitch types, and a wider range of biomechanics and performance outcomes. It will also be important for future studies to continue examining how data sufficiency, serial dependence, and uneven outcome availability influence comparative model performance across different athlete-monitoring contexts.

Conclusion

In conclusion, individualized N-of-1 models demonstrated stronger performance than group-based models for quantifying rolling pitch count associations in collegiate baseball pitchers. Across both velocity and the selected biomechanics outcomes, N-of-1 models showed lower prediction error and stronger within-pitcher explanatory performance, indicating that pitcher-specific workload-response patterns were not adequately represented by population-level averages alone. At the same time, the findings of this chapter support that group-based and individualized models should be viewed as complementary rather than interchangeable approaches, with each contributing different types of information depending on the inferential goal. Overall, these results support the use of individualized modeling when the objective is to understand or monitor workload response within a specific pitcher.

References

1. Bakshi NK, Inclan PM, Kirsch JM, Bedi A, Agresta C, Freehill MT. Current Workload Recommendations in Baseball Pitchers: A Systematic Review. *Am J Sports Med.* 2020;48(1):229-241. doi:10.1177/0363546519831010
2. Birfer R, Sonne MW, Holmes MW. Manifestations of muscle fatigue in baseball pitchers: a systematic review. *PeerJ.* 2019;7:e7390. doi:10.7717/peerj.7390
3. Dick R, Sauers EL, Agel J, et al. Descriptive Epidemiology of Collegiate Men's Baseball Injuries: National Collegiate Athletic Association Injury Surveillance System, 1988–1989 Through 2003–2004. *J Athl Train.* 2007;42(2):183-193.
4. Clarsen B, Myklebust G, Bahr R. Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) overuse injury questionnaire. *Br J Sports Med.* 2013;47(8):495-502. doi:10.1136/bjsports-2012-091524
5. Nicholson KF, Mylott JA, Waterman BR, Bullock GS. Pitching Biomechanics Normative Values and Kinetic Differences by Competition Level. *J Surg Orthop Adv.* 2022;31(3):177-180.
6. Daza EJ. Causal Analysis of Self-tracked Time Series Data Using a Counterfactual Framework for N-of-1 Trials*. *Methods Inf Med.* 2018;57(Suppl 1):e10-e21. doi:10.3414/ME16-02-0044
7. Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *J Clin Epidemiol.* 1997;50(4):401-410. doi:10.1016/s0895-4356(96)00429-5

8. Kravitz RL, Duan N, Braslow J. Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages. *Milbank Quarterly*. 2004;82(4):661-687. doi:10.1111/j.0887-378x.2004.00327.x
9. Araujo A, Julious S, Senn S. Understanding Variation in Sets of N-of-1 Trials. *PLoS One*. 2016;11(12):e0167167. doi:10.1371/journal.pone.0167167
10. Vieira R, McDonald S, Araújo-Soares V, Sniehotta FF, Henderson R. Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies. *Health Psychology Review*. 2017;11(3):222-234. doi:10.1080/17437199.2017.1343680
11. Dowling B, McNally MP, Chaudhari AMW, Oñate JA. A Review of Workload-Monitoring Considerations for Baseball Pitchers. *J Athl Train*. 2020;55(9):911-917. doi:10.4085/1062-6050-0511-19
12. Reinold MM, Dowling B, Fleisig GS, et al. An Interval Throwing Program for Baseball Pitchers Based upon Workload Data. *IJSPT*. 2024;19(3):326-336. doi:10.26603/001c.94146
13. Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, Robertson SJ. Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *Int J Sports Physiol Perform*. 2017;12(2):230-234. doi:10.1123/ijsp.2015-0791
14. Trasolini NA, Nicholson KF, Mylott J, Bullock GS, Hulburt TC, Waterman BR. Biomechanical Analysis of the Throwing Athlete and Its Impact on Return to Sport. *Arthrosc Sports Med Rehabil*. 2022;4(1):e83-e91. doi:10.1016/j.asmr.2021.09.027
15. Diffendaffer AZ, Bagwell MS, Fleisig GS, et al. The Clinician's Guide to Baseball Pitching Biomechanics. *Sports Health: A Multidisciplinary Approach*. 2023;15(2):274-281. doi:10.1177/19417381221078537

16. Yang L, Armijo-Olivo S, Gross DP. Single-Case Experimental Design in Rehabilitation: Basic Concepts, Advantages, and Challenges. *Am J Phys Med Rehabil*. 2023;102(5):454-460. doi:10.1097/PHM.0000000000002215
17. Schmid C, Yang J. Bayesian Models for N-of-1 Trials. *Harvard Data Science Review*. 2022;(Special Issue 3). doi:10.1162/99608f92.3f1772ce
18. Daza EJ. Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Track Health Data. *arXiv*. Preprint posted online January 22, 2019:arXiv:1901.03423. doi:10.48550/arXiv.1901.03423
19. Chalmers PN, Erickson BJ, Ball B, Romeo AA, Verma NN. Fastball Pitch Velocity Helps Predict Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers. *Am J Sports Med*. 2016;44(8):2130-2135. doi:10.1177/0363546516634305
20. Manzi JE, Yu JS, Sudah SY, et al. A Systematic Review of Lower-Body Kinematic and Strength Factors Associated With Pitch Velocity in Adult Baseball Pitchers. *Journal of Sport Rehabilitation*. 2023;32(4):440-448. doi:10.1123/jsr.2022-0059
21. Werner SL, Suri M, Guido JA, Meister K, Jones DG. Relationships between ball velocity and throwing mechanics in collegiate baseball pitchers. *J Shoulder Elbow Surg*. 2008;17(6):905-908. doi:10.1016/j.jse.2008.04.002
22. Luera MJ, Dowling B, Muddle TWD, Jenkins NDM. Differences in Rotational Kinetics and Kinematics for Professional Baseball Pitchers With Higher Versus Lower Pitch Velocities. *J Appl Biomech*. 2020;36(2):68-75. doi:10.1123/jab.2019-0235
23. Stodden DF, Fleisig GS, McLean SP, Andrews JR. Relationship of biomechanical factors to baseball pitching velocity: within pitcher variation. *J Appl Biomech*. 2005;21(1):44-56. doi:10.1123/jab.21.1.44

24. Shanley E, Bailey L, Sandago MP, Pinkerton A, Singleton SB, Thigpen CA. The use of a pitch count estimator to calculate exposure in collegiate baseball pitchers. *Phys Ther Sport*. 2015;16(4):344-348. doi:10.1016/j.ptsp.2015.02.002
25. International Olympic Committee Injury and Illness Epidemiology Consensus Group, Bahr R, Clarsen B, et al. International Olympic Committee Consensus Statement: Methods for Recording and Reporting of Epidemiological Data on Injury and Illness in Sports 2020 (Including the STROBE Extension for Sports Injury and Illness Surveillance (STROBE-SIIS)). *Orthop J Sports Med*. 2020;8(2):2325967120902908. doi:10.1177/2325967120902908
26. Van Calster B, van Smeden M, De Cock B, Steyerberg EW. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. *Stat Methods Med Res*. 2020;29(11):3166-3178. doi:10.1177/0962280220921415
27. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010;33:1-22. doi:10.18637/jss.v033.i01
28. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Published online October 4, 2000:1.9-4. doi:10.32614/CRAN.package.mgcv
29. Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA. A linear exponent AR(1) family of correlation structures. *Stat Med*. 2010;29(17):1825-1838. doi:10.1002/sim.3928
30. Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*. 2019;11(3):e1460. doi:10.1002/wics.1460
31. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 23rd printing. Cambridge Univ. Press; 2007.

32. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014;7(3):1247-1250. doi:10.5194/gmd-7-1247-2014
33. Raykov T, DiStefano C. Evaluating Change in Adjusted R-Square and R-Square Indices: A Latent Variable Method Application. *Educ Psychol Meas*. Published online April 11, 2025:00131644251329178. doi:10.1177/00131644251329178
34. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Published online October 4, 2000:1.9-4. doi:10.32614/CRAN.package.mgcv

CHAPTER V: Practical Applications of Individualized Pitch Count Models for Monitoring Collegiate Baseball Pitchers

Garrett Fernandez¹, Garrett Bullock^{1,3}, Patrick Ward⁴, Joel Stitzel^{1,2}, Ryan
McGinnis¹, Tanner Filben⁵, Kristen Nicholson^{1,3}

¹Department of Biomedical Engineering, Wake Forest School of Medicine, Winston-Salem, NC, ²Virginia Tech-Wake Forest School of Biomedical Engineering and Sciences, Winston-Salem, NC, ³Department of Orthopedic Surgery and Rehabilitation, Wake Forest School of Medicine, Winston Salem, NC, ⁴Seattle Seahawks, Seattle, WA, USA, ⁵Elemance, LLC, Winston-Salem, NC

Abstract

The purpose of this chapter was to translate individualized fastball velocity models into a practical monitoring framework for collegiate baseball pitchers using appearance-level cumulative sum (CUSUM) procedures. Pitcher-season specific expected velocity values generated from the finalized N-of-1 models in Chapter 2 were used as the reference process and observed-minus-expected residuals were monitored sequentially to detect sustained below-expected performance across a season. Initial back testing and grid-search tuning evaluated baseline reference windows of 3 to 5 appearances, minimum appearance sizes of 10 to 20 pitches, k values of 0.25 to 0.75 standard deviations, h values of 3 to 5 standard deviations, and both raw and standardized residual specifications. The final selected implementation was an appearance-level, lower-sided CUSUM using raw residuals, with a baseline window of 3 appearances, a minimum of 10 pitches per appearance, and control parameters of $k = 0.75$ standard deviations and $h = 5$ standard deviations. Of the 35 pitcher-seasons with finalized individualized velocity models, 26 were retained for the final appearance-level monitoring procedure, with 9 excluded after monitoring-specific requirements were applied. Across the retained cohort, 7 pitcher-seasons generated at least one appearance-level signal, 19 did not generate a signal, and the median first-signal appearance among signaled pitcher-seasons was appearance 5. Representative examples demonstrated three distinct monitoring patterns under the final specification: stable non-signal behavior, transient downward deviation with later recovery, and more sustained below-expected performance. Overall, this chapter demonstrated that individualized velocity models could be extended into a practical appearance-level monitoring framework and that the resulting CUSUM outputs captured heterogeneous pitcher-specific surveillance patterns across a competitive season.

Introduction

Chapters 2 through 4 established the analytical foundation for this dissertation. Chapter 2 demonstrated that rolling pitch count associations with fastball performance metrics could be modeled at the pitcher-season level using individualized N-of-1 methods, with fastball velocity emerging as the most consistently interpretable performance outcome.¹⁻³ Chapter 3 extended this individualized framework to pitching biomechanics and further supported that workload-related responses were not expressed uniformly across athletes or across outcome domains. Chapter 4 then showed that when the inferential target is the individual pitcher, N-of-1 models provided stronger representation of within-pitcher workload-response structure than traditional population-level alternatives. Together, these findings support the broader premise that workload response in baseball is heterogeneous and that individualized models are better aligned with player-level decision-making than approaches centered primarily on the group average.⁴

At the same time, establishing individualized workload-response associations is not itself the final applied objective.^{1,2,5} In practice, coaches, analysts, and sports medicine staff are not only interested in whether workload-related patterns can be modeled, but whether those models can be used to identify meaningful changes in a pitcher over time and whether those changes are associated with subsequent changes in performance or injury risk. This distinction is important because the practical problem in baseball is not simply to describe an average workload effect in retrospect, but to determine whether a specific pitcher is showing signs of atypical response or accumulating deviation from their expected pattern in ways that may have downstream performance or health implications.

This issue is especially relevant in baseball because commonly used workload indicators such as pitch count remain descriptive rather than decisional on their own. Pitch count provides a simple estimate of throwing exposure, but it does not indicate whether a

pitcher is responding normally to that exposure or deviating from his expected pattern.¹⁻³ Prior chapters also showed that no single population-average relationship adequately characterizes all pitchers, and even within individualized analyses, workload-related responses may differ in direction, magnitude, and outcome domain. As a result, practical monitoring requires more than simply tracking how many pitches were thrown. It requires a framework that can compare observed performance against pitcher-specific expectations and identify when departures from those expectations become systematic enough to warrant attention.

Among the modeled outcomes in this dissertation, fastball velocity provides the most practical starting point for that application. Velocity is routinely tracked, readily interpretable to applied stakeholders, and showed the strongest and most consistent individualized model performance in the earlier analyses.^{2,6} For that reason, velocity provides a useful anchor for translating individualized modeling into a monitoring framework. Importantly, such a framework should not be interpreted as a direct measure of fatigue or injury risk. Rather, it is intended to detect meaningful deviations from expected pitcher-specific performance, which may then be interpreted alongside broader contextual, clinical, and workload information.

Accordingly, the purpose of this chapter was to translate the individualized modeling framework developed in the previous chapters into a practical monitoring approach for collegiate baseball pitchers.⁷⁻¹¹ Using individualized fastball velocity models as the prediction foundation, this chapter applies cumulative sum monitoring methods to evaluate whether systematic deviations from expected performance can be identified over time at the pitcher level. In doing so, this chapter shifts the dissertation from model development and comparison toward application, with the broader goal of illustrating how

individualized workload models may be used to support ongoing monitoring and decision-making in applied baseball settings.

Monitoring Implementation

This study was approved by the Wake Forest University School of Medicine Institutional Review Board. Chapter 5 used the same eligible pitcher-season fastball data and finalized individualized velocity models described previously in Chapter 2. Consistent with the dissertation's broader analytic framework, monitoring was conducted at the pitcher-season level using repeated fastball observations ordered within appearances. The individualized models used here were carried forward from the prior modeling pipeline rather than redeveloped for the present chapter.

Source of Individualized Predictions

The individualized expected values used for monitoring in this chapter were obtained from the final pitcher-season specific fastball velocity models developed in Chapter 2. These finalized N-of-1 velocity models were carried forward as the prediction foundation for the monitoring procedures implemented in Chapter 5.^{7,8,10} As described previously, final individualized models were fit separately for each pitcher-season using generalized additive mixed models, with rolling pitch count retained as the primary exposure and additional contextual predictors incorporated according to the prior screening and model-selection process.¹² Candidate models were also evaluated with and without AR(1) residual correlation structures specified within appearance, and final individualized model selection was based primarily on Akaike information criterion with consideration of model fit and interpretability.^{13,14} As a result, final model form was allowed to vary across pitcher-seasons according to the predictor structure supported within each individual series. Expected velocity was not generated from a single common equation and was not based on pitch count alone. Rather, each expected value reflected the fitted

value from that pitcher-season's own finalized model, conditional on the specific covariate structure retained for that series. In this way, the monitoring reference preserved the same individualized and context-adjusted prediction framework established in Chapter 2.

For the purposes of monitoring, the saved finalized models were applied back to the underlying pitch-level data to generate fitted values for each eligible fastball observation. These fitted values served as the individualized expected velocity reference for subsequent residual construction. A separate downstream export step was then used to assemble monitoring inputs from the observed and predicted values generated from those finalized models. The export procedure produced both pitch-level and appearance-level monitoring files derived from the same individualized model outputs.

Construction of Monitoring Inputs

Monitoring inputs were constructed by pairing the observed fastball velocity series with the individualized fitted values generated from the finalized pitcher-season specific velocity models. At the pitch level, this produced an observed value, a model-predicted value, and an observed-minus-expected residual for each eligible fastball observation. These pitch-level residuals provided the most granular monitoring series and were retained as a secondary surveillance stream. The export script was structured specifically to preserve pitch-level observed velocity, predicted velocity, residuals, and related residual summaries for downstream monitoring use.

In addition to the pitch-level residual series, an appearance-level monitoring series was derived by aggregating pitch-level quantities within each appearance. For each pitcher-season and appearance, pitch-level values were summarized to generate appearance-level monitoring inputs, including mean observed velocity, mean predicted velocity, and mean residual across pitches within that appearance. This appearance-level residual series served as the primary monitoring input for the final implementation. The

appearance-level focus was chosen because the practical monitoring objective was to identify whether small deviations were emerging from one appearance to the next rather than reacting to pitch-to-pitch fluctuation alone.

The appearance-level series was prioritized because it reduced short-term pitch-to-pitch variability and aligned more closely with the applied setting in which pitcher evaluation typically occurs across outings or appearances.^{15,16} Pitch-level monitoring was retained as a secondary comparison in order to examine whether a denser residual stream produced substantively different signal behavior. This two-level structure therefore allowed the same individualized model outputs to be evaluated both at the level of within-appearance pitch sequences and at the level of appearance-to-appearance monitoring.

Residual-based monitoring was used at both levels rather than monitoring observed velocity directly. This ensured that the surveillance target reflected deviation from the individualized expected value generated by the prior models rather than raw velocity alone. Negative residual values indicated observed velocity below the pitcher-specific expected value, whereas positive residual values indicated observed velocity above expectation. The final exported monitoring inputs therefore consisted of two linked residual series for each eligible pitcher-season: a pitch-level observed-minus-expected series and an appearance-level mean residual series.

CUSUM Implementation and Tuning

Cumulative sum (CUSUM) monitoring is a sequential surveillance method in which deviations from an expected reference process are accumulated over time.^{15–17} At each new observation, the cumulative statistics are updated according to the size and direction of the deviation, so that repeated small departures in the same direction can eventually exceed a decision threshold even when no single observation appears extreme on its own.

In the present study, CUSUM was applied to residual-based monitoring series to detect sustained negative drift relative to individualized expected velocity.

CUSUM monitoring was applied to the residual series in order to detect systematic negative drift relative to pitcher-specific expected performance.^{15,16} Initial back testing procedures were implemented using both appearance-level and pitch-level monitoring inputs and preserved both raw and standardized residual forms. These runs also retained appearance-level as the primary monitoring level, pitch-level as a secondary monitoring level, a lower-sided primary direction, a two-sided secondary comparison, and exported the associated settings for review. In the initial back test script, baseline residual behavior was estimated from an early reference window and monitoring was then applied to subsequent observations using standard deviation-scaled decision parameters.

After these initial runs, a limited grid search was used to tune the CUSUM specification.¹⁸ The grid-search validation procedure was explicitly designed with appearance-level, one-sided lower CUSUM as the primary target, with secondary comparisons including pitch-level monitoring and two-sided variants. The primary grid varied the number of baseline appearances, the minimum pitches required per appearance, the allowance parameter k , the decision threshold h , and whether monitoring used raw or standardized residuals. Specifically, the primary search evaluated baseline windows of 3, 4, or 5 appearances; minimum appearance sizes of 10, 15, or 20 pitches; k values of 0.25, 0.50, or 0.75 standard deviations; h values of 3, 4, or 5 standard deviations; and both raw and standardized residual specifications. Secondary comparisons were intentionally more restricted and examined both appearance-level and pitch-level monitoring, lower-sided and two-sided directions, smaller pitch-count thresholds, and a narrower parameter range.

The grid search used injected-shift simulations to assess practical detection performance under multiple decline sizes and both step and gradual shift structures.¹⁸ Ranking of candidate primary specifications emphasized false-alarm behavior, detection performance, and monitoring coverage using a weighted score. The final selected implementation for the primary monitoring framework was an appearance-level, lower-sided CUSUM using raw residuals, with a baseline window of 3 appearances, a minimum of 10 pitches per appearance, and control parameters of $k = 0.75$ standard deviations and $h = 5$ standard deviations. For visual interpretation of representative plots, separate practical severity bands at cumulative values of -4, -6, and -8 were overlaid based on a priori coach-informed thresholds; these bands were not used as the formal CUSUM signaling threshold. These values were then carried forward into the final appearance-level summaries and signal counts.

Signal Definition

Within the final implementation, the primary monitored quantity was the appearance-level mean residual, defined as the average difference between observed and individualized expected velocity across pitches within an appearance.^{15,16} Under the lower-sided CUSUM framework, a signal occurred when repeated negative residual deviations accumulated sufficiently to exceed the selected decision threshold. Thus, the primary signal definition reflected sustained below-expected appearance-level performance relative to the pitcher-specific model reference.

Pitch-level CUSUM procedures were also examined secondarily using the same observed-minus-expected residual logic at a finer temporal scale. However, because the final selected monitoring framework was appearance-level, subsequent emphasis in this chapter remained centered on appearance-to-appearance surveillance. Pitch-level signal

behavior was retained as a secondary comparison rather than the primary basis for interpretation.

This signal definition was applied consistently across eligible pitcher-seasons using the finalized parameter specification described above. As a result, the monitoring outputs presented in the following section reflect whether recent residual behavior remained within the expected range of short-term variation under the selected appearance-level CUSUM procedure or instead accumulated to a level meeting the predefined signal criterion.

Applied Monitoring Results

Cohort-level Monitoring Summary

Of the 35 pitcher-seasons with finalized individualized velocity models available from Chapter 2, 26 were retained for the final appearance-level monitoring procedure after applying the monitoring-specific eligibility requirements (Table XII). This reduction occurred because the final appearance-level CUSUM specification, selected through the grid-search tuning procedure, required not only a finalized individualized model but also sufficient appearance structure to support sequential monitoring. In particular, pitcher-seasons needed enough early appearances to establish the reference window and enough pitch volume within appearances to satisfy the minimum appearance-level requirement used in the final implementation. Table XII therefore summarizes both the final selected monitoring specification and the resulting retained cohort under that appearance-level framework.

Table XII: Cohort-level summary of the final appearance-level CUSUM monitoring sample.

Component	Value
Initial pitcher-seasons with finalized N-of-1 velocity models	35
Pitcher-seasons retained for final appearance-level monitoring	26
Pitcher-seasons excluded after monitoring requirements	9
Primary monitoring level	Appearance-level
Secondary monitoring level	Pitch-level
Final signal direction	Lower-sided
Residual form used in final primary specification	Raw residuals
Baseline reference window	3 appearances
Minimum pitches required per appearance	10
Final allowance parameter (k)	0.75 SD
Final decision threshold (h)	5 SD
Pitcher-seasons with at least one appearance-level signal	7
Pitcher-seasons without an appearance-level signal	19

Median first-signal appearance among signaled pitcher-seasons	5

Representative Pitcher Examples

The appearance-level CUSUM examples shown in Figures 18 through 20 illustrate three distinct monitoring patterns observed under the final selected specification. These examples were chosen to reflect a non-signaled series, a signaled series with later recovery, and a signaled series with more sustained below-expected performance. Together, they demonstrate that the selected appearance-level procedure did not produce a uniform monitoring pattern across pitcher-seasons. Additional appearance-level CUSUM plots for the remaining monitored pitcher-seasons are provided in Appendix C.

Player 8 provides an example of a retained pitcher-season that did not generate a signal under the final appearance-level specification (Figure 18). Although the series showed modest negative movement through the middle portion of the season, the cumulative residual pattern remained above the minimum practical threshold and later returned toward baseline. This example reflects a monitoring series in which observed

appearance-level performance remained broadly consistent with individualized expectation despite short-term fluctuation.

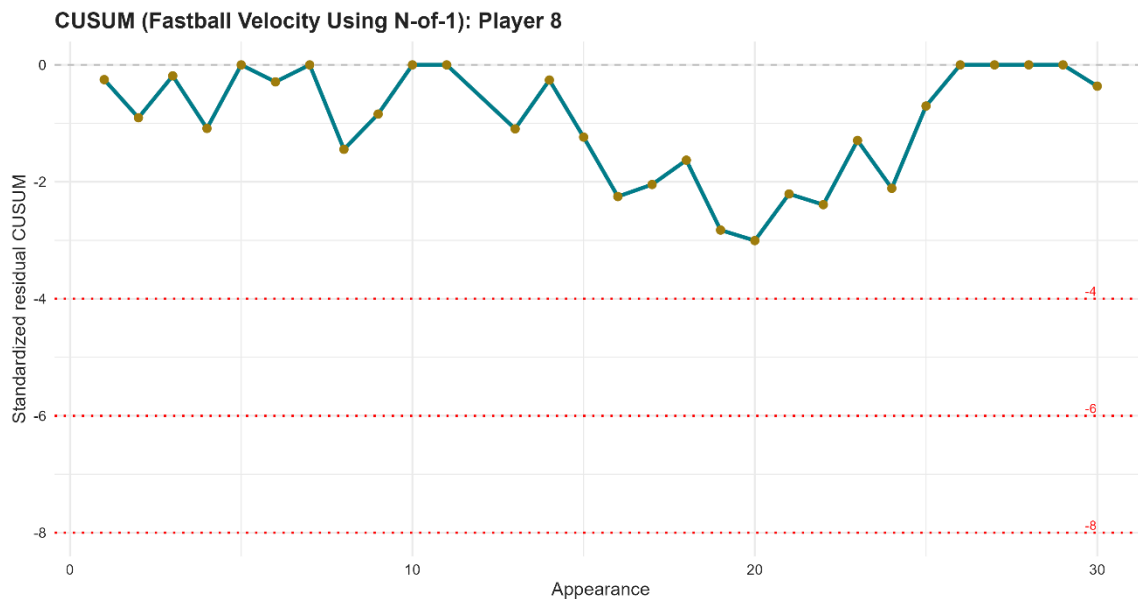


Figure 18: Appearance-level CUSUM for Player 8 under the final selected monitoring specification. The cumulative residual series remained above the practical reference threshold throughout the monitored sequence and did not generate a signal. Dashed horizontal lines at -4, -6, and -8 indicate practical severity bands, defined a priori with coach input, and are distinct from the formal CUSUM signaling threshold ($h = 5$ SD).

Player 13 illustrates a signal pattern characterized by pronounced downward movement followed by later recovery (Figure 19). In this series, cumulative negative residuals increased through the middle appearances and crossed the minimum practical threshold before subsequently returning upward toward baseline late in the season. This pattern shows that a signal could emerge under the final monitoring procedure without remaining permanently depressed across the full monitored sequence.

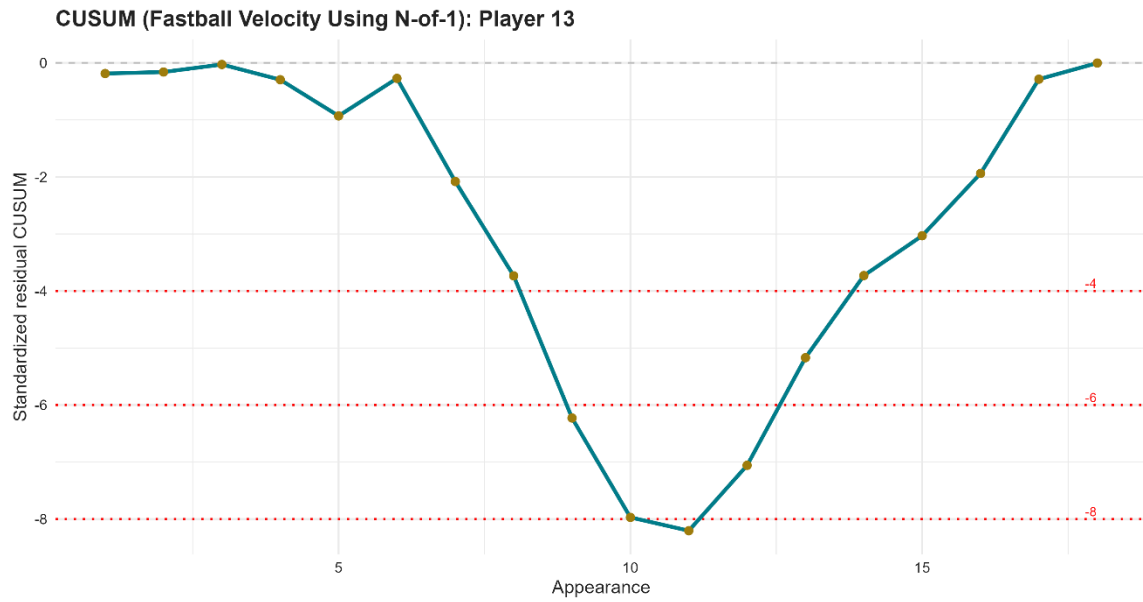


Figure 19: Appearance-level CUSUM for Player 13 under the final selected monitoring specification. The cumulative residual series crossed the practical reference threshold during the middle portion of the monitored sequence and later returned toward baseline. Dashed horizontal lines at -4, -6, and -8 indicate practical severity bands, defined a priori with coach input, and are distinct from the formal CUSUM signaling threshold ($h = 5$ SD).

Player 25 provides an example of a more sustained below-expected pattern (Figure 5.3). After remaining near baseline during the early appearances, the cumulative residual series declined steadily, crossed the minimum practical threshold, and continued downward across much of the remaining sequence. Unlike Player 13, this series did not return toward baseline by the end of monitoring and instead remained well below the practical reference bands. This example therefore represents a more persistent appearance-level deviation from individualized expected velocity under the selected monitoring framework. Although the present example emphasizes a negative CUSUM trend, consistent with the broader focus on potential fatigue-related decline across this dissertation, positive trends were also observed in some pitcher-seasons, reflecting sustained performance above individualized expectations. From an applied perspective, such patterns may indicate favorable adaptation or recovery progression, although the

present work primarily emphasizes deviations associated with potential performance decline.

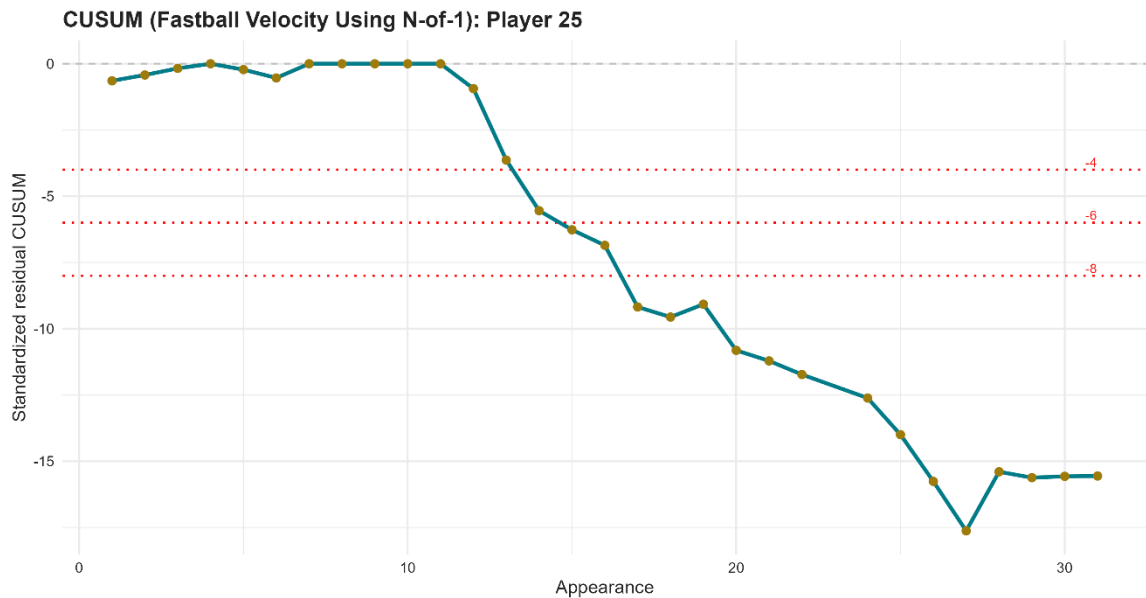


Figure 20: Appearance-level CUSUM for Player 25 under the final selected monitoring specification. The cumulative residual series crossed the practical reference threshold and remained below baseline across the later portion of the monitored sequence. Dashed horizontal lines at -4, -6, and -8 indicate practical severity bands, defined a priori with coach input, and are distinct from the formal CUSUM signaling threshold ($h = 5$ SD).

Discussion

This chapter extended the earlier dissertation findings by translating individualized fastball velocity models into a practical monitoring framework. Whereas Chapters 2 through 4 focused on identifying pitcher-specific workload-response associations and comparing individualized versus population-level approaches, the present chapter shifted to downstream application. Using finalized pitcher-season specific N-of-1 velocity models as the reference process, appearance-level CUSUM monitoring was used to evaluate

whether observed performance remained consistent with individualized expected velocity over time. The results showed that this framework could be implemented across a substantial subset of the modeled pitcher-seasons and that the resulting monitoring outputs captured multiple distinct patterns, including stable non-sigaled series, transient downward deviation with later recovery, and more sustained below-expected performance.

One important feature of the present framework was the decision to prioritize appearance-level monitoring rather than pitch-level monitoring as the primary implementation.^{15,16} Pitch-level residual series were retained and evaluated secondarily, but the final selected specification emphasized appearance-to-appearance change because that level of aggregation better aligned with the practical decision context in which pitchers are typically reviewed. In applied settings, intervention decisions are rarely made based on a single pitch. Instead, pitchers are more often evaluated across outings, appearances, or broader short-term performance windows. The appearance-level approach also reduced pitch-to-pitch variability, allowing the monitoring signal to reflect more sustained deviation from expected performance rather than isolated within-appearance noise.

The representative examples reinforced that the selected monitoring procedure did not produce a single uniform signal pattern across all pitchers. Instead, the retained series showed heterogeneous monitoring behavior, which is consistent with the broader dissertation findings that workload-response patterns are individualized rather than interchangeable across athletes. Player 8 demonstrated that modest fluctuation could occur without crossing the minimum practical threshold, while Players 13 and 25 illustrated that below-expected performance could emerge with different temporal structures. In one case, the cumulative deviation returned toward baseline, whereas in the other the

downward drift remained sustained across the later portion of the monitored sequence. These examples suggest that the practical value of the monitoring framework lies not only in whether a threshold is crossed, but also in the shape and persistence of the deviation pattern over time.

At the same time, the present framework should be interpreted carefully. A CUSUM signal in this context does not directly identify fatigue, injury risk, or mechanical breakdown.^{5,6,19,20} Rather, it indicates that observed velocity has become persistently lower than individualized expected velocity under the selected model-based reference process. That distinction is important because the same downward deviation could plausibly arise from multiple sources, including workload accumulation, recovery disruption, role changes, strategic pacing, mechanical adaptation, or other contextual influences not fully captured by the monitoring series itself.²¹ Accordingly, the output of this framework is best understood as an analytic surveillance signal that may prompt closer review, rather than as a standalone explanation of why performance changed.

Viewed in relation to the earlier chapters, the present findings extend the dissertation from individualized model development into applied monitoring. Chapters 2 and 3 established that rolling pitch count associations could be modeled at the pitcher-season level and that these associations varied across pitchers and outcome domains. Chapter 4 then showed that individualized models better preserved within-pitcher structure than traditional group-based alternatives when the primary inferential target was the individual athlete. Building from that foundation, the present chapter demonstrates that those individualized models can then be used to construct a practical surveillance framework based on deviation from pitcher-specific expected performance over time. This application does not replace the earlier explanatory work but instead provides one

example of how the individualized modeling framework developed in the prior chapters may be extended into an applied monitoring context.

Limitations

Several limitations should be considered when interpreting this monitoring framework. First, the primary surveillance outcome in this chapter was fastball velocity. Although velocity showed the strongest individualized model performance in the earlier analyses, monitoring based on velocity alone may not capture all relevant workload-related responses.^{2,6,19,22} A pitcher could remain near expected velocity while showing meaningful change in other domains, including biomechanics, command, perceived effort, or recovery-related measures.^{6,21,23}

Second, the monitoring procedure was dependent on the quality and structure of the underlying individualized models. Expected values were derived from the finalized pitcher-season specific velocity models carried forward from Chapter 2, meaning that the surveillance signal was only as informative as the individualized reference process on which it was based. As reflected by the reduction from 35 to 26 pitcher-seasons in the final appearance-level monitoring cohort, not all modeled pitcher-seasons supported the additional structural requirements needed for the selected CUSUM implementation.

Third, the final CUSUM specification depended on selected design choices, including the baseline reference window, minimum pitches per appearance, and decision parameters. Although these values were informed by back testing and grid-search tuning, different specifications could alter signal frequency, signal timing, or the balance between sensitivity and false alarms. In addition, the practical reference bands used in the representative plots were informed in part by coach input, which improved applied relevance but means that some thresholds reflected practical judgment rather than purely statistical criterion.

Finally, a signal in this framework did not identify the cause of the observed deviation. The CUSUM output indicated that recent performance had become inconsistent with individualized expected velocity, but it could not distinguish between possible explanations such as workload accumulation, recovery disruption, role change, strategic pacing, or other contextual influences. All analyses were also conducted within a single collegiate baseball population, which limits generalizability to other levels of play and organizations. This constraint highlights a broader practical challenge, as individualized monitoring frameworks may be less applicable in pitcher-seasons with limited observational history.

Future Directions

Future work should evaluate whether this monitoring framework can be extended beyond velocity-only surveillance to incorporate additional indicators such as biomechanics, pitch characteristics, or recovery-related measures within the same monitoring structure. Although the earlier chapters modeled multiple outcomes separately, the present chapter monitored only fastball velocity. A logical next step will therefore be to determine whether combining velocity-based monitoring with other relevant signals improves sensitivity to meaningful workload-related change. It will also be important to determine whether monitoring signals correspond with applied outcomes such as reduced performance, altered role, soreness, or recovery disruption.

An additional next step will be to evaluate how this framework can be adapted for pitchers with limited individualized history. The present implementation was best suited to pitcher-seasons with sufficient repeated data, so future studies should examine whether hybrid approaches can provide a practical transition for new or data-sparse pitchers before fully individualized monitoring is possible. External evaluation in other baseball settings

will also be needed to determine how well the current implementation generalizes beyond this single collegiate program.

Conclusion

This chapter illustrated how individualized fastball velocity models can be extended into an applied monitoring framework using appearance-level CUSUM procedures. Using pitcher-season specific expected values as the reference process, the selected monitoring approach was able to characterize different appearance-level monitoring patterns, including stable series, transient downward deviation, and more sustained below-expected performance in representative examples. Taken together, these findings provide a practical demonstration of how the individualized modeling framework developed in the earlier chapters may be translated into pitcher-level surveillance across a season.

References

1. Bakshi NK, Inclan PM, Kirsch JM, Bedi A, Agresta C, Freehill MT. Current Workload Recommendations in Baseball Pitchers: A Systematic Review. *Am J Sports Med.* 2020;48(1):229-241. doi:10.1177/0363546519831010
2. Dowling B, McNally MP, Chaudhari AMW, Oñate JA. A Review of Workload-Monitoring Considerations for Baseball Pitchers. *J Athl Train.* 2020;55(9):911-917. doi:10.4085/1062-6050-0511-19
3. Shanley E, Bailey L, Sandago MP, Pinkerton A, Singleton SB, Thigpen CA. The use of a pitch count estimator to calculate exposure in collegiate baseball pitchers. *Phys Ther Sport.* 2015;16(4):344-348. doi:10.1016/j.ptsp.2015.02.002
4. Araujo A, Julious S, Senn S. Understanding Variation in Sets of N-of-1 Trials. *PLoS One.* 2016;11(12):e0167167. doi:10.1371/journal.pone.0167167
5. Halson SL. Monitoring Training Load to Understand Fatigue in Athletes. *Sports Med.* 2014;44(S2):139-147. doi:10.1007/s40279-014-0253-z
6. Birfer R, Sonne MW, Holmes MW. Manifestations of muscle fatigue in baseball pitchers: a systematic review. *PeerJ.* 2019;7:e7390. doi:10.7717/peerj.7390
7. Vieira R, McDonald S, Araújo-Soares V, Sniehotta FF, Henderson R. Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies. *Health Psychology Review.* 2017;11(3):222-234. doi:10.1080/17437199.2017.1343680
8. McDonald S, Vieira R, Johnston DW. Analysing N-of-1 observational data in health psychology and behavioural medicine: a 10-step SPSS tutorial for beginners. *Health Psychol Behav Med.* 8(1):32-54. doi:10.1080/21642850.2019.1711096
9. Senn S. The analysis of continuous data from n-of-1 trials using paired cycles: a simple tutorial. *Trials.* 2024;25:128. doi:10.1186/s13063-024-07964-7

10. Daza EJ. Causal Analysis of Self-tracked Time Series Data Using a Counterfactual Framework for N-of-1 Trials*. *Methods Inf Med*. 2018;57(Suppl 1):e10-e21. doi:10.3414/ME16-02-0044
11. Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. Published online May 14, 2015. doi:10.1136/bmj.h1738
12. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Published online October 4, 2000:1.9-4. doi:10.32614/CRAN.package.mgcv
13. Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*. 2019;11(3):e1460. doi:10.1002/wics.1460
14. Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA. A linear exponent AR(1) family of correlation structures. *Stat Med*. 2010;29(17):1825-1838. doi:10.1002/sim.3928
15. Noyez L. Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg*. 2009;9(3):494-499. doi:10.1510/icvts.2009.204768
16. Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. *British Journal of Surgery*. 2008;95(7):925-929. doi:10.1002/bjs.6056
17. Page ES. Continuous Inspection Schemes. *Biometrika*. 1954;41(1/2):100-115. doi:10.2307/2333009
18. Liashchynskyi P, Liashchynskyi P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. arXiv.org. December 12, 2019. Accessed March 31, 2026. <https://arxiv.org/abs/1912.06059v1>
19. Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, Robertson SJ. Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an

- Individualized Approach. *Int J Sports Physiol Perform.* 2017;12(2):230-234. doi:10.1123/ijsp.2015-0791
20. Cornell DJ, Paxson JL, Caplinger RA, Seligman JR, Davis NA, Ebersole KT. Resting Heart Rate Variability Among Professional Baseball Starting Pitchers. *Journal of Strength and Conditioning Research.* 2017;31(3):575-581. doi:10.1519/JSC.0000000000001538
21. Grantham WJ, Byram IR, Meadows MC, Ahmad CS. The Impact of Fatigue on the Kinematics of Collegiate Baseball Pitchers. *Orthop J Sports Med.* 2014;2(6):2325967114537032. doi:10.1177/2325967114537032
22. Erickson BJ, Chalmers PN, D'Angelo J, Ma K, Romeo AA. Performance and Return to Sport After Latissimus Dorsi and Teres Major Tears Among Professional Baseball Pitchers. *Am J Sports Med.* 2019;47(5):1090-1095. doi:10.1177/0363546519829086
23. Okoroha KR, Meldau JE, Lizzio VA, et al. Effect of Fatigue on Medial Elbow Torque in Baseball Pitchers: A Simulated Game Analysis. *Am J Sports Med.* 2018;46(10):2509-2513. doi:10.1177/0363546518782451

CHAPTER VI: SUMMARY OF RESEARCH

The research presented in this dissertation examined rolling pitch count associations with fastball performance metrics and pitching biomechanics in collegiate baseball pitchers using individualized N-of-1 modeling approaches, compared those individualized models with traditional group-based alternatives, and extended individualized model outputs into a practical monitoring framework. Collectively, this work was designed to evaluate whether pitcher-specific modeling could provide a more informative framework for understanding workload-response patterns in collegiate baseball pitchers across performance, biomechanics, and applied surveillance contexts. The overall objective of this dissertation was to apply individualized N-of-1 generalized additive mixed models to investigate how rolling pitch count was associated with key pitching outcomes and to evaluate the practical value of these individualized models relative to traditional analytic approaches. The specific aims of this dissertation were as follows:

1. Quantify within-pitcher associations between rolling pitch count and fastball performance metrics, including velocity, spin rate, and vertical break.
2. Quantify within-pitcher associations between rolling pitch count and key pitching biomechanics.
3. Compare individualized N-of-1 models with traditional population-level mixed-effects models in explanatory performance and residual independence.
4. Apply cumulative-sum (CUSUM) monitoring procedures to individualized model outputs to detect early, systematic declines in performance across a season.

Together, these aims were designed to evaluate individualized workload-response patterns in collegiate baseball pitchers across performance, biomechanics, comparative modeling, and applied monitoring domains.

Table XIV: Publication plans for research presented in this dissertation

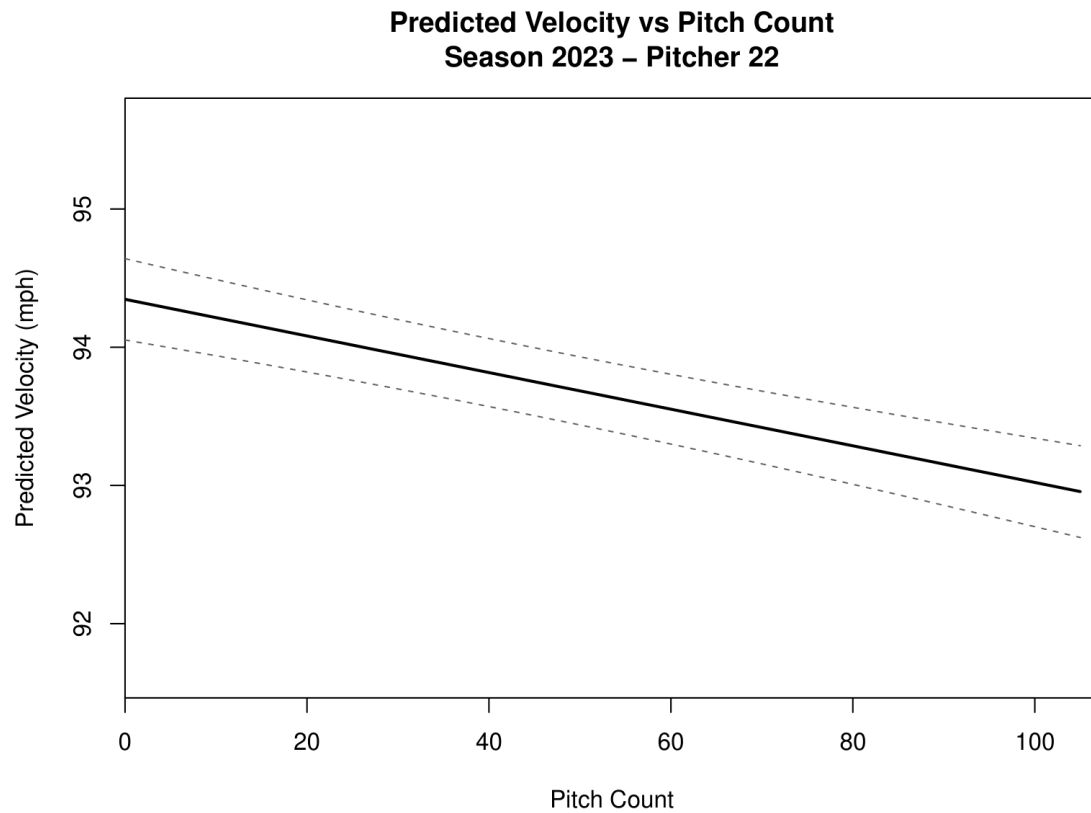
Chapter	Topic	Journal
II	N-of-1 Modeling of Rolling Pitch Count Associations with Fastball Performance Metrics in Collegiate Baseball Pitchers	<i>American Journal of Sports Medicine (AJSM)</i> **
III	N-of-1 Modeling of Rolling Fastball Pitch Count Associations with Pitching Biomechanics in Collegiate Baseball Pitchers	<i>Journal of Sports Biomechanics</i> **
IV	Comparative Evaluation of Group-Based and Individualized N-of-1 Models for Pitch Count Associations in Baseball Pitchers	TBD**
V	Practical Applications of Individualized Pitch Count Models for Monitoring Collegiate Baseball Pitchers	TBD

* Published

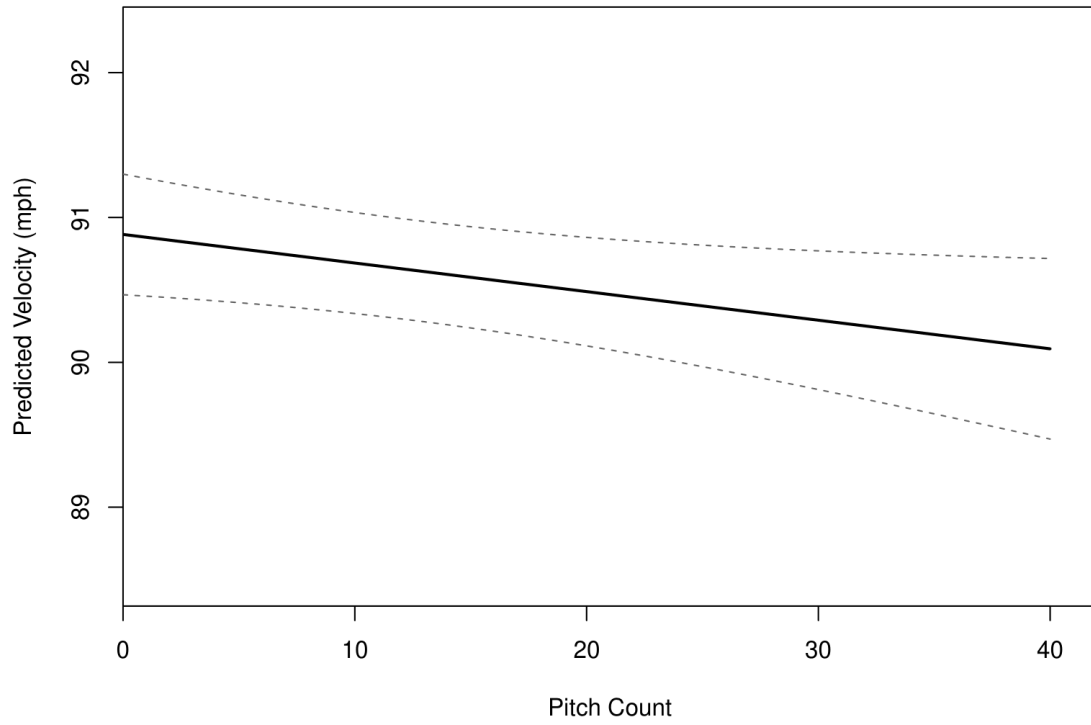
**Manuscript in preparation

Appendix A – Performance Models Prediction Curves

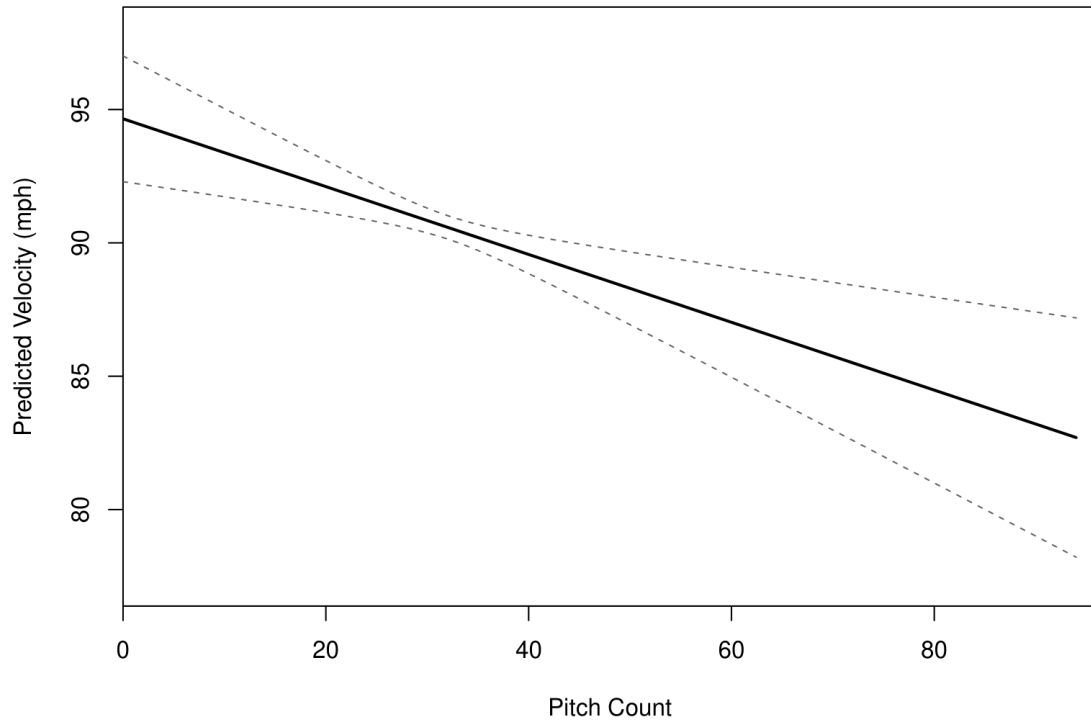
Velocity Prediction Curves



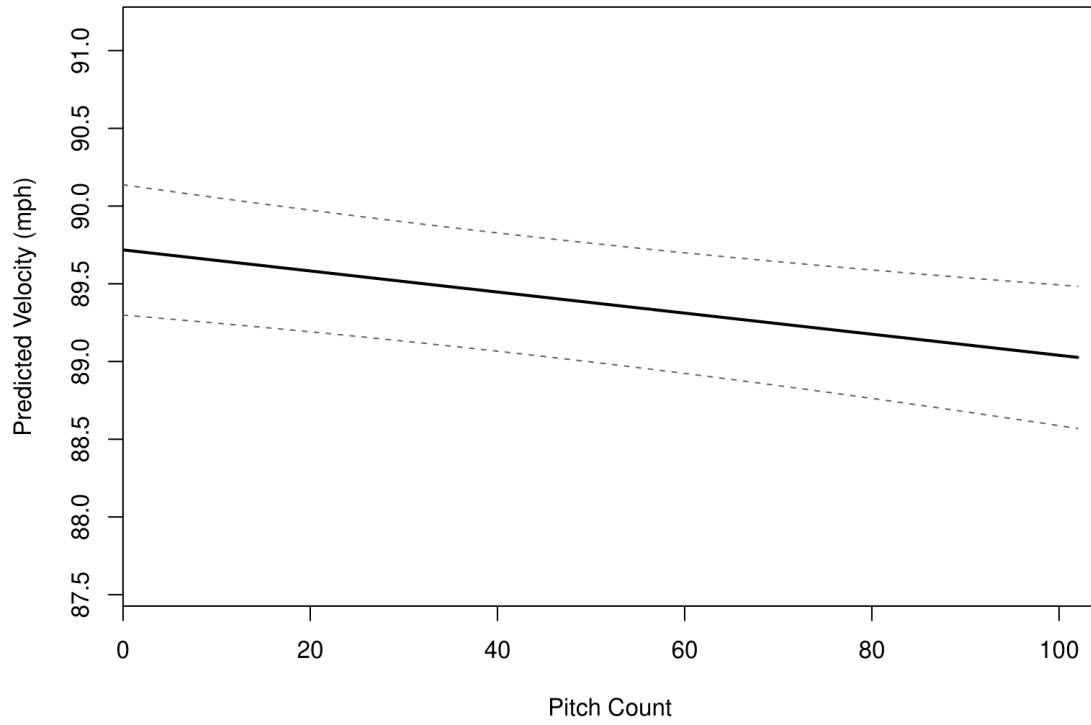
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 23



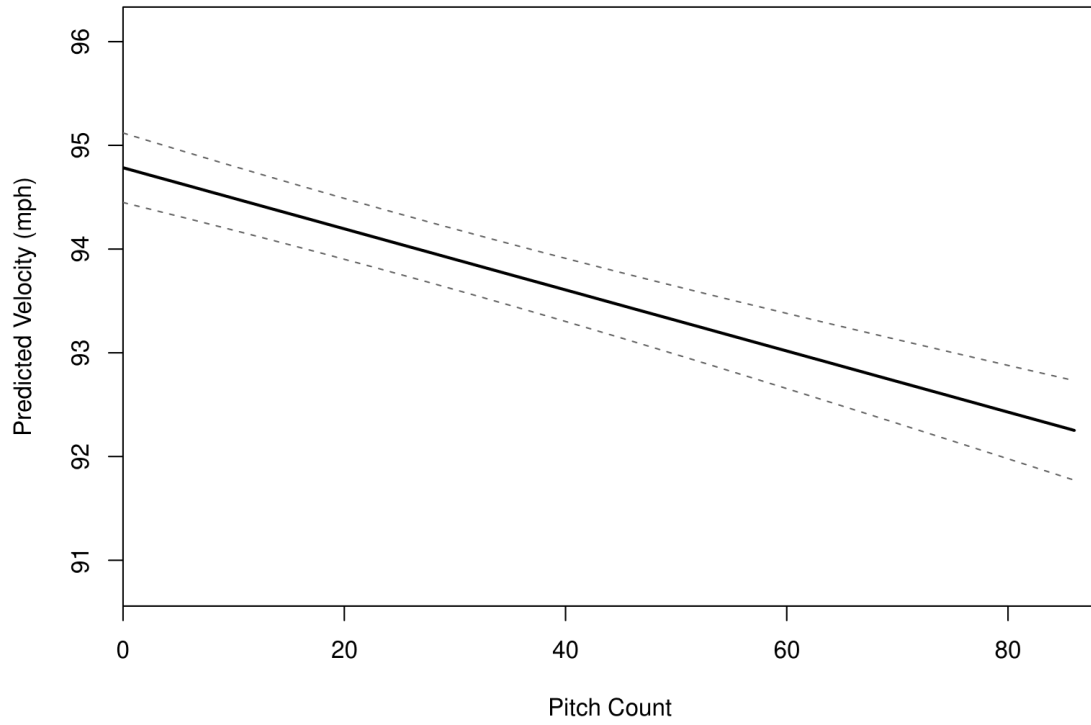
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 24



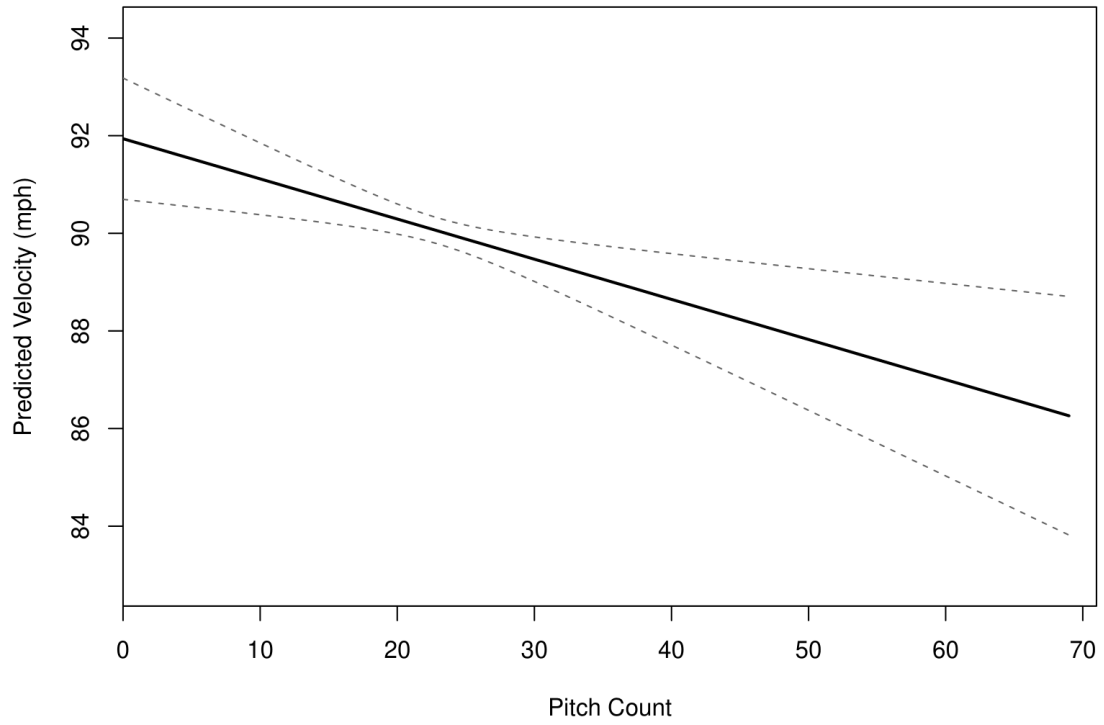
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 25



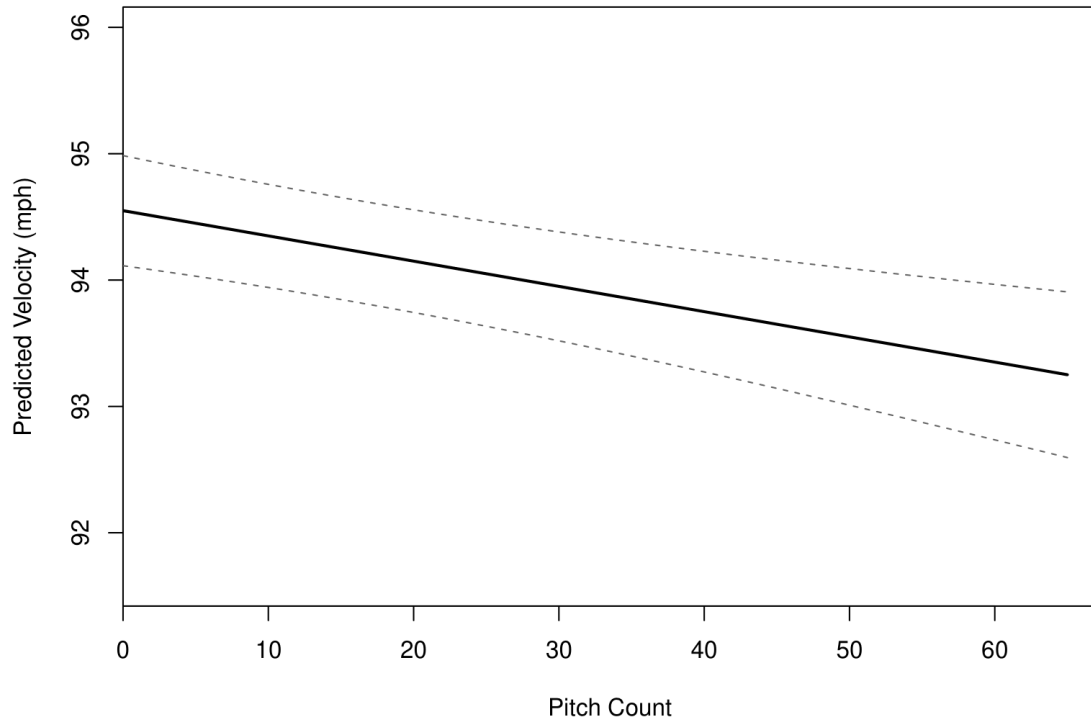
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 26



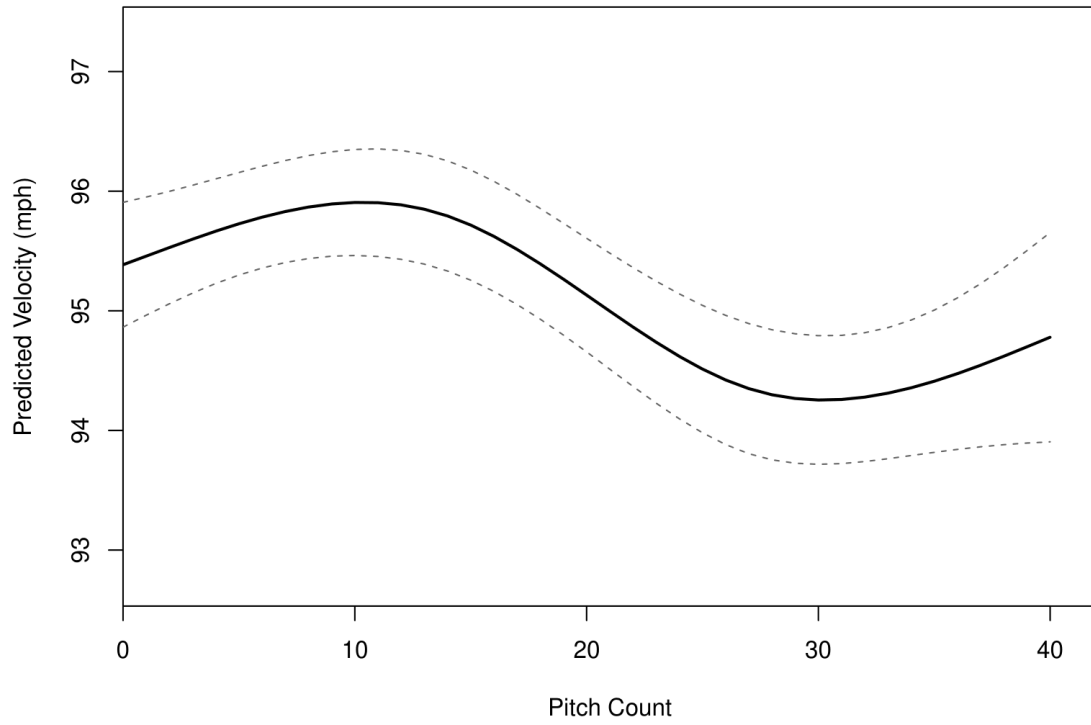
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 27



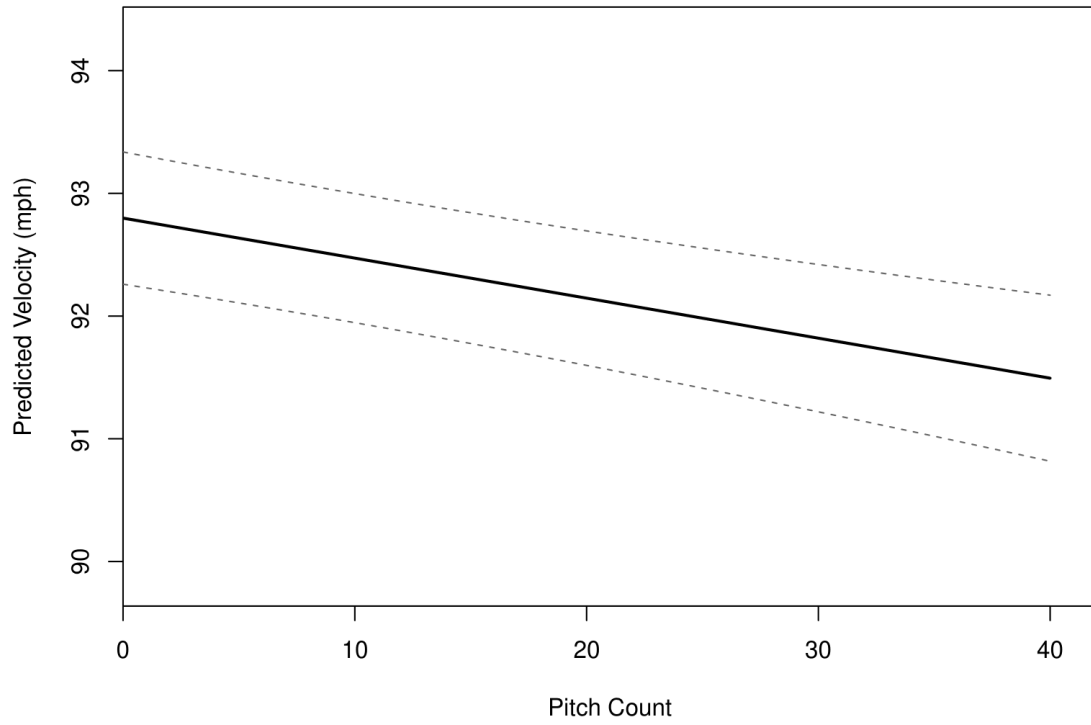
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 28



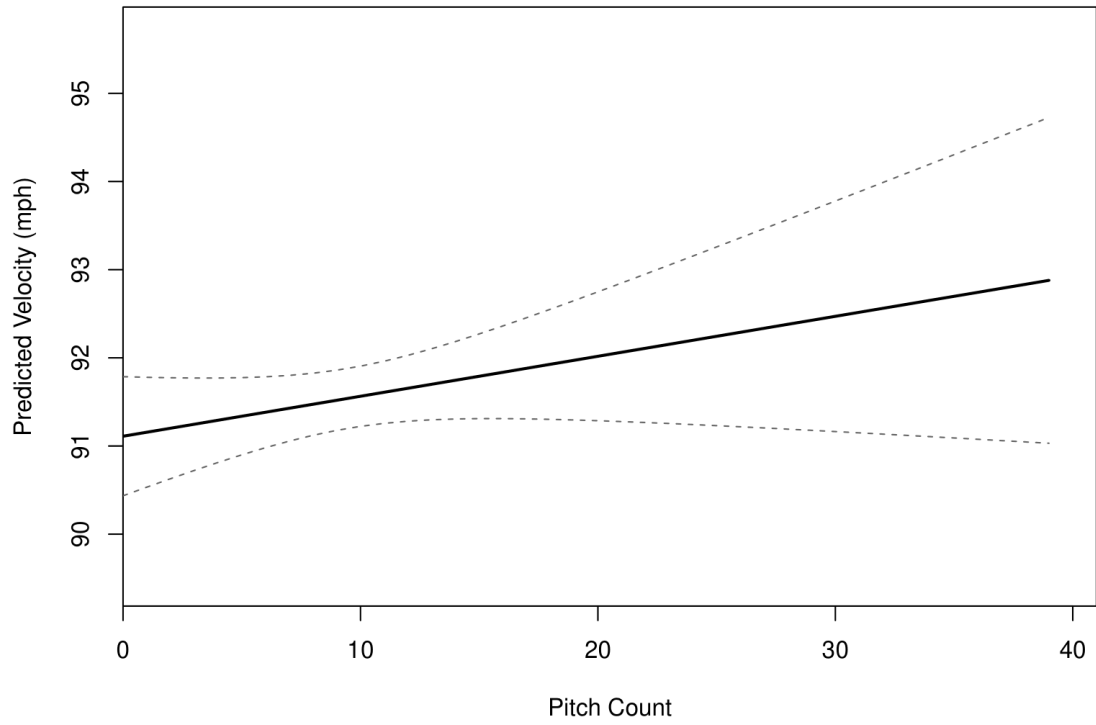
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 29



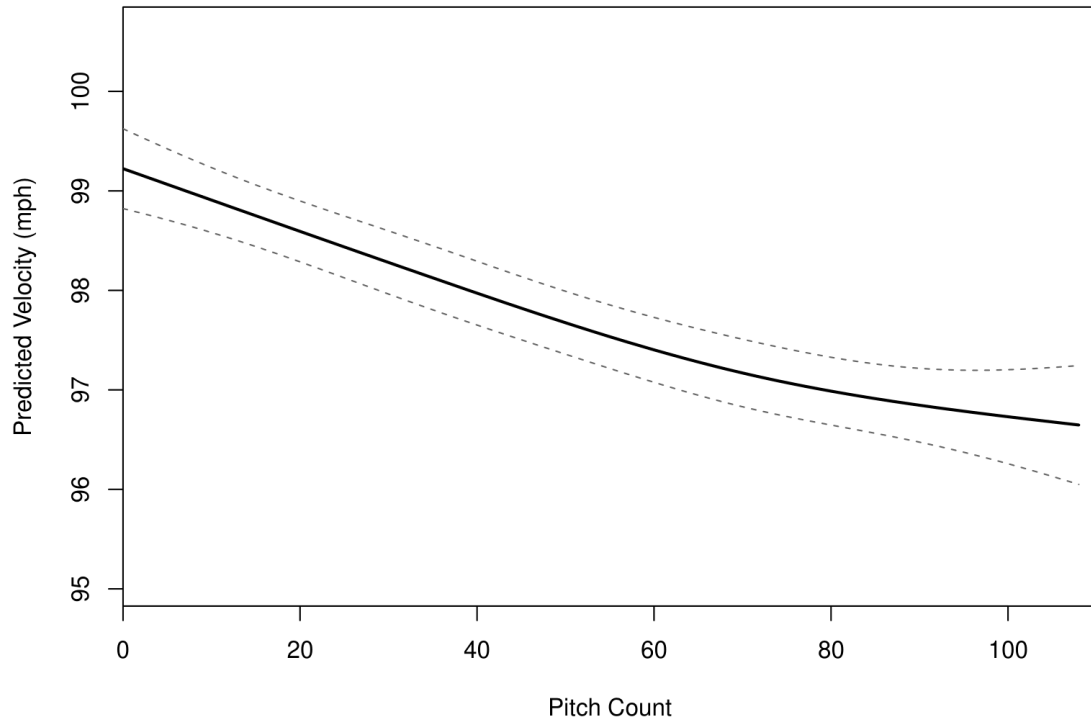
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 30



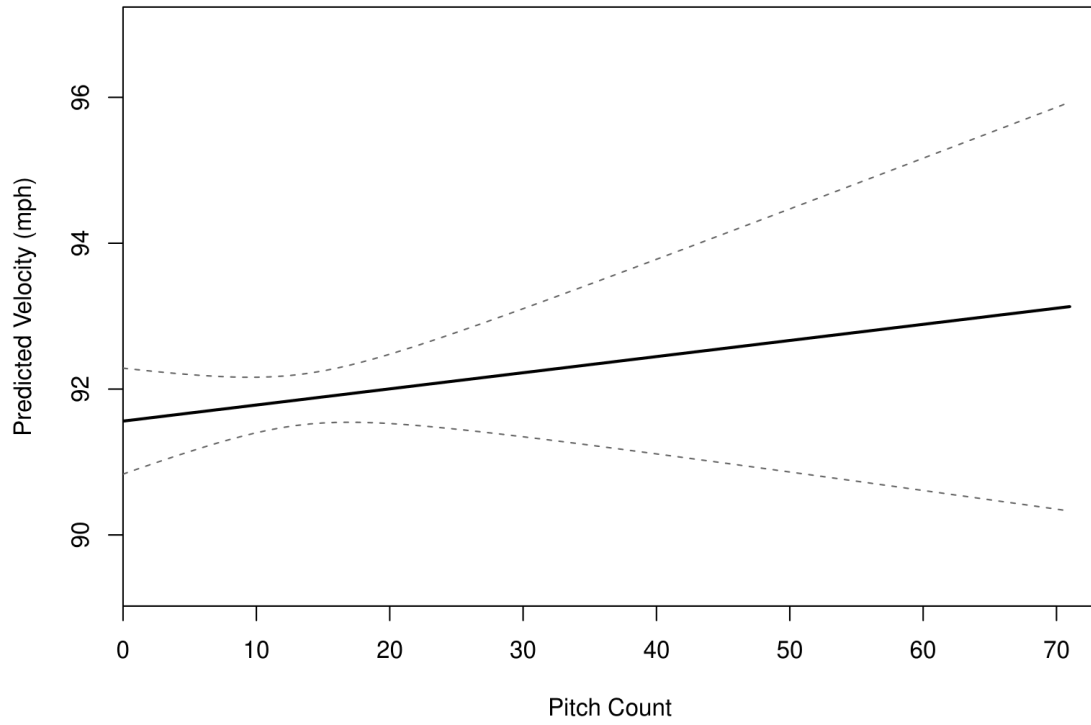
Predicted Velocity vs Pitch Count
Season 2023 – Pitcher 31



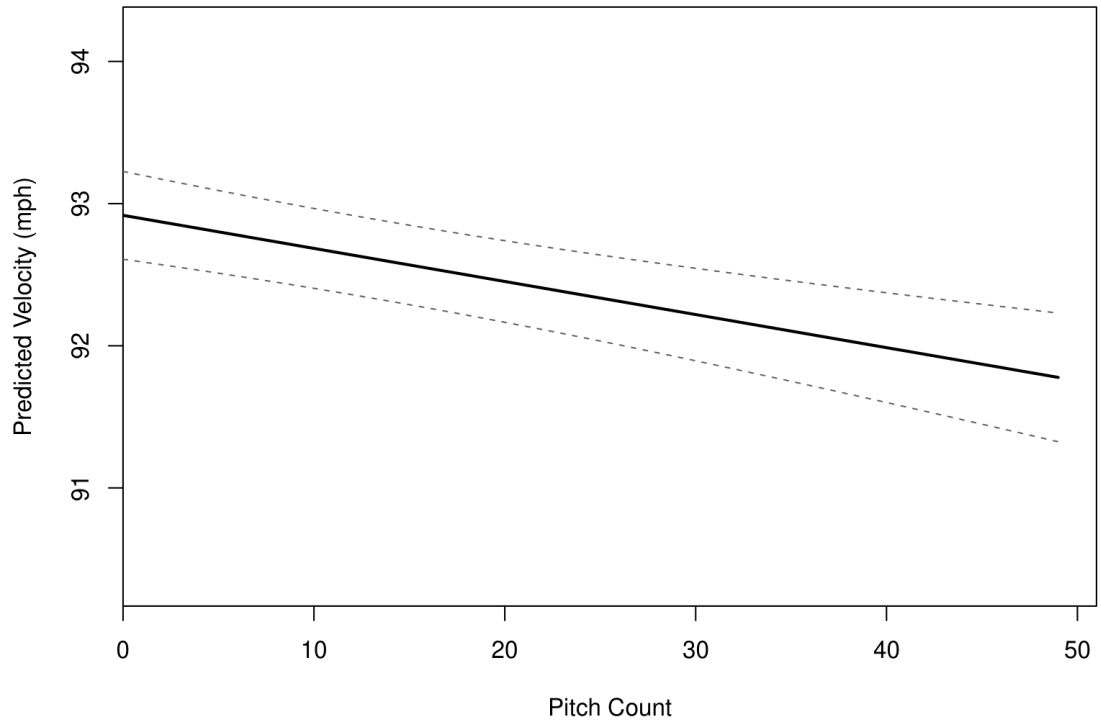
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 5



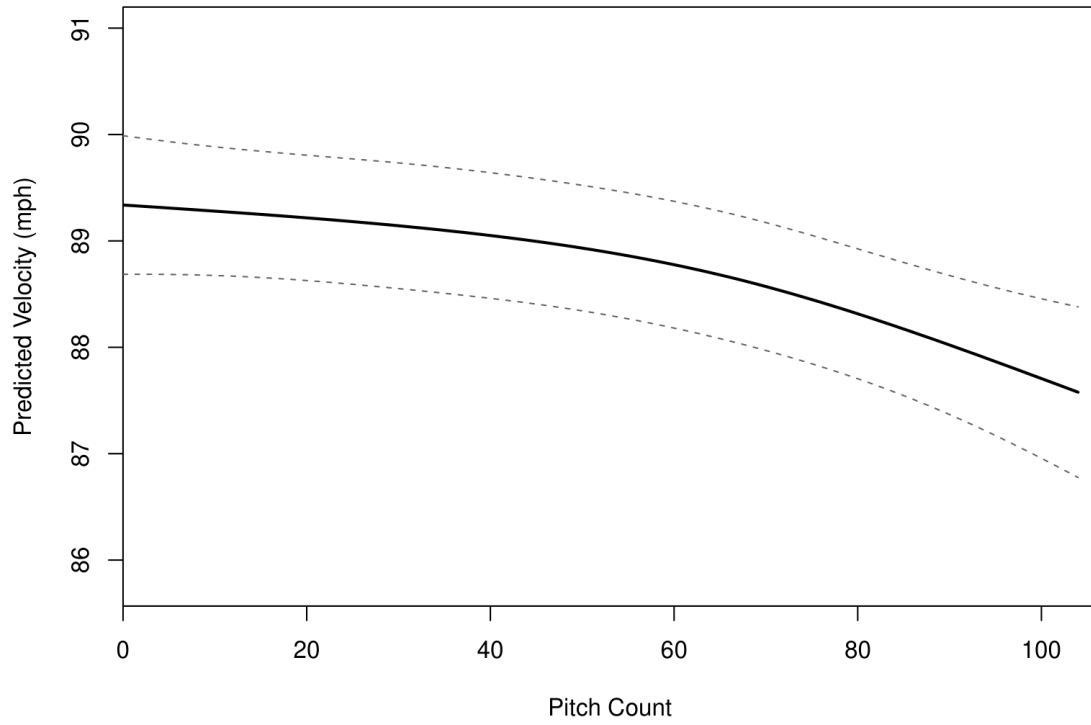
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 8



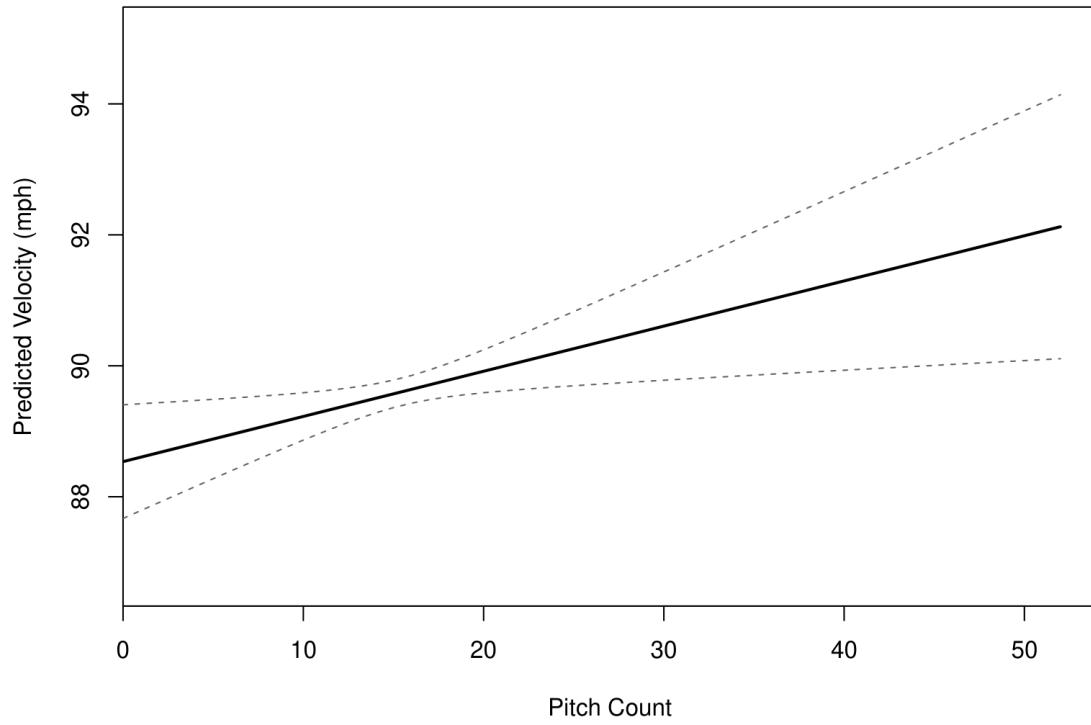
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 11



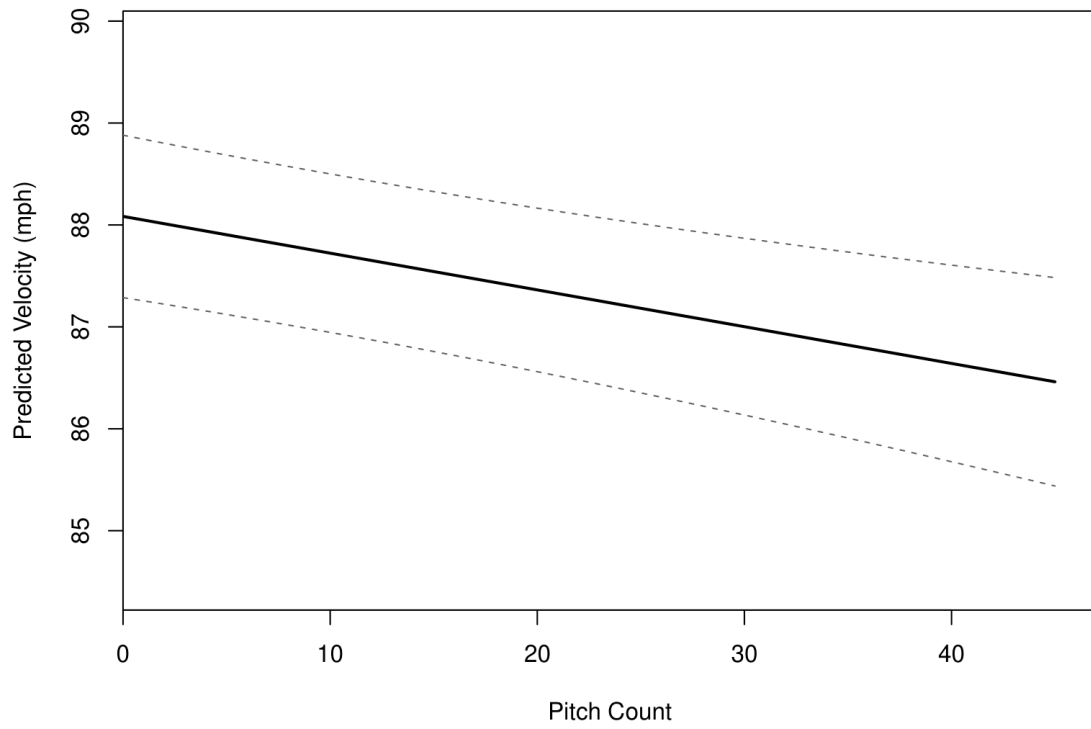
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 12



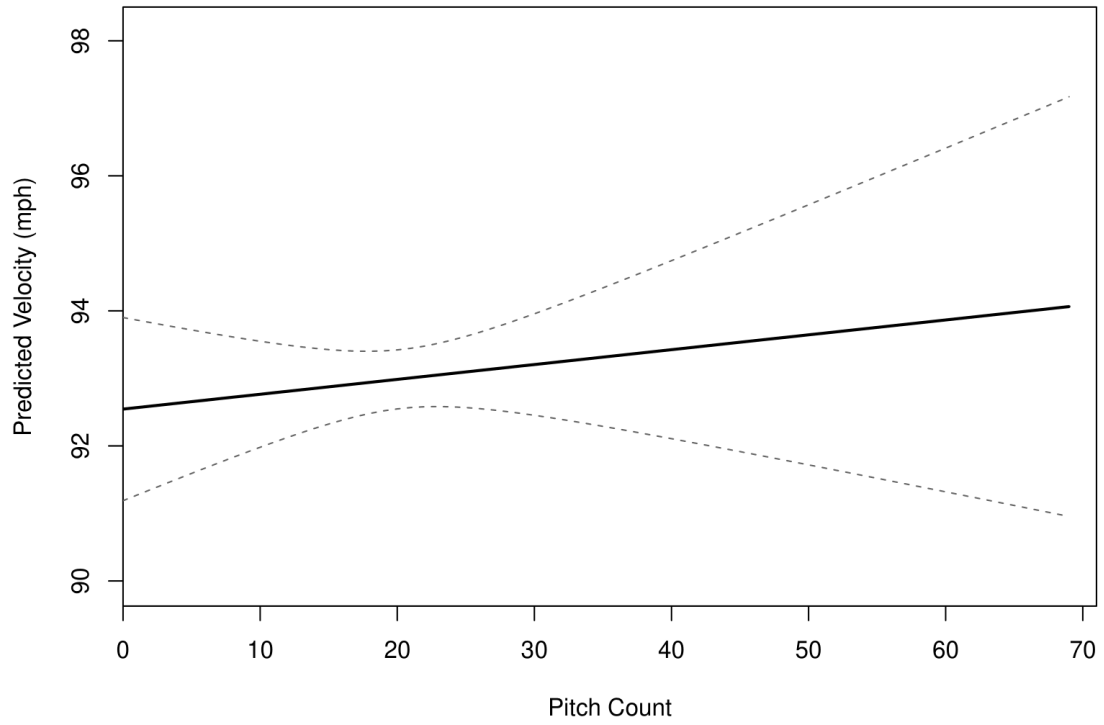
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 14



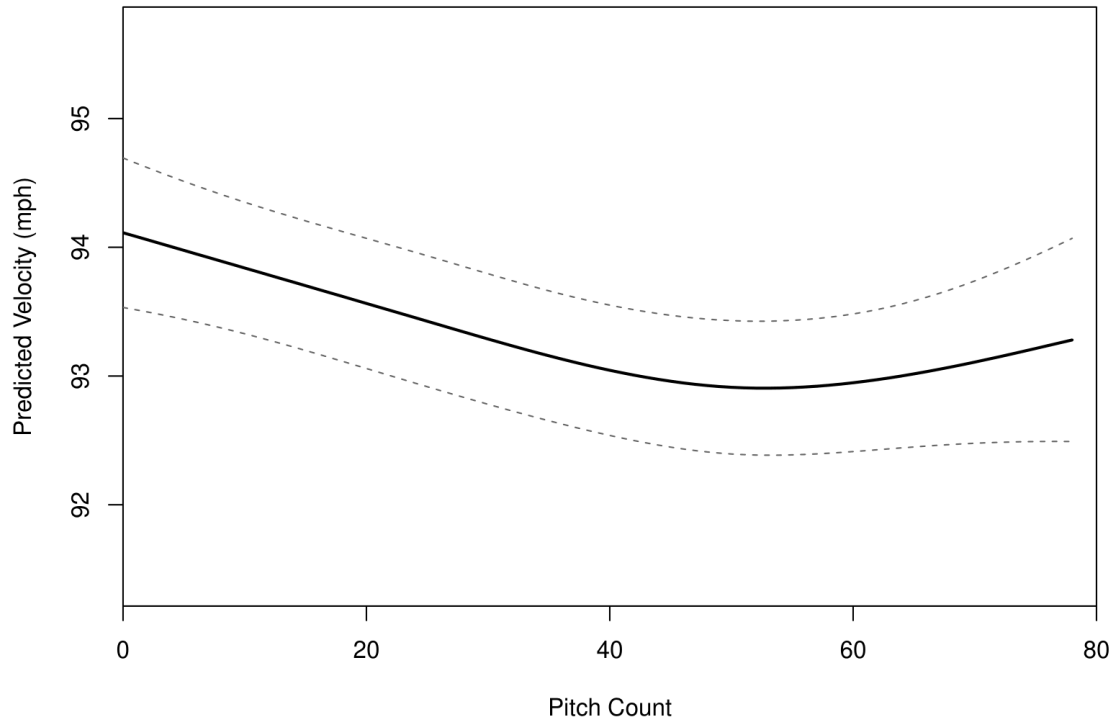
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 15



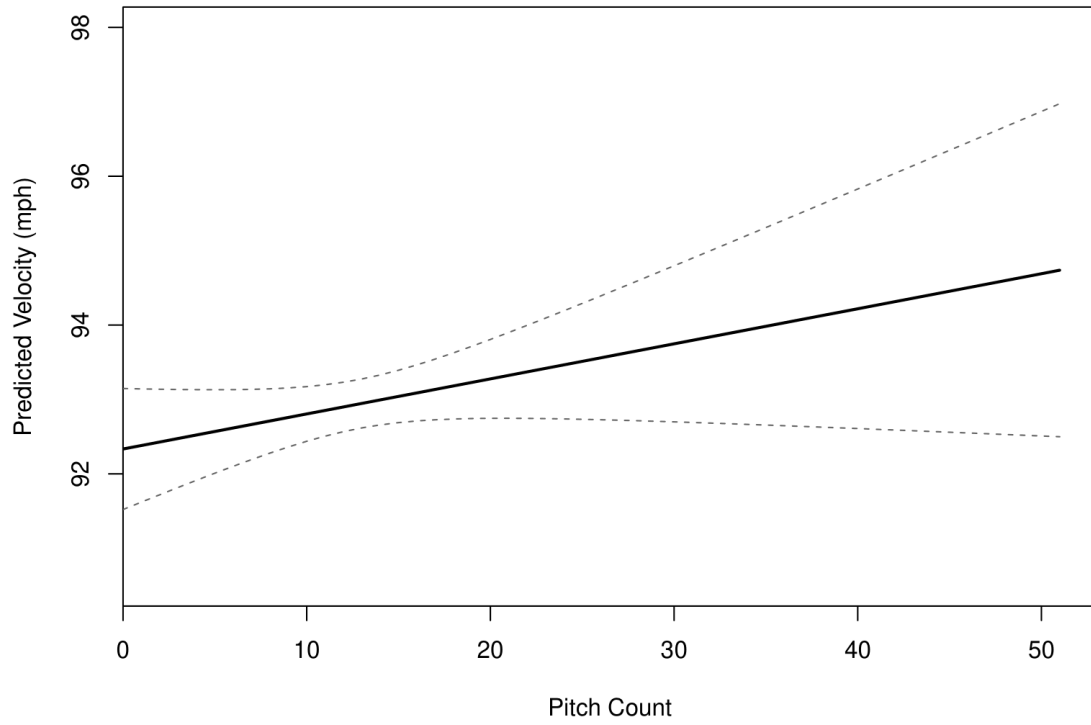
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 17



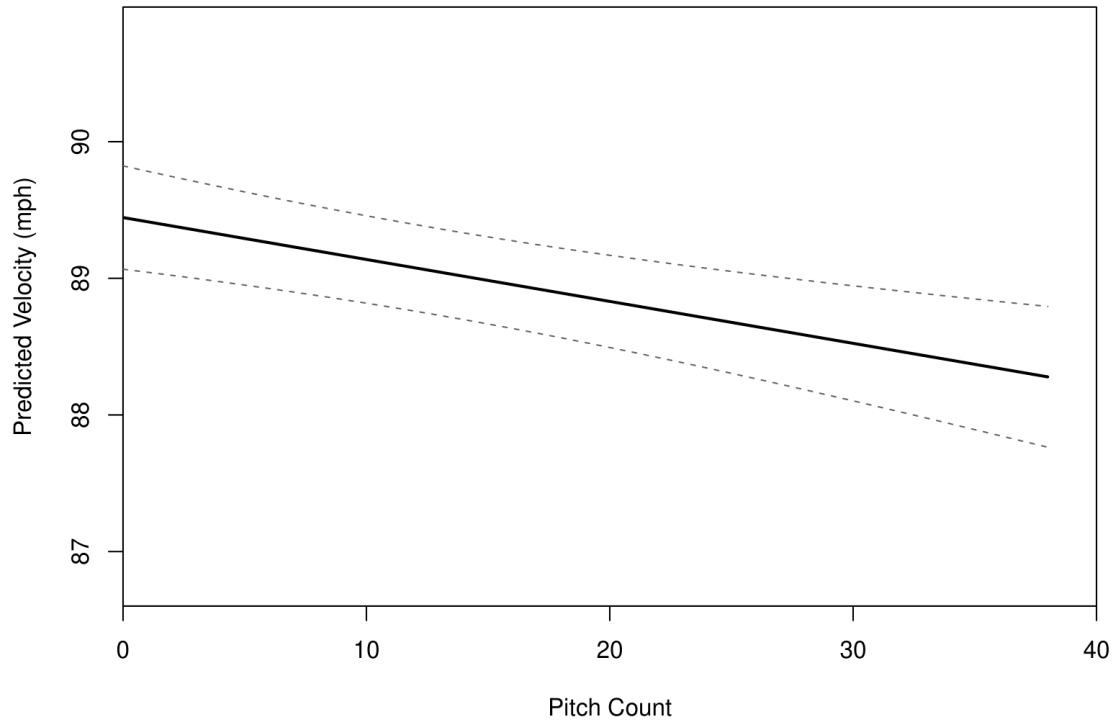
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 19



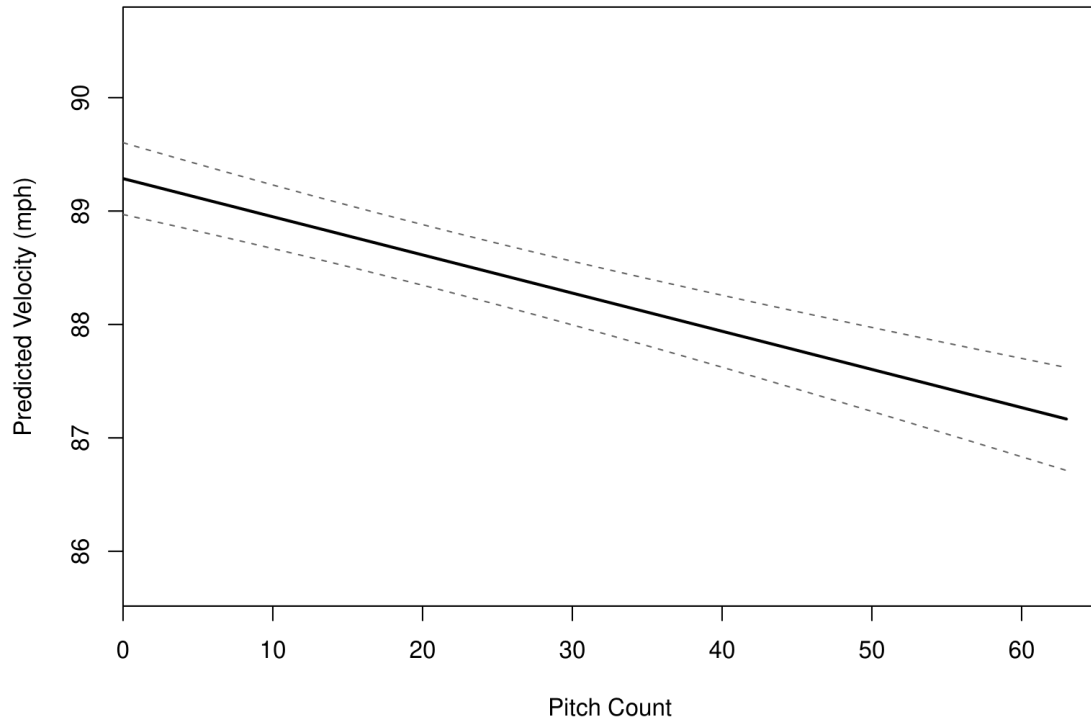
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 21



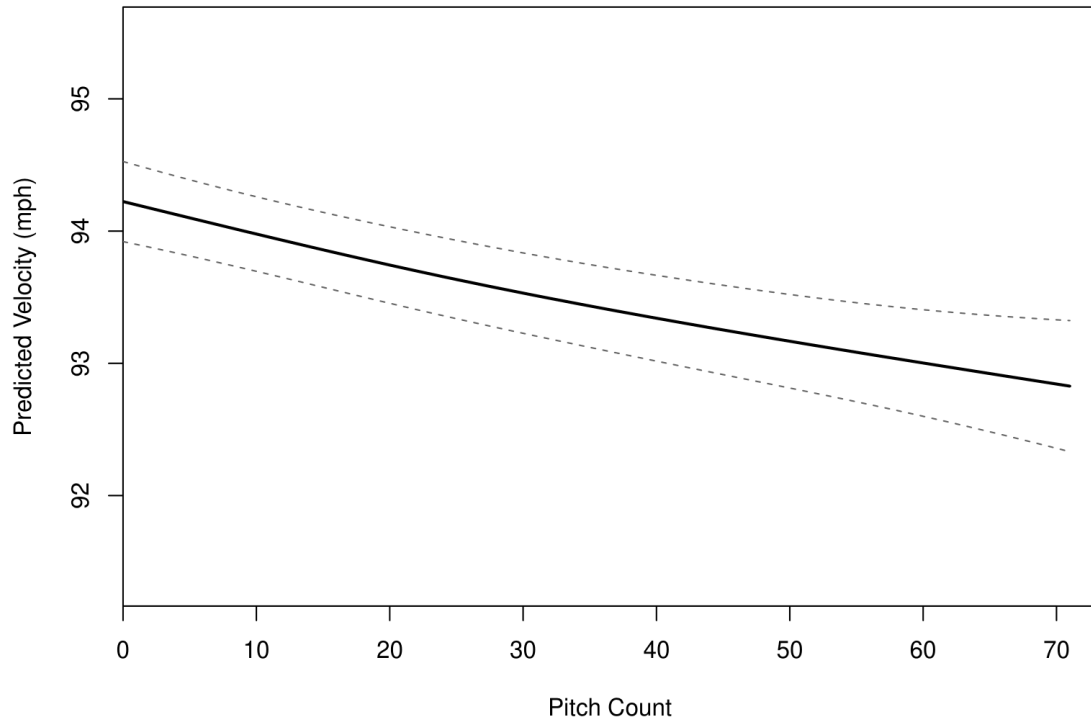
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 33



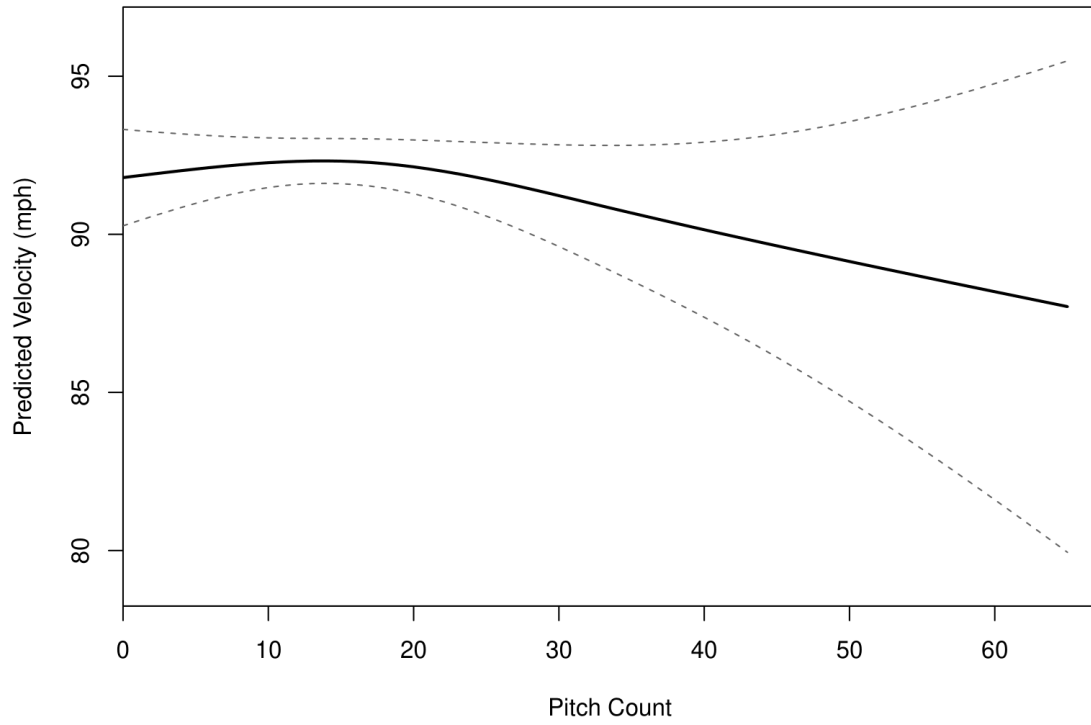
Predicted Velocity vs Pitch Count
Season 2024 – Pitcher 35



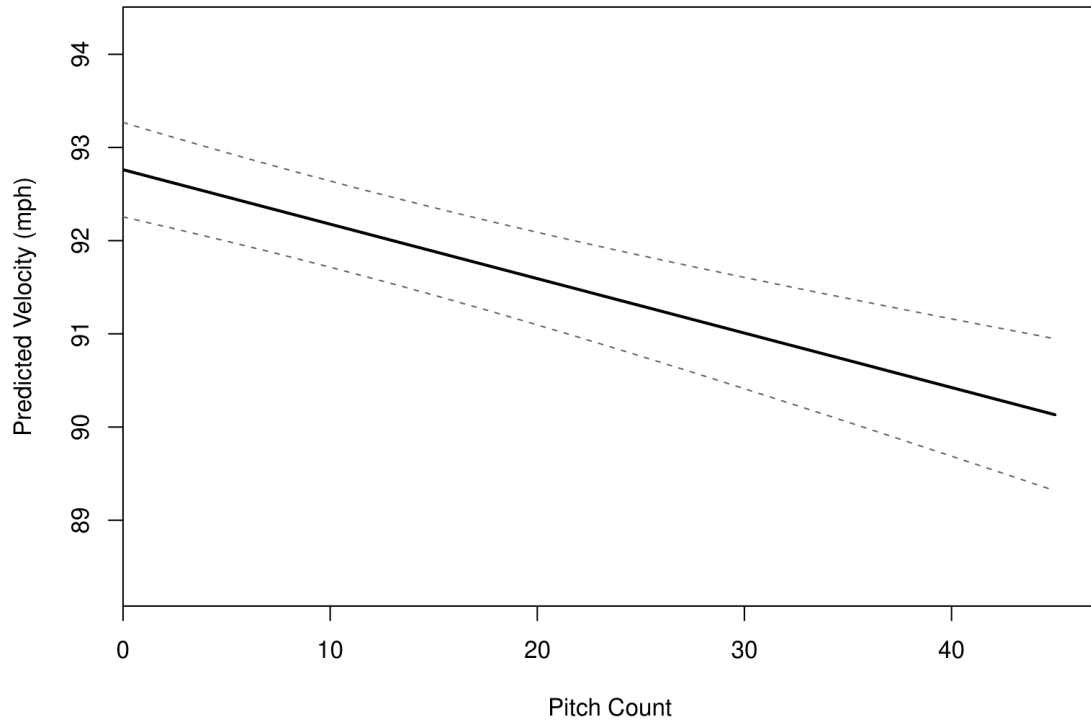
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 1



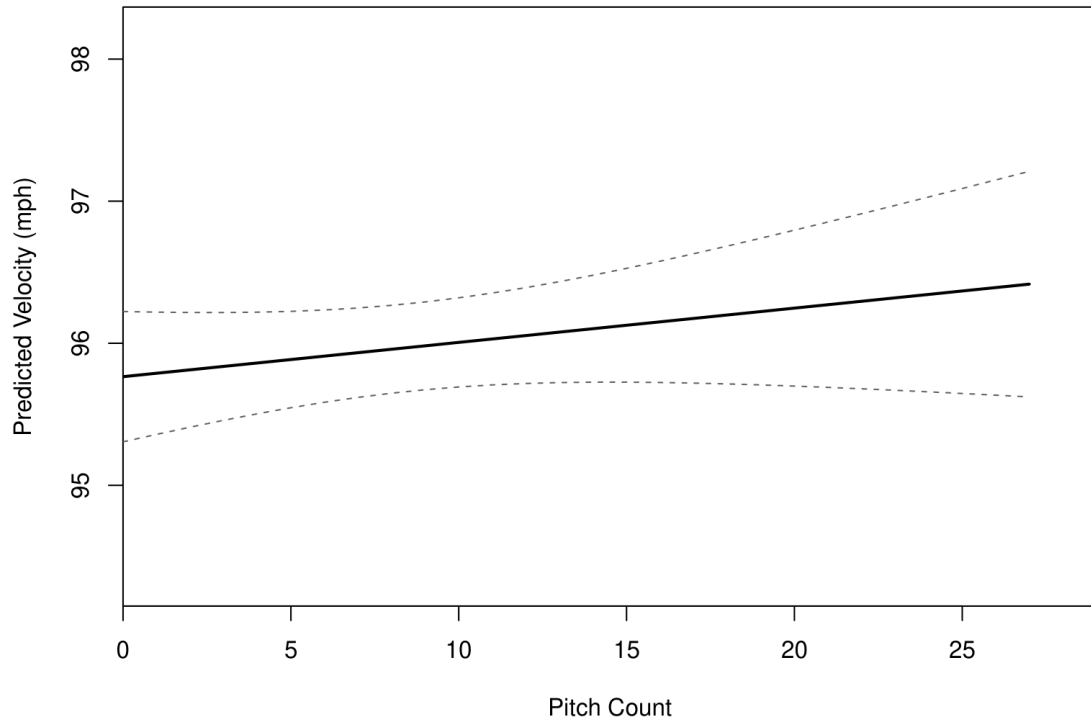
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 2



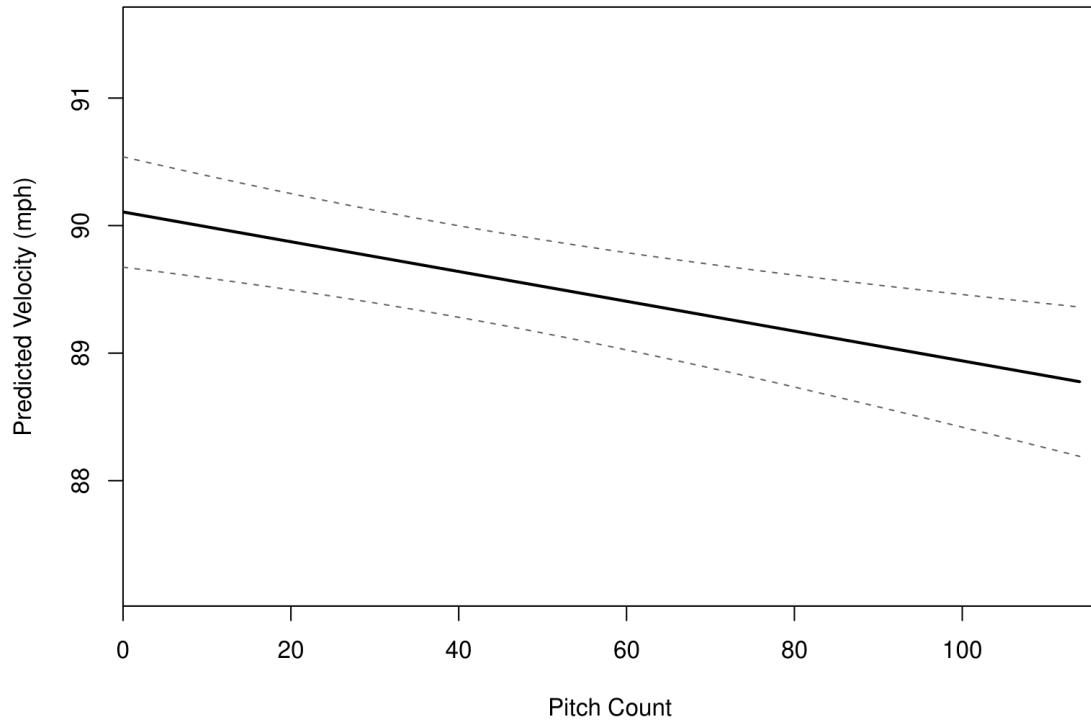
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 3



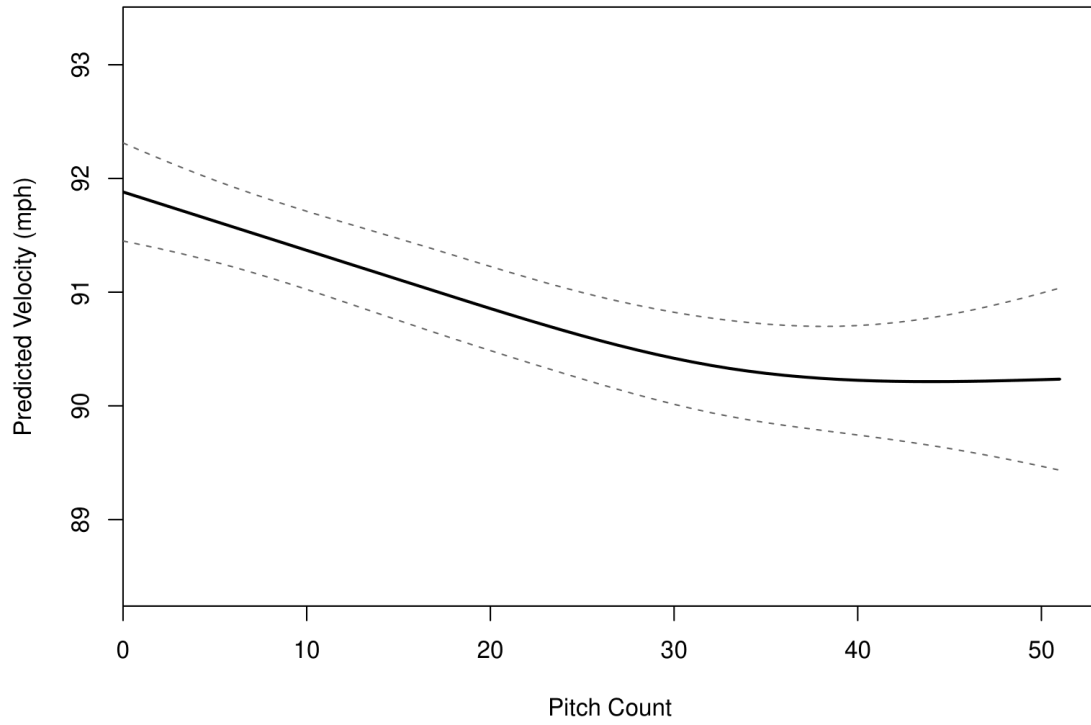
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 4



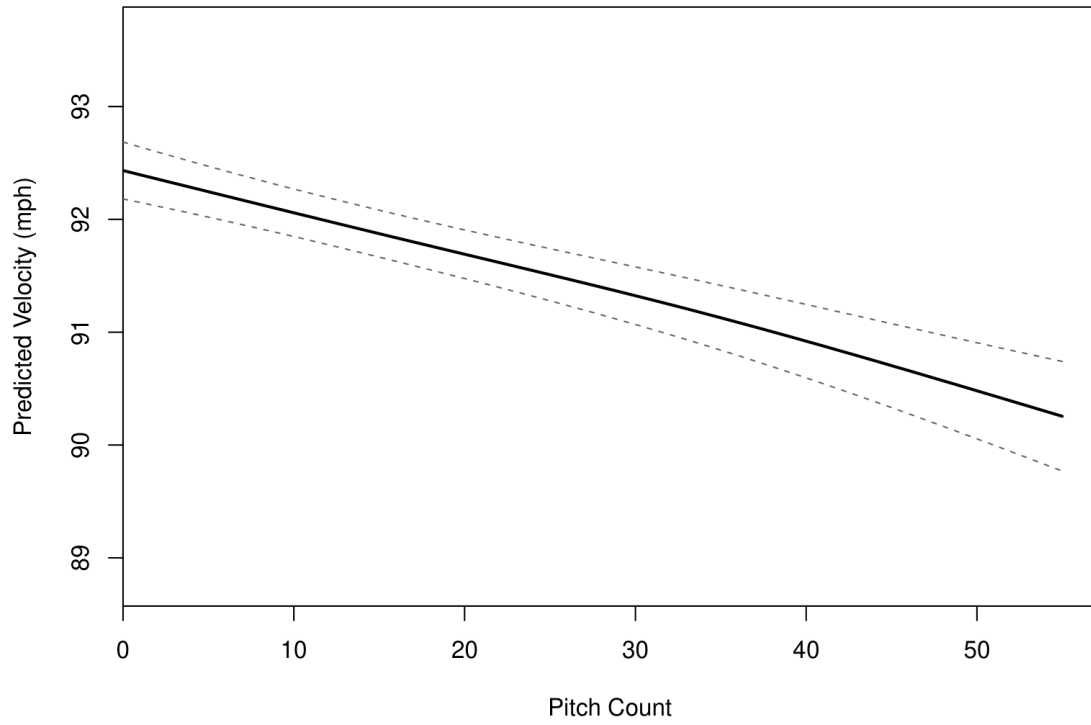
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 6



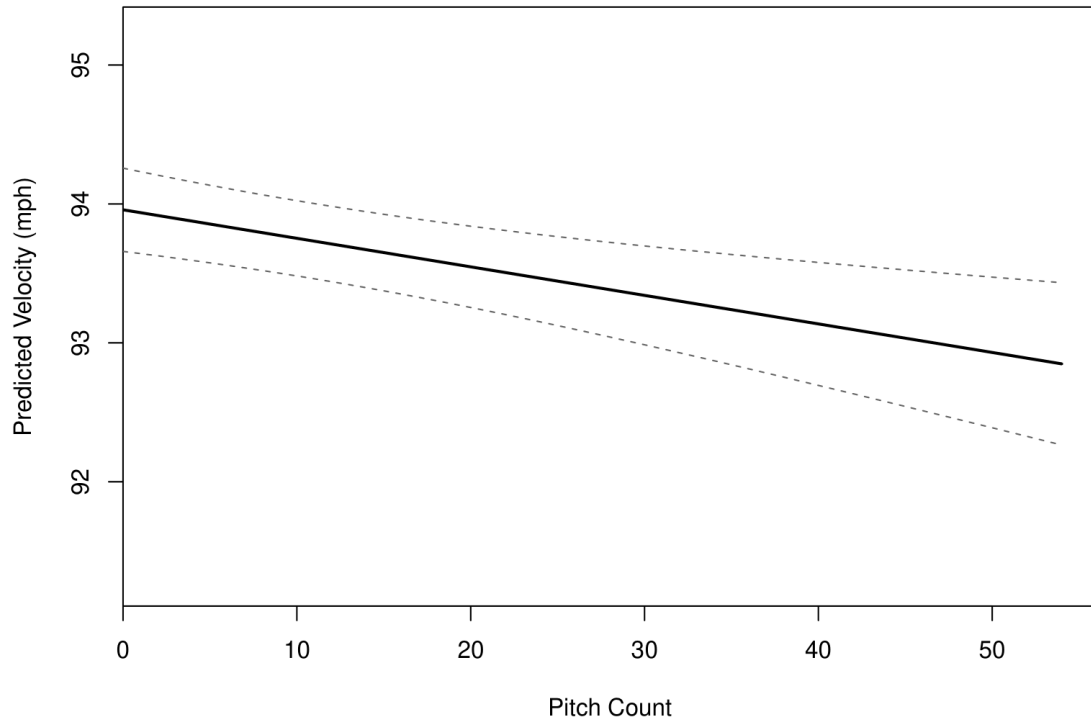
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 7



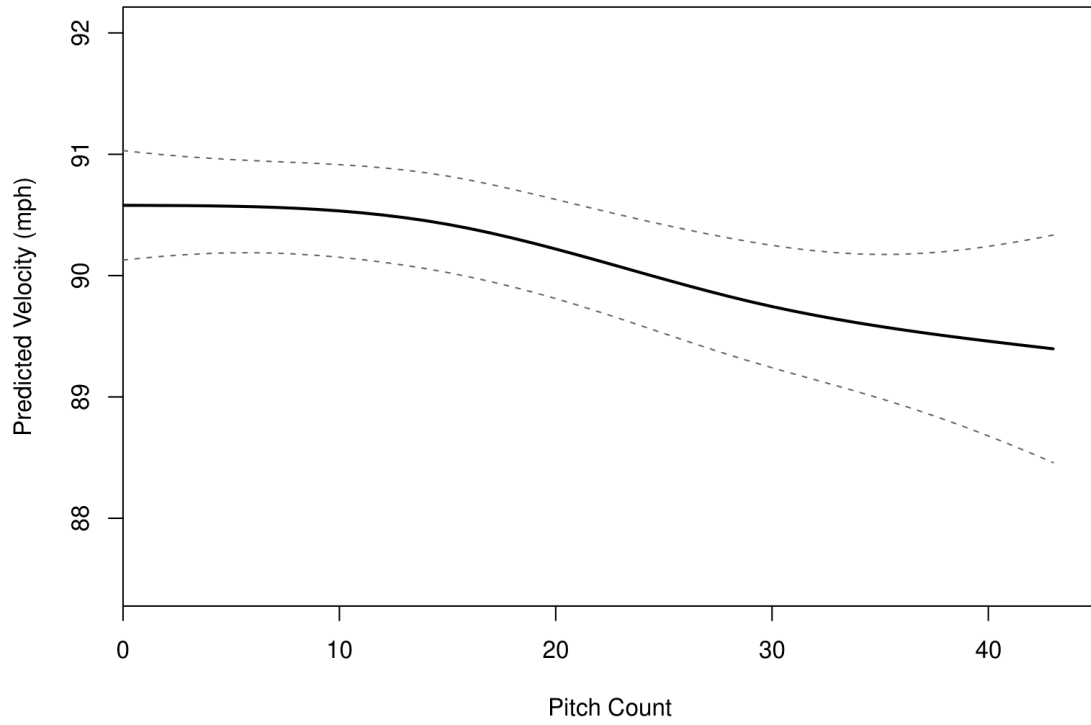
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 9



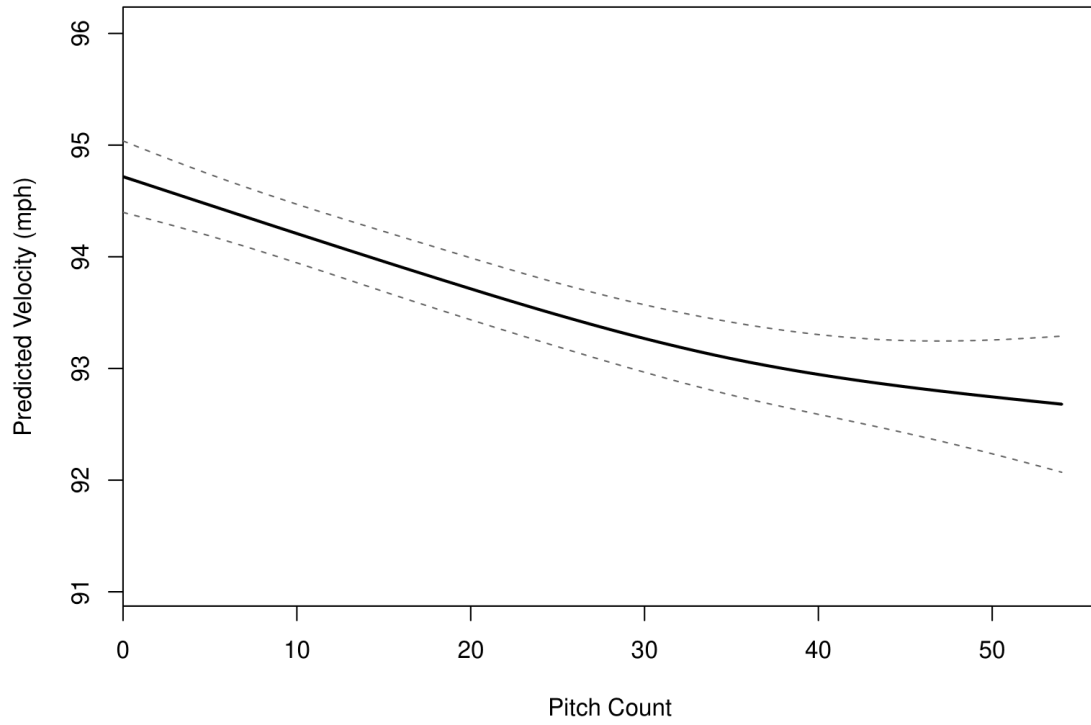
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 10



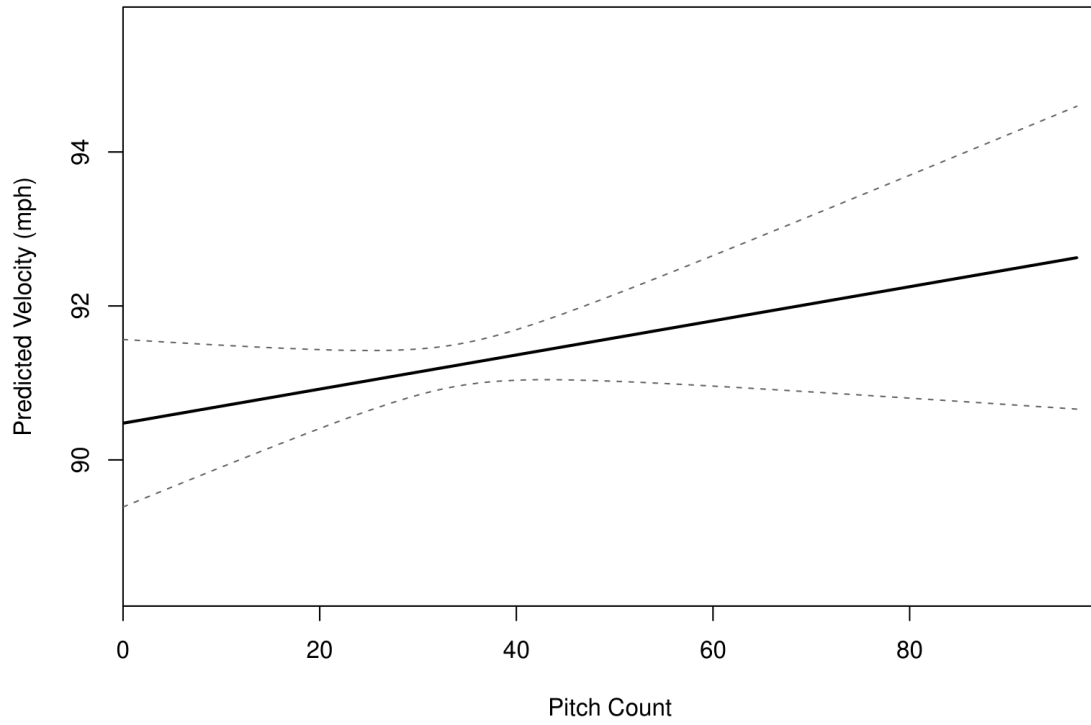
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 13



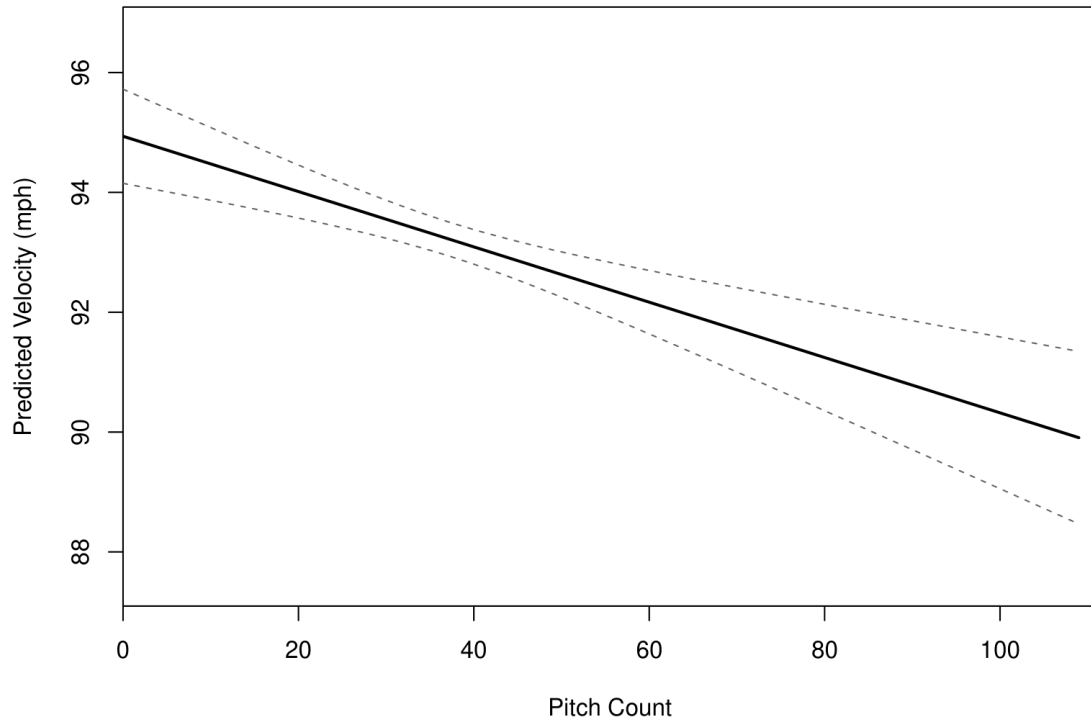
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 16



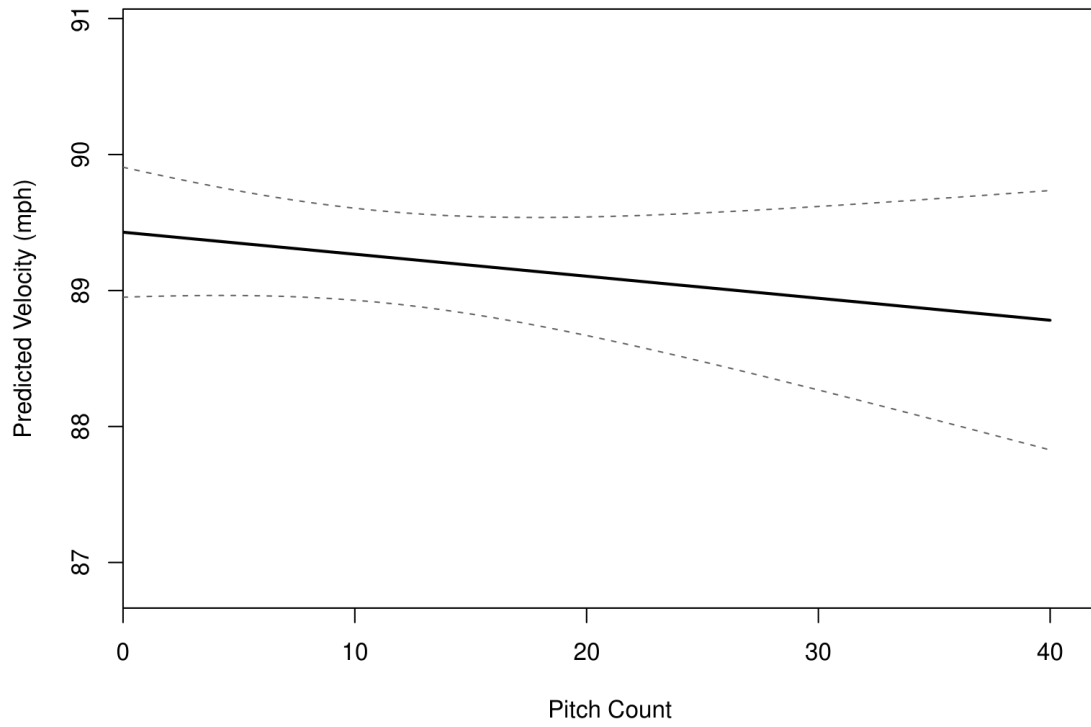
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 18



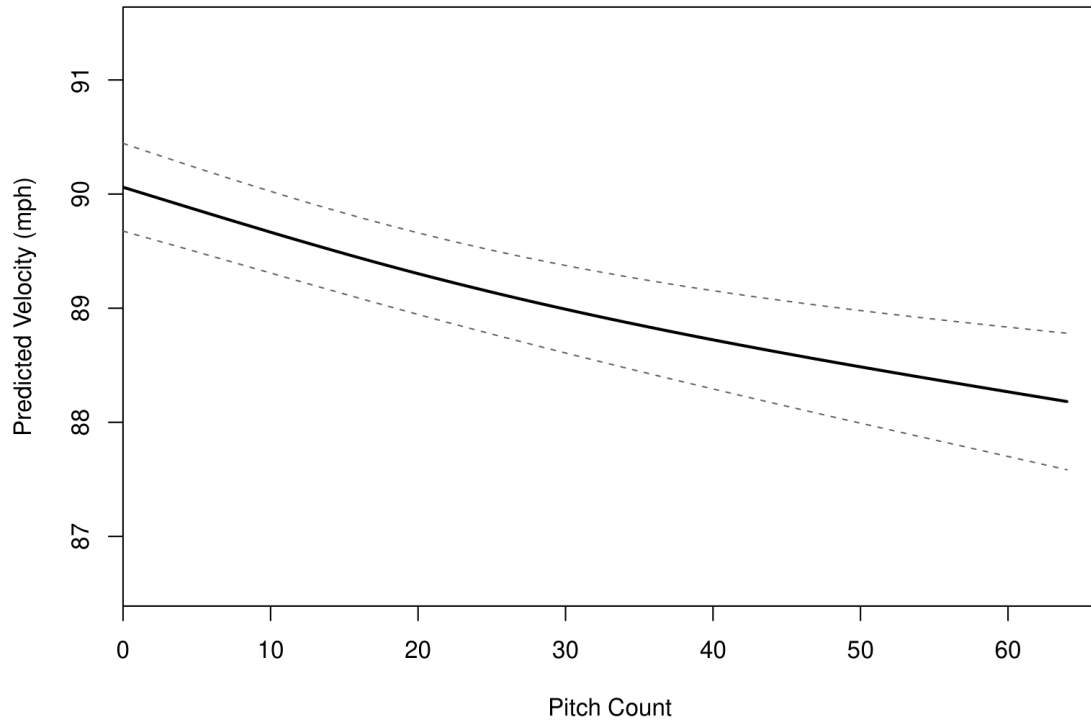
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 20



Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 32

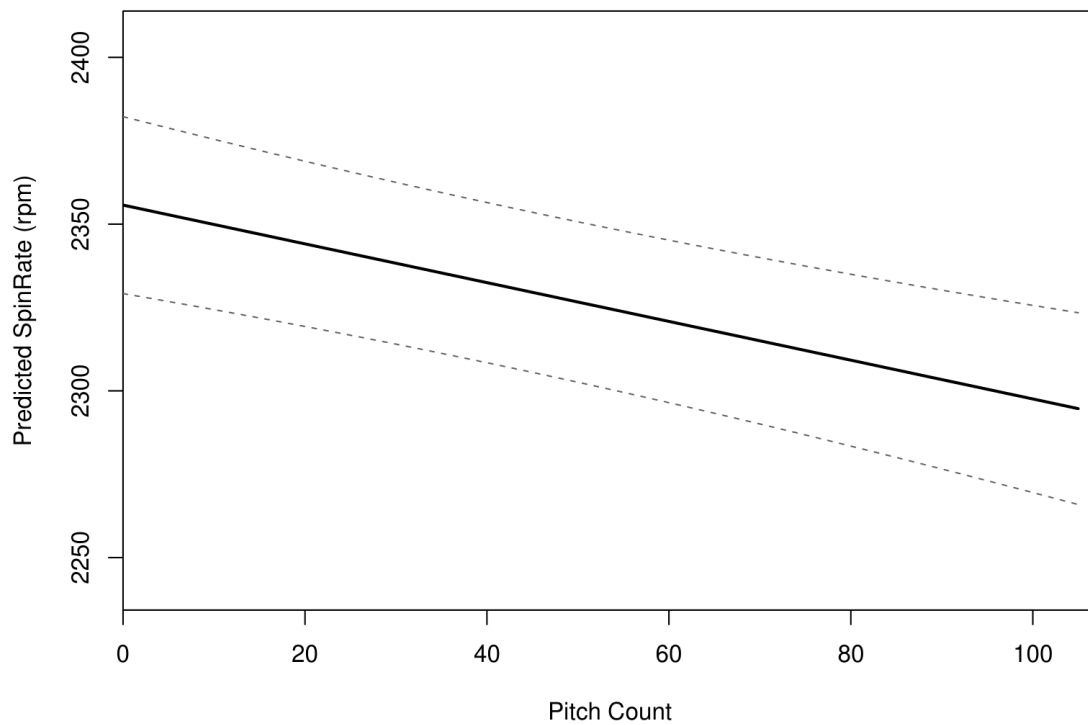


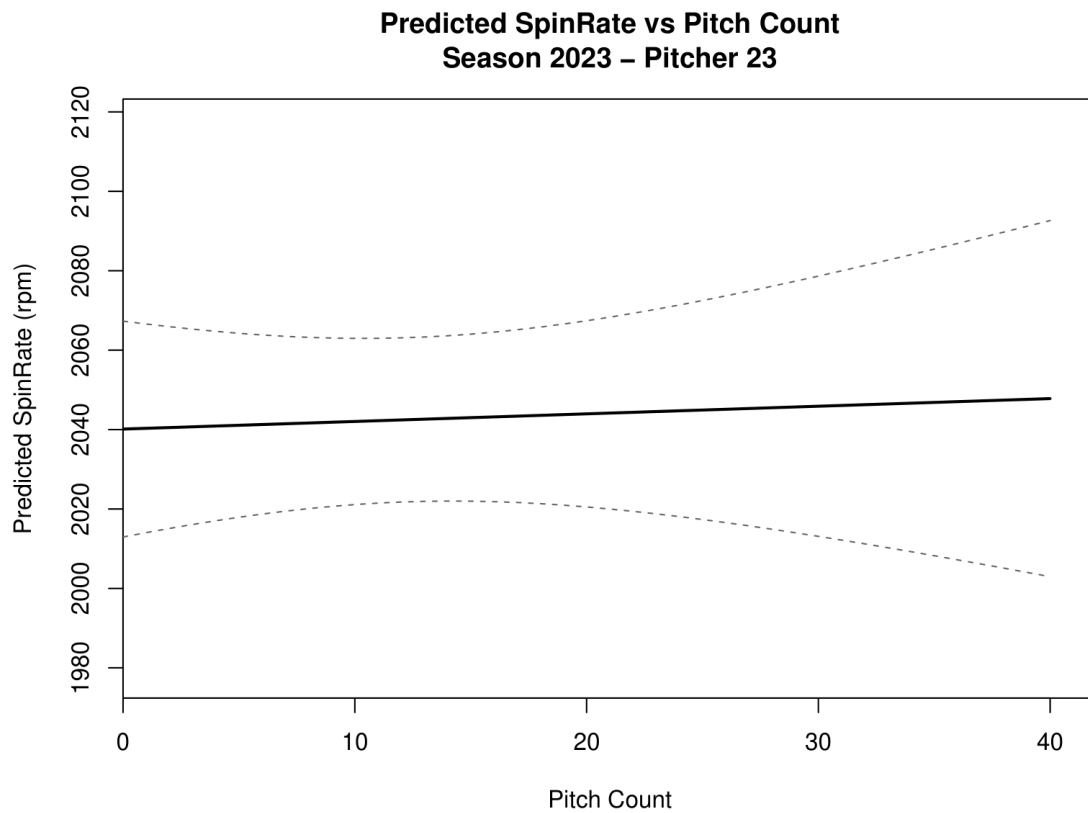
Predicted Velocity vs Pitch Count
Season 2025 – Pitcher 34



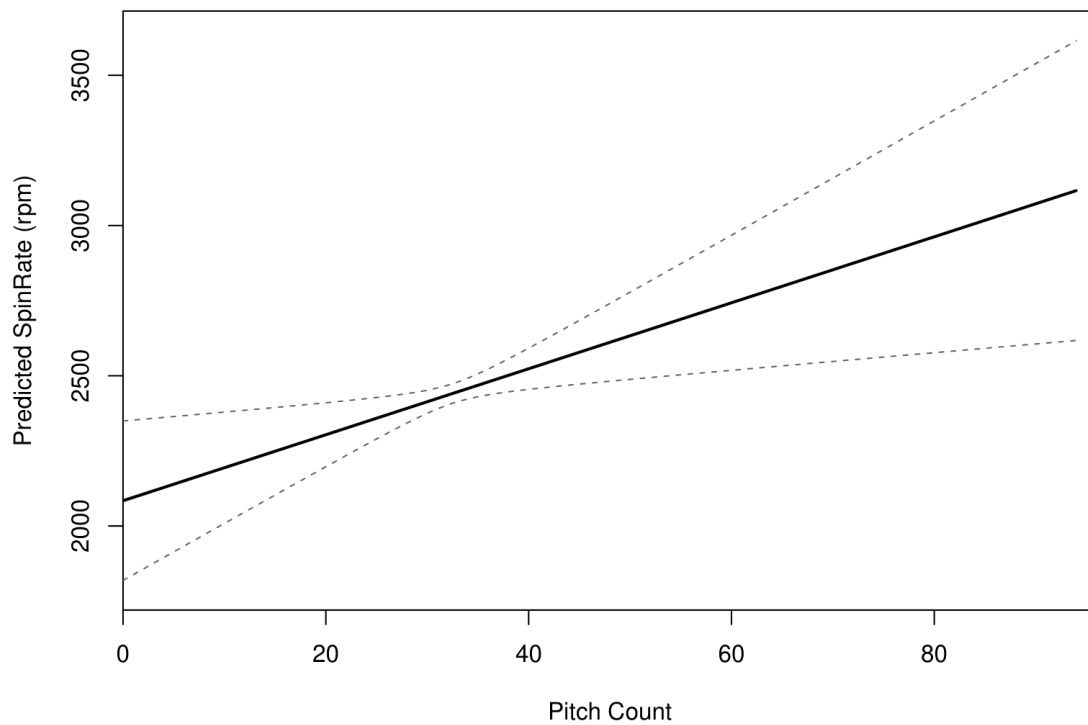
Spin Rate Prediction Curves

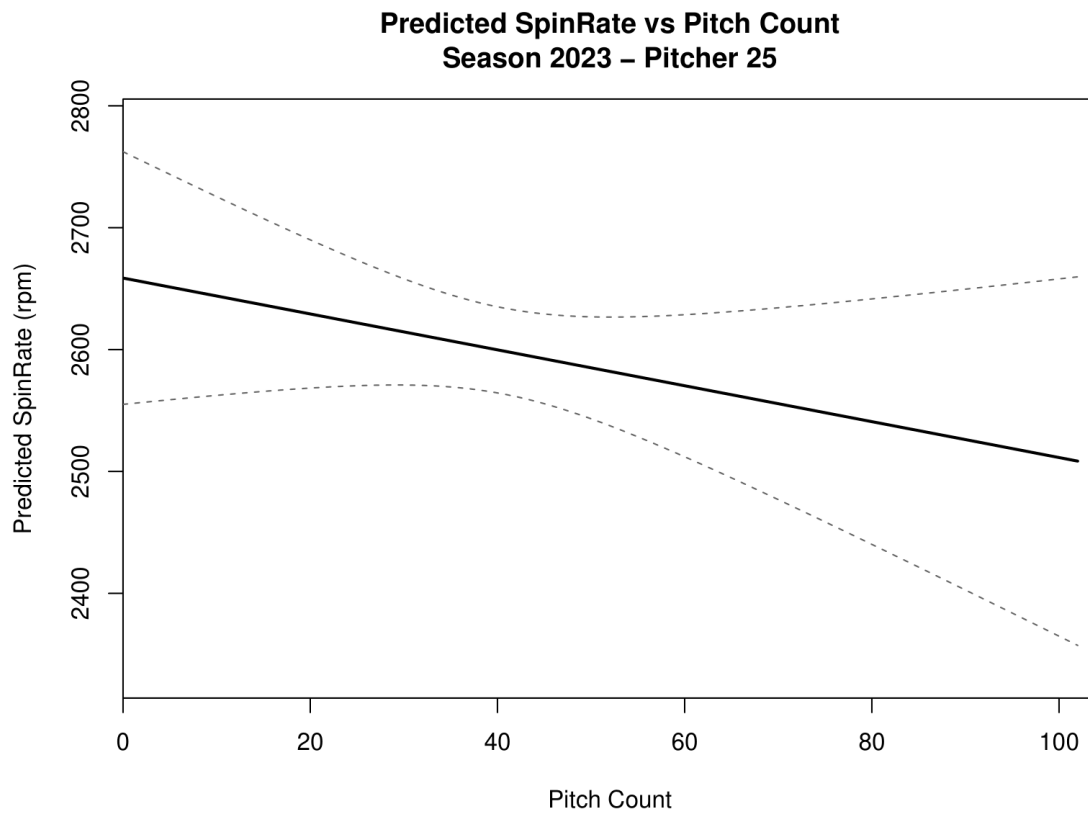
Predicted SpinRate vs Pitch Count
Season 2023 – Pitcher 22

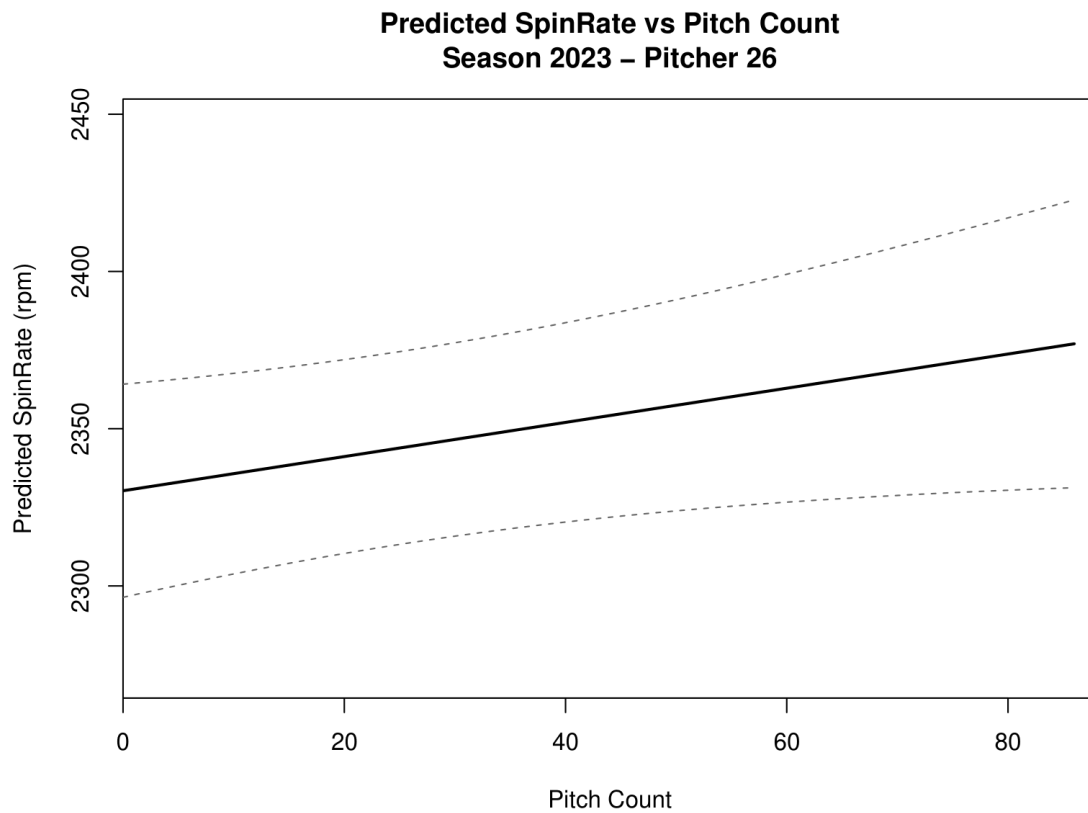


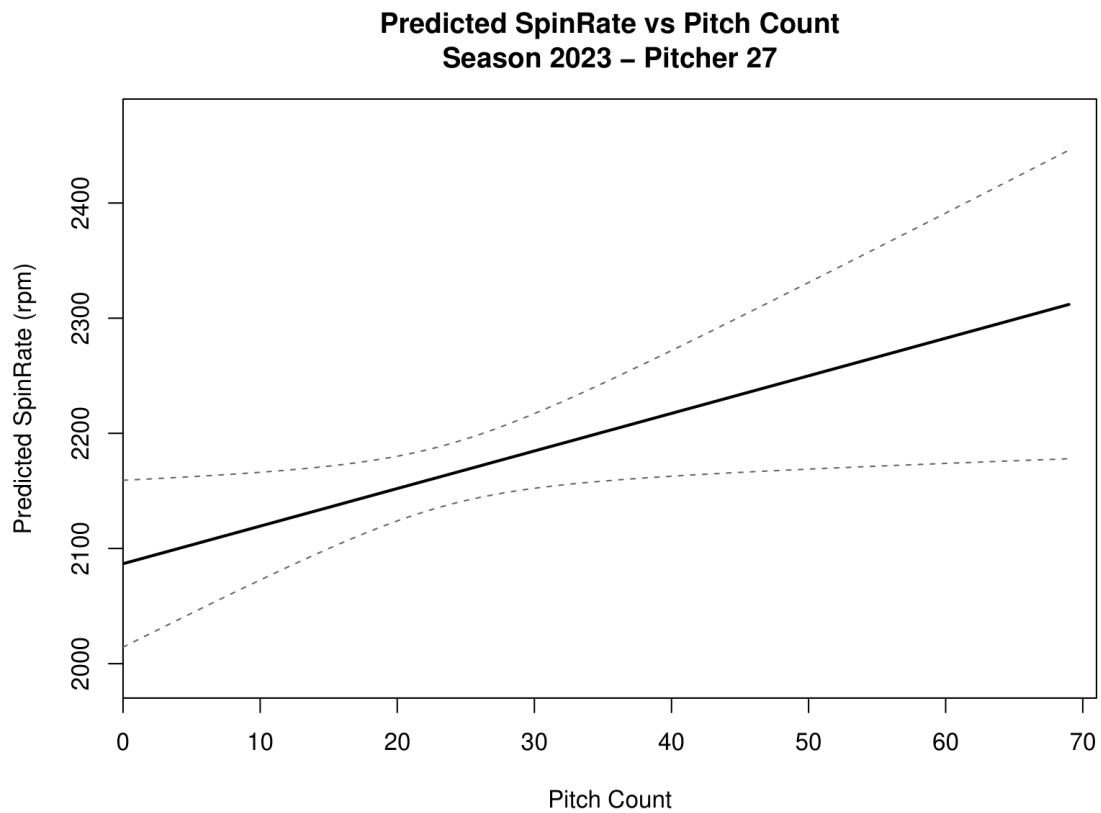


Predicted SpinRate vs Pitch Count
Season 2023 – Pitcher 24

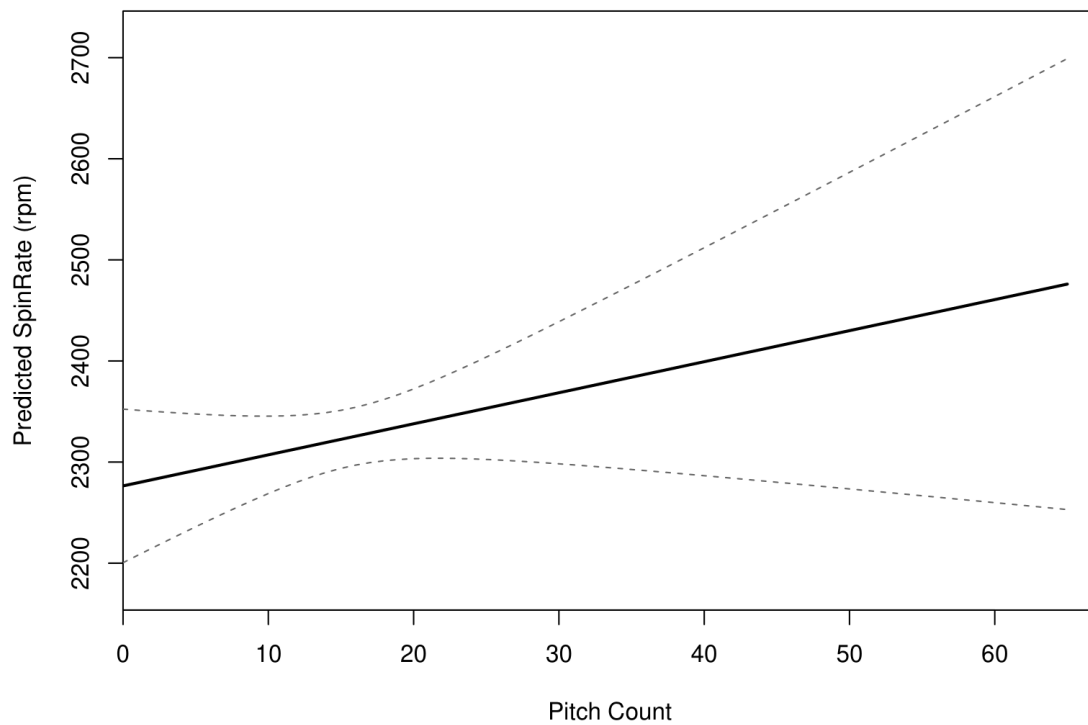




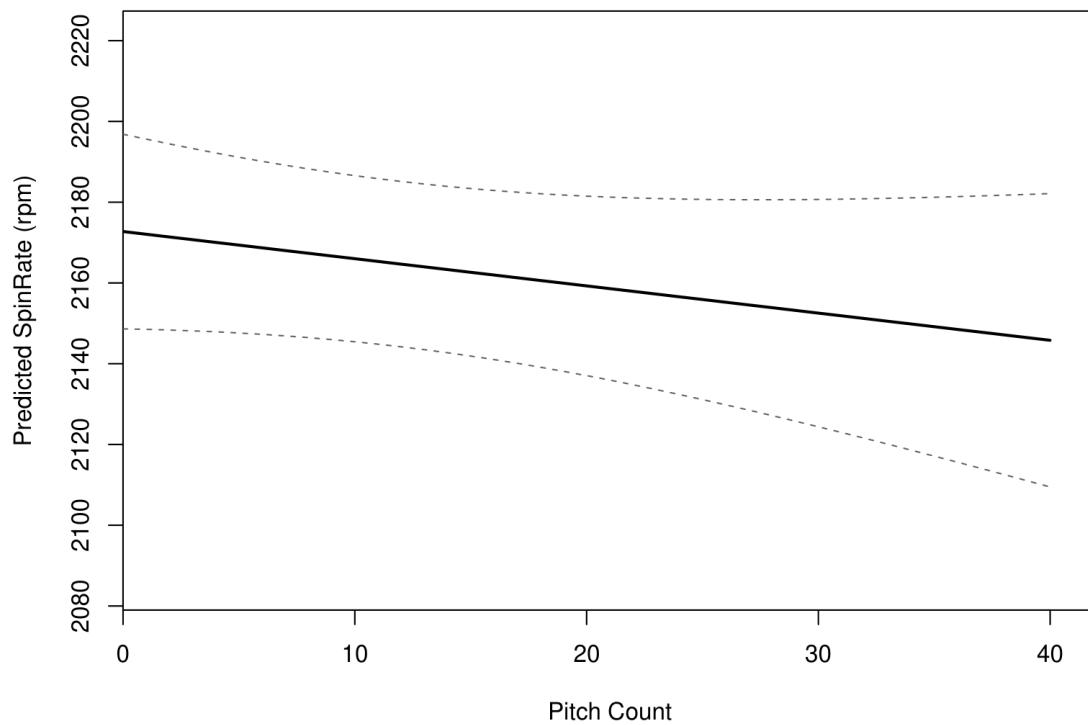


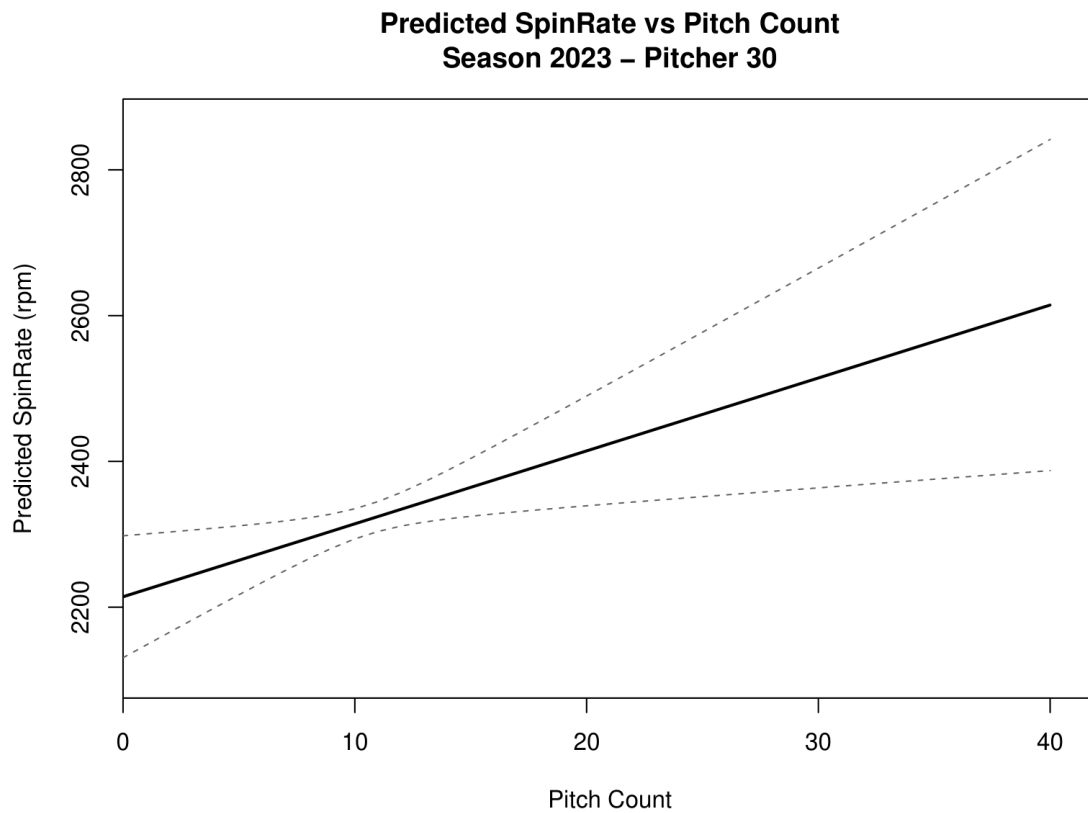


Predicted SpinRate vs Pitch Count
Season 2023 – Pitcher 28

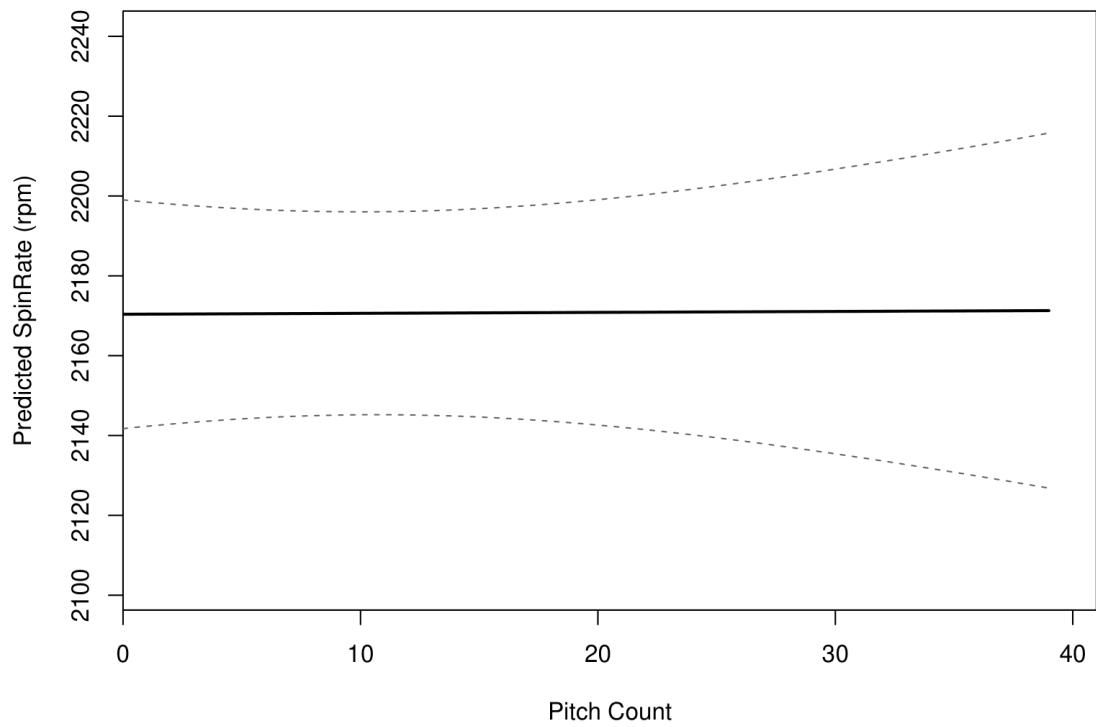


Predicted SpinRate vs Pitch Count
Season 2023 – Pitcher 29

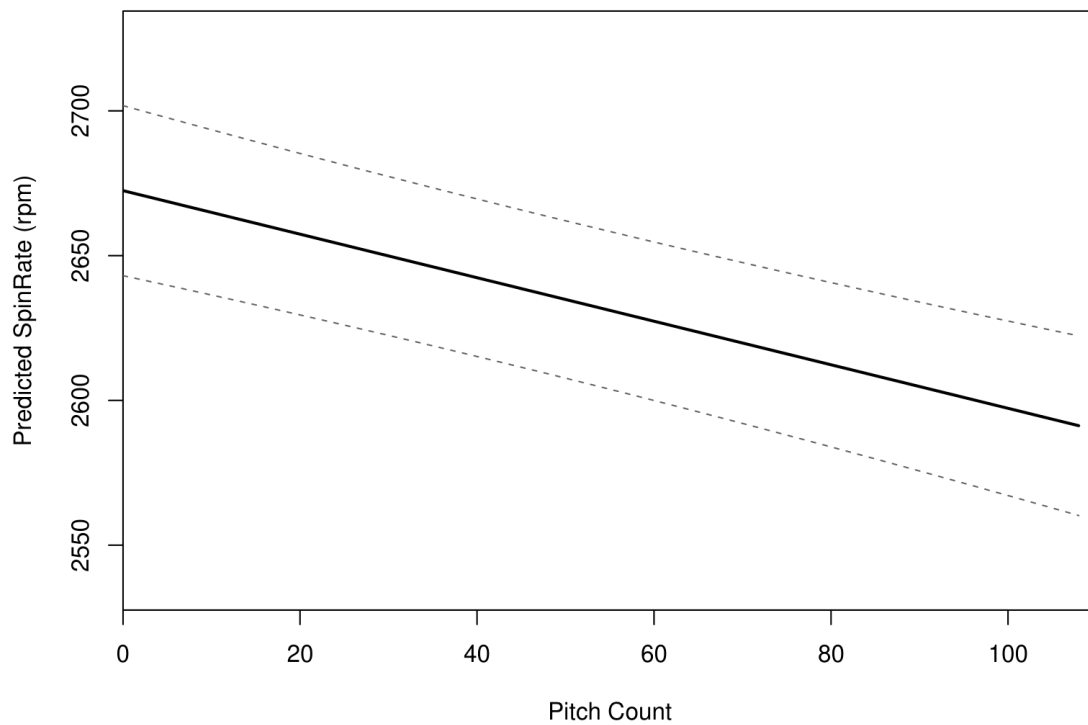


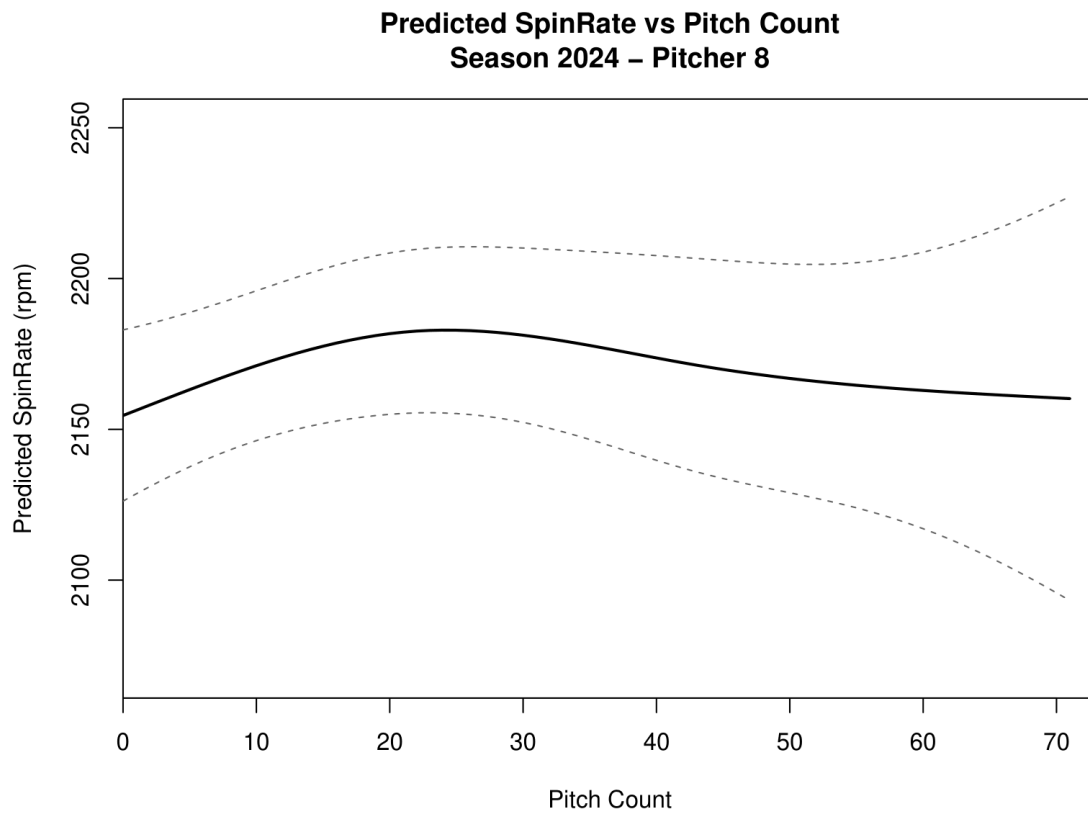


Predicted SpinRate vs Pitch Count
Season 2023 – Pitcher 31

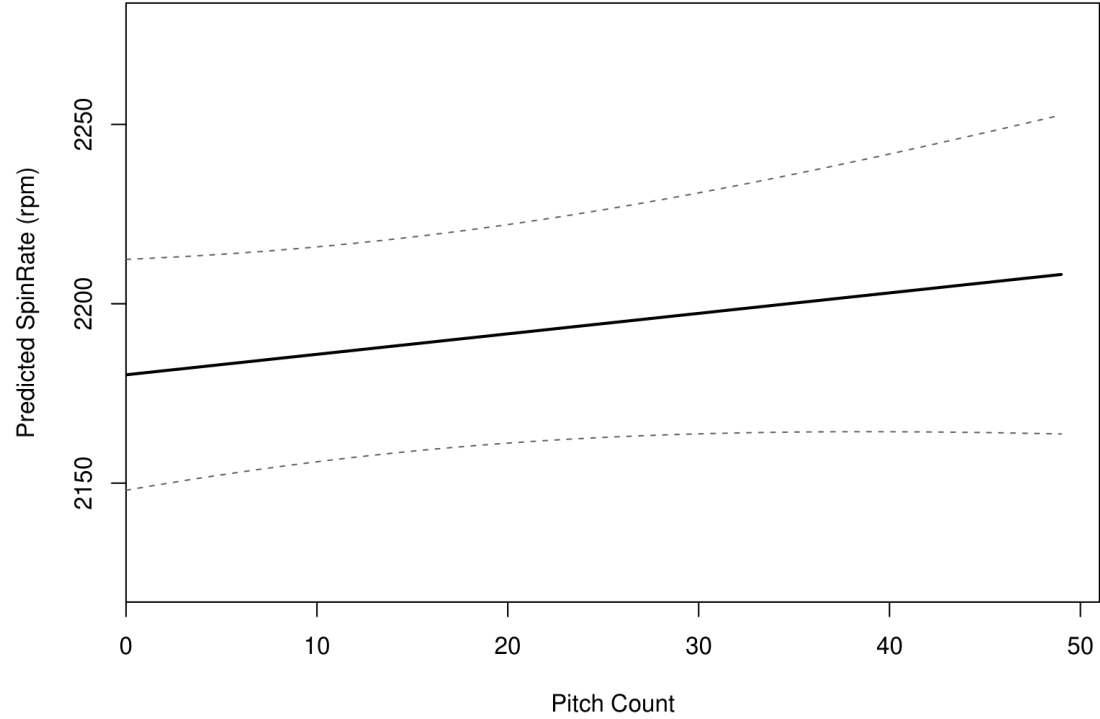


Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 5

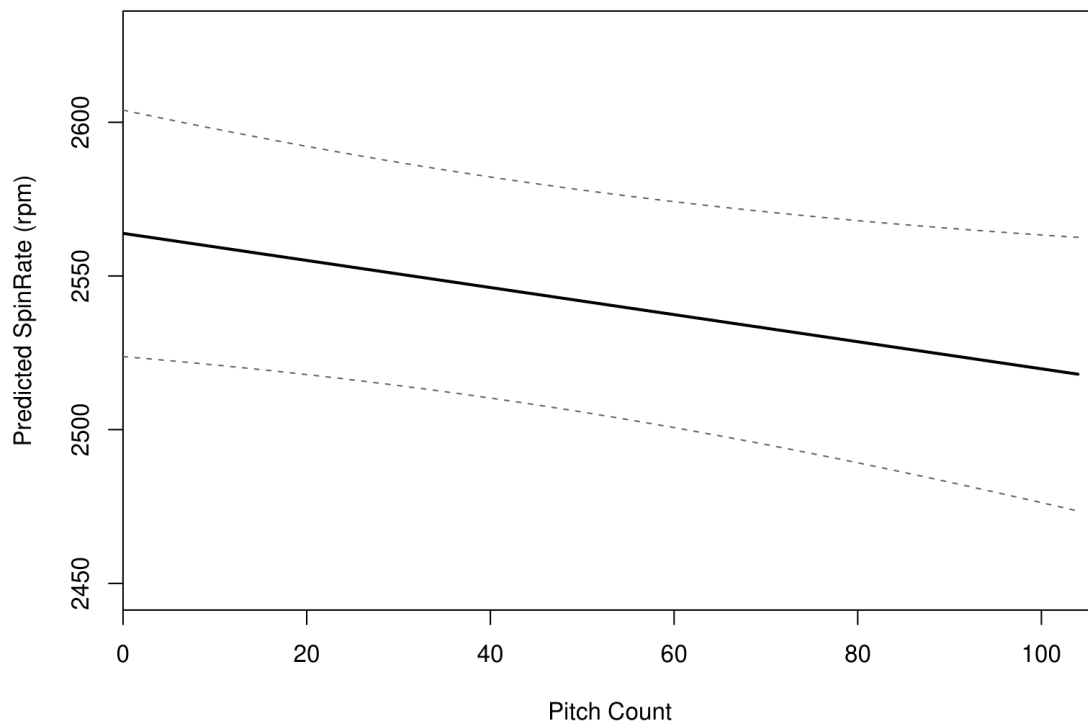




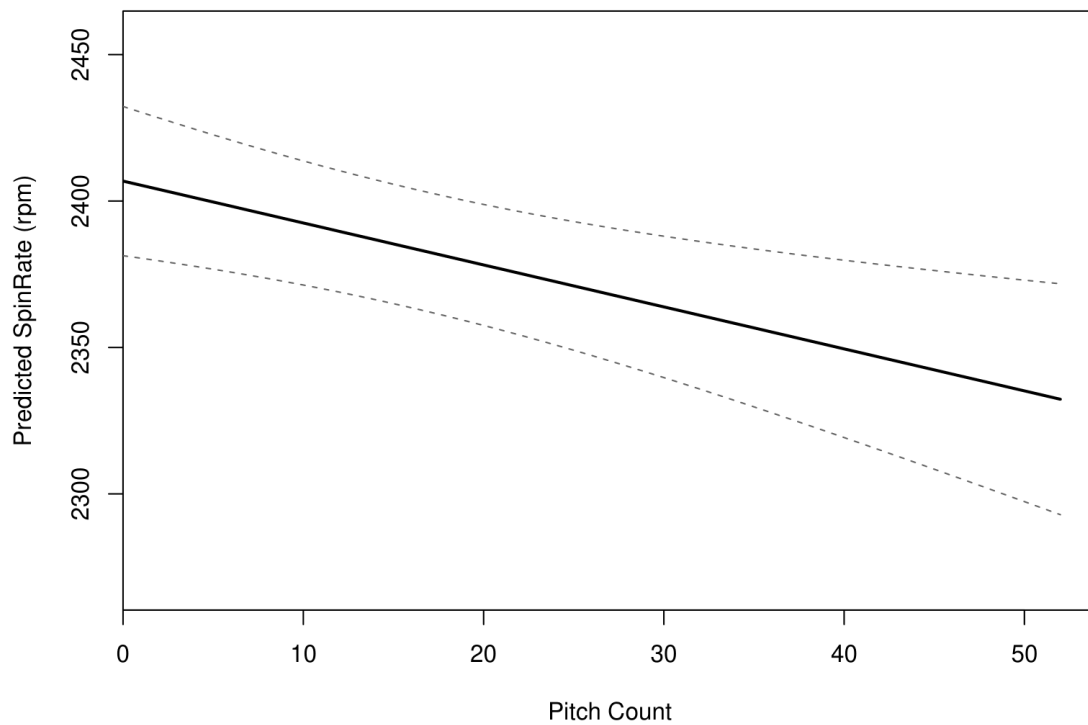
Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 11



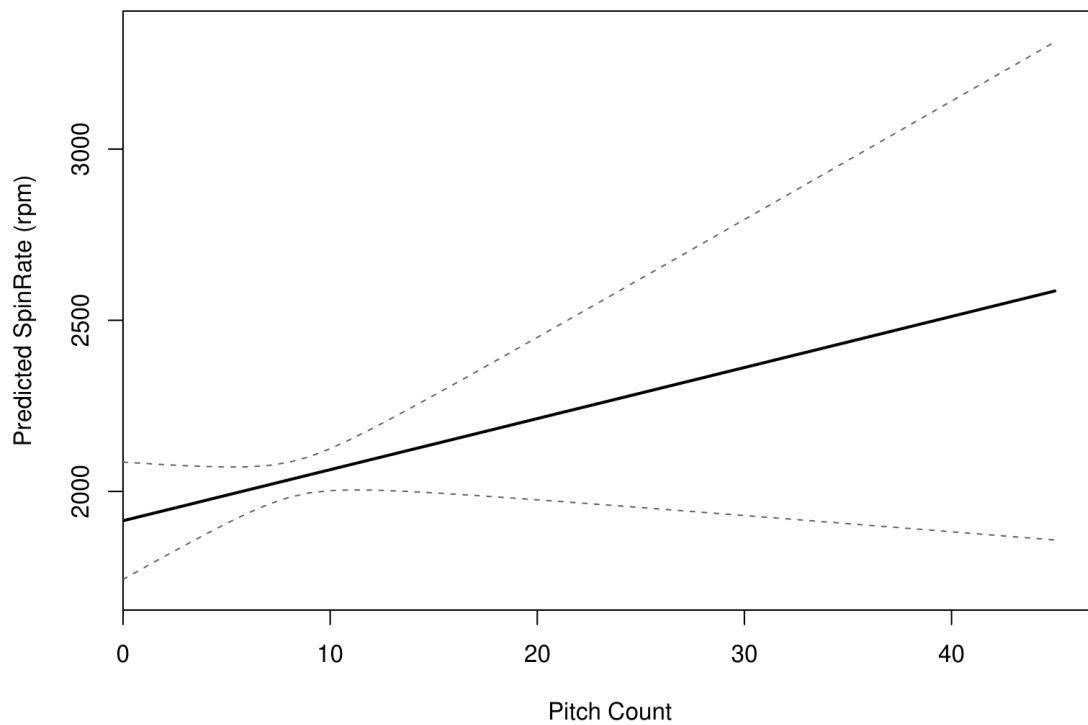
Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 12

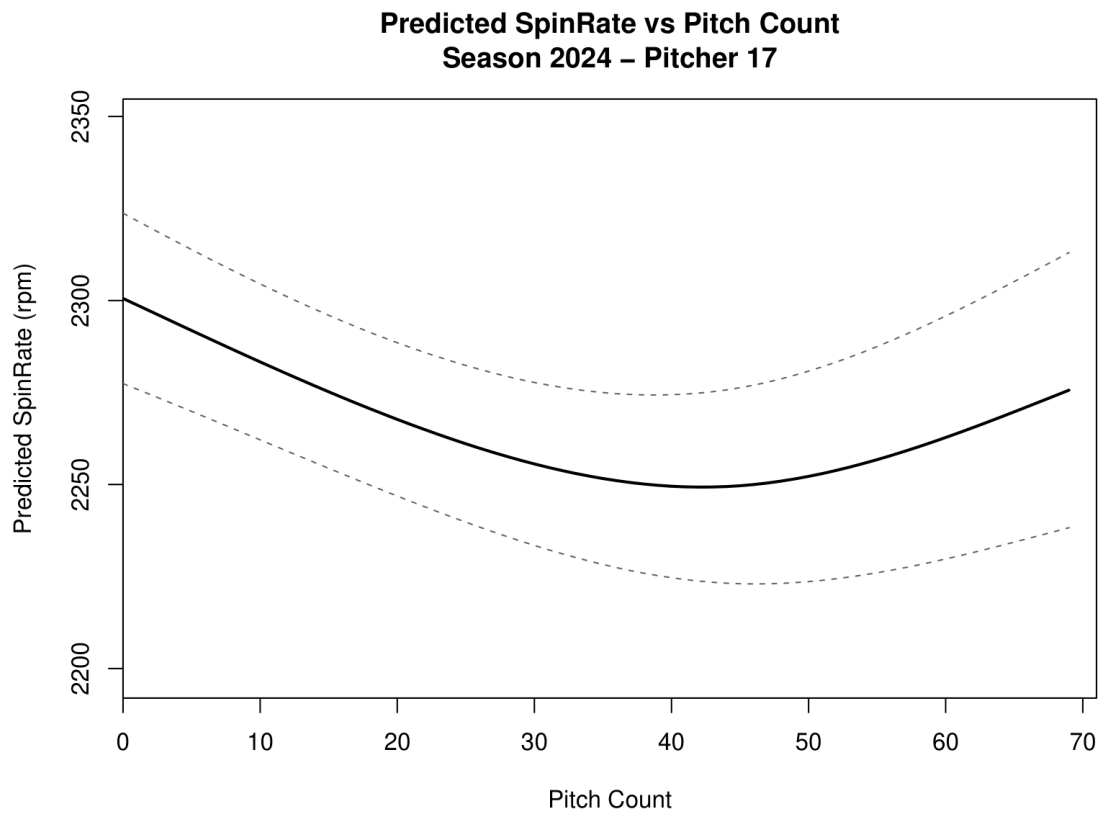


Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 14

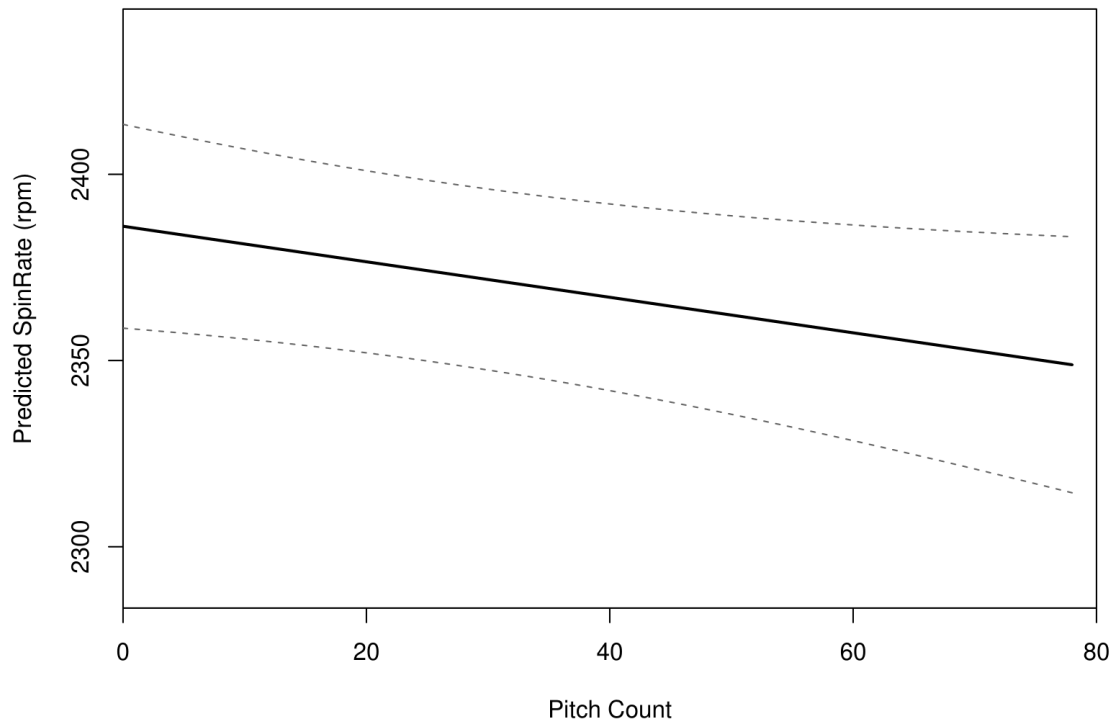


Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 15

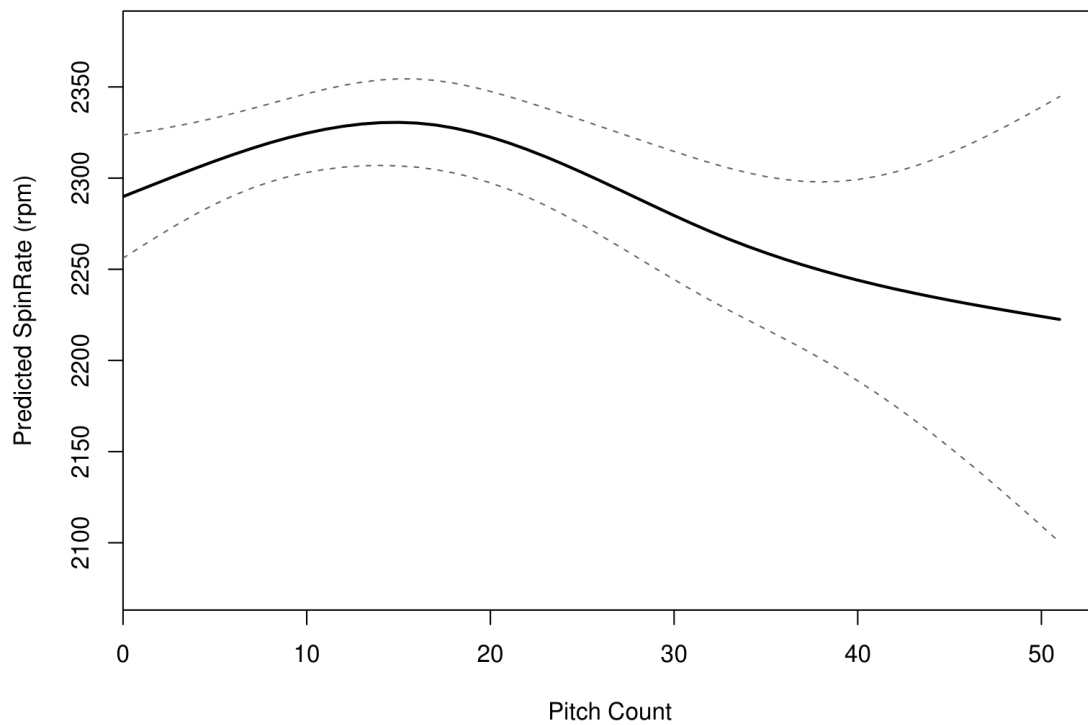




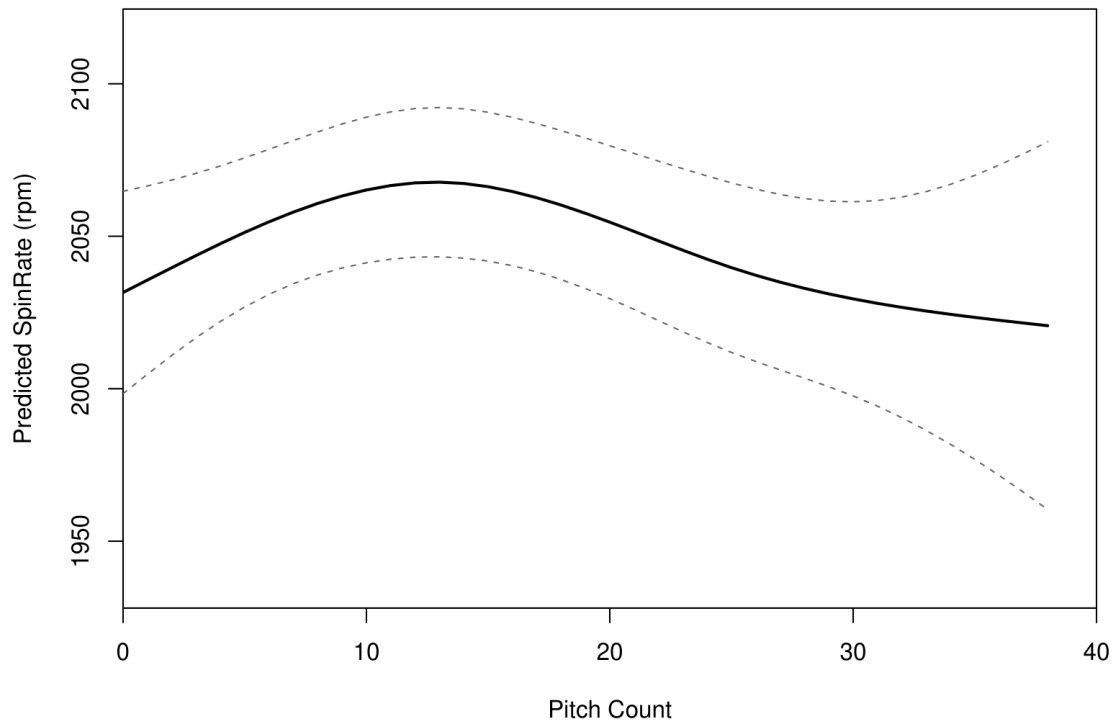
Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 19

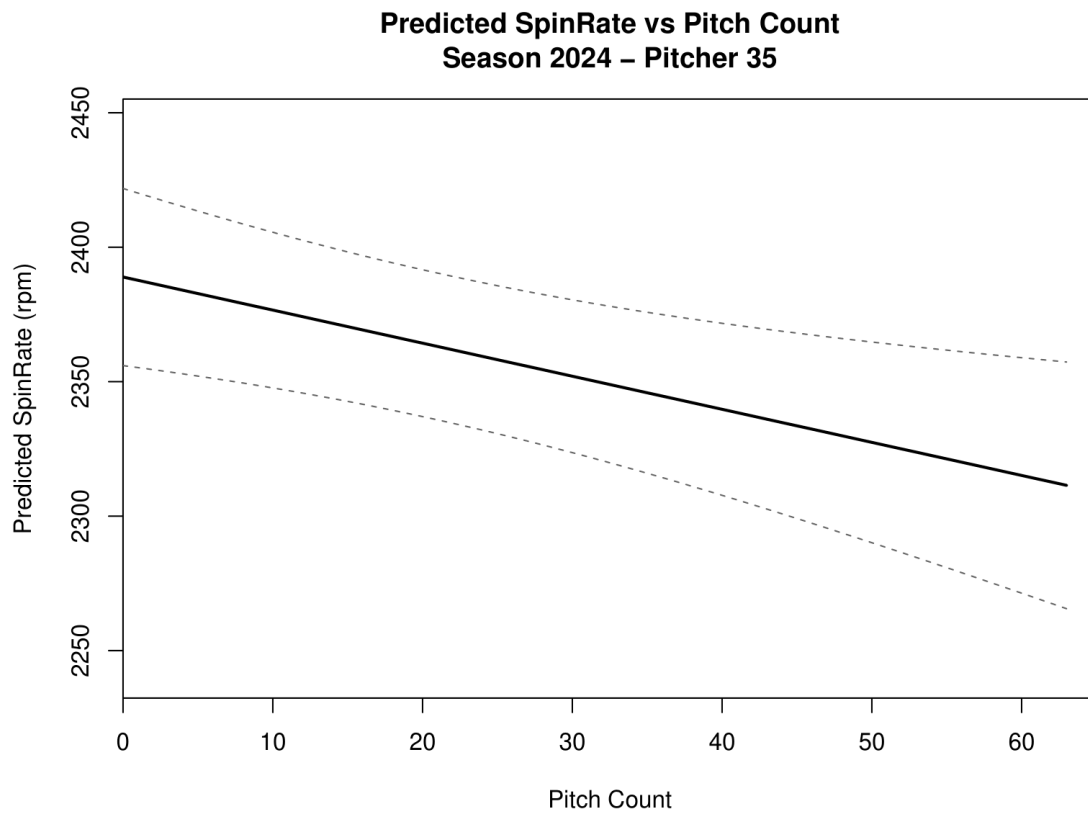


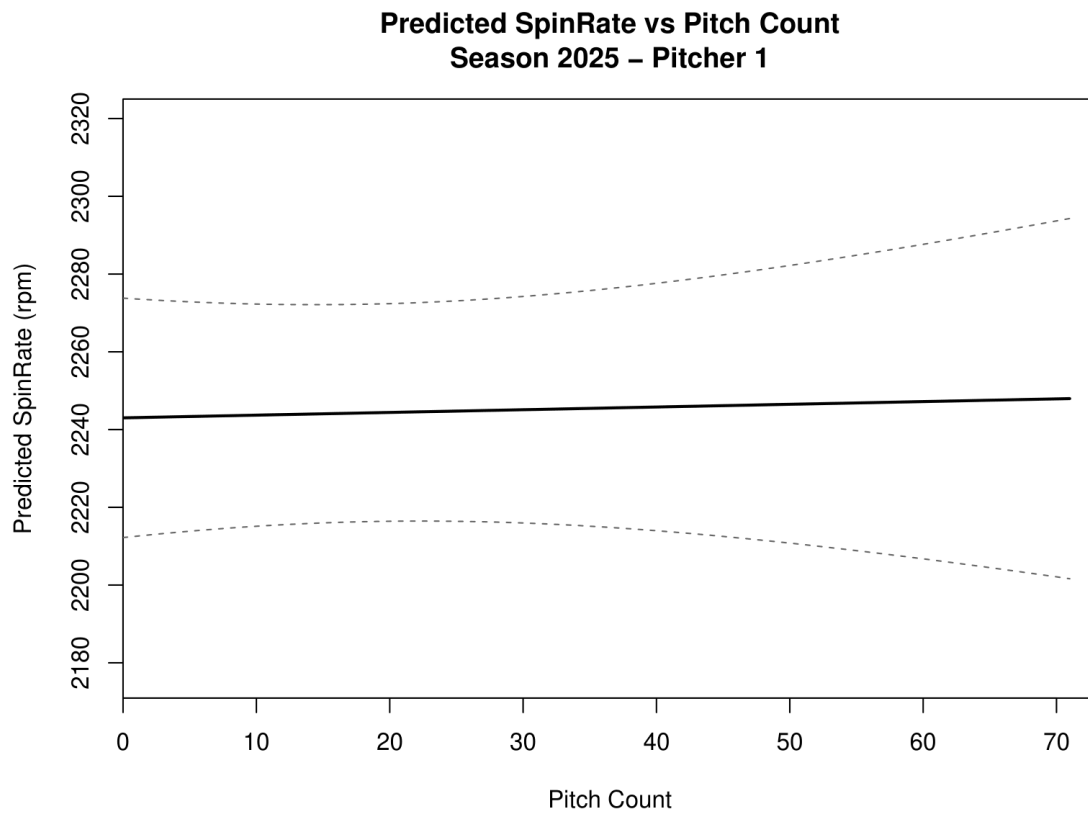
Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 21



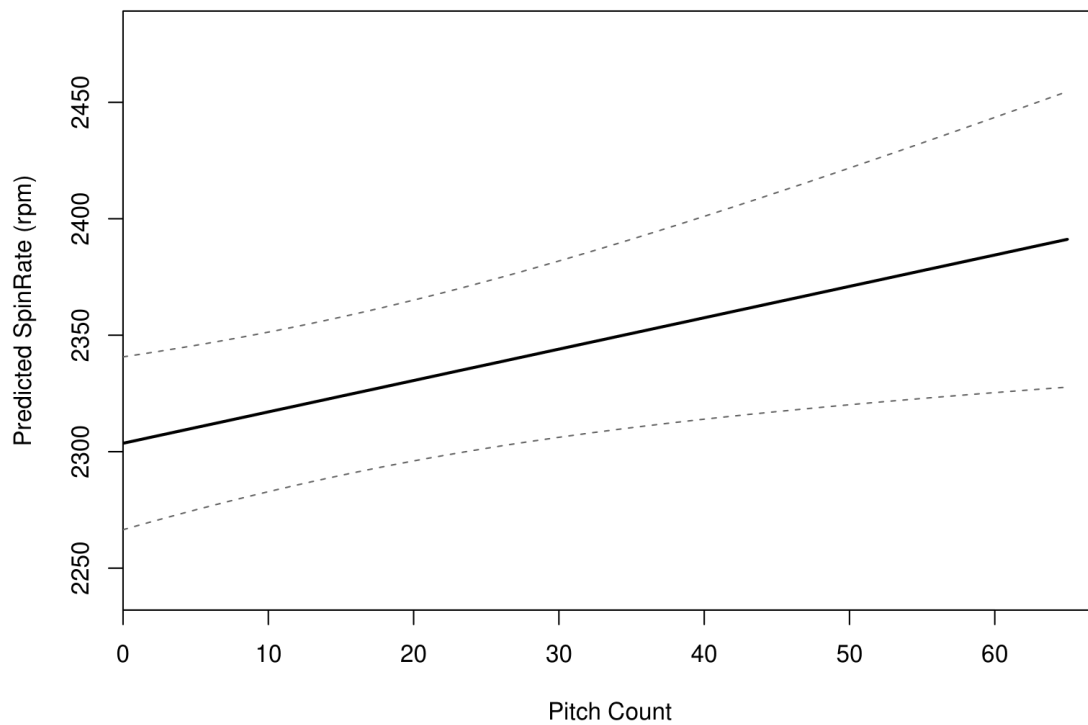
Predicted SpinRate vs Pitch Count
Season 2024 – Pitcher 33

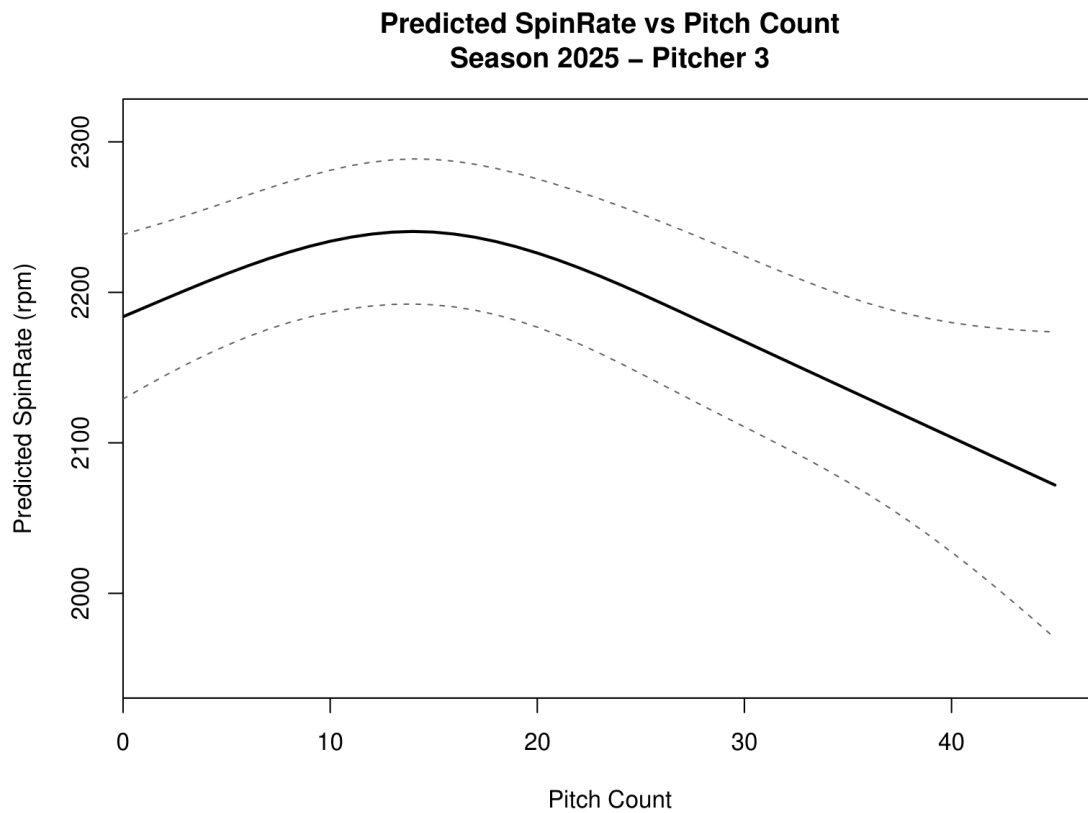




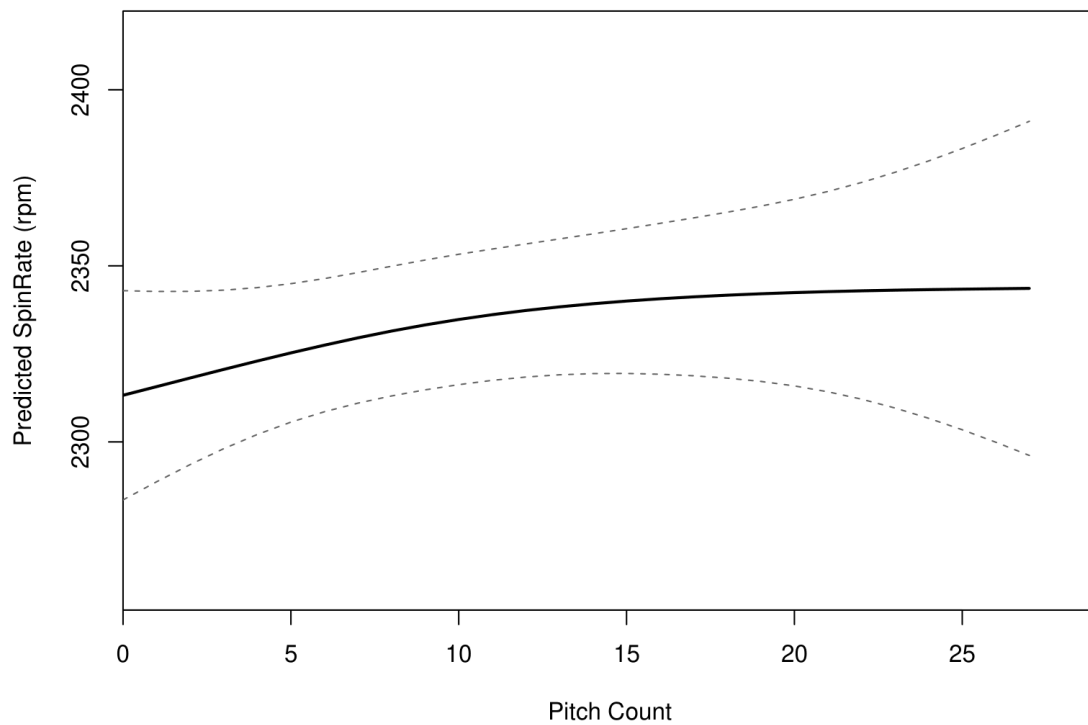


Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 2

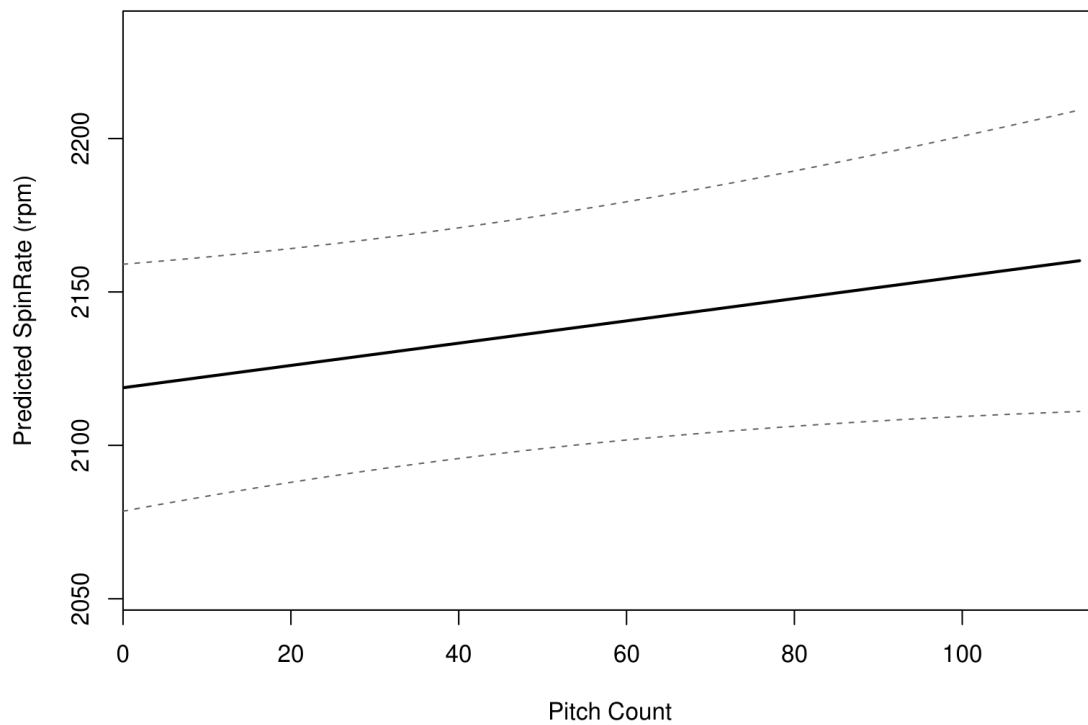




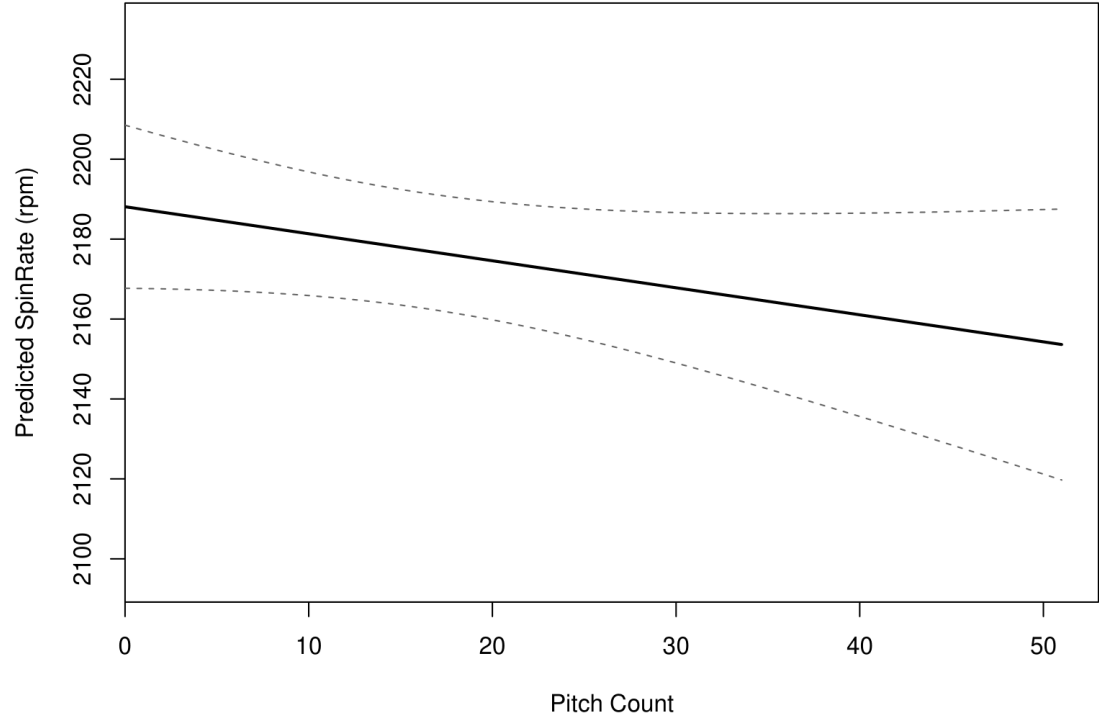
Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 4



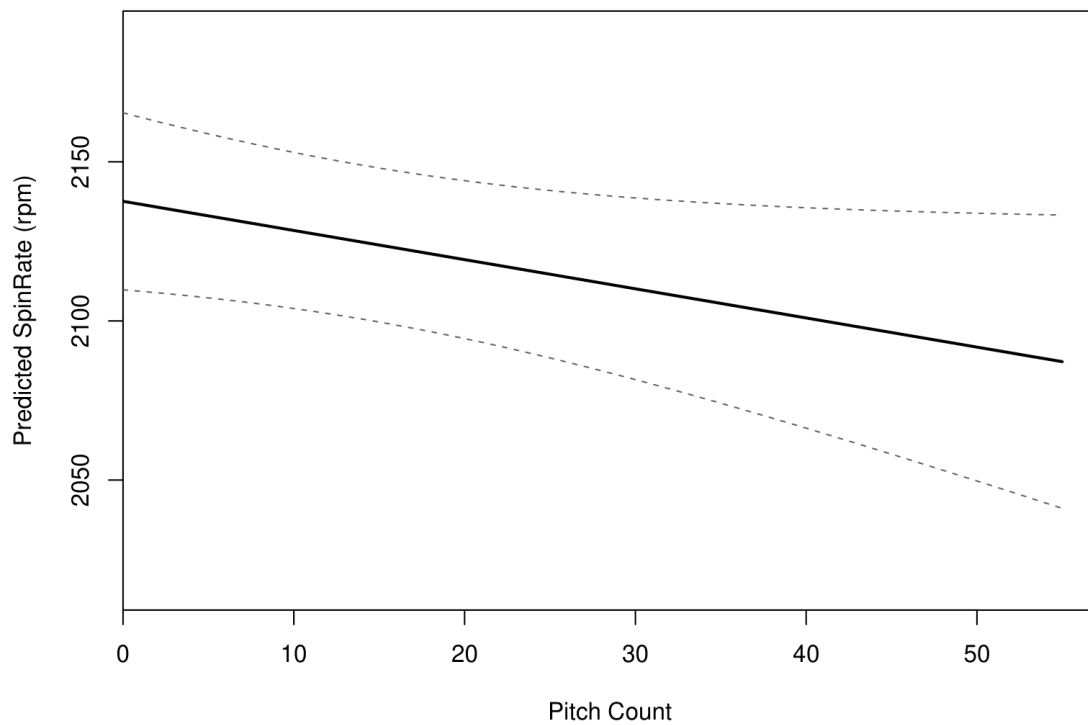
Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 6

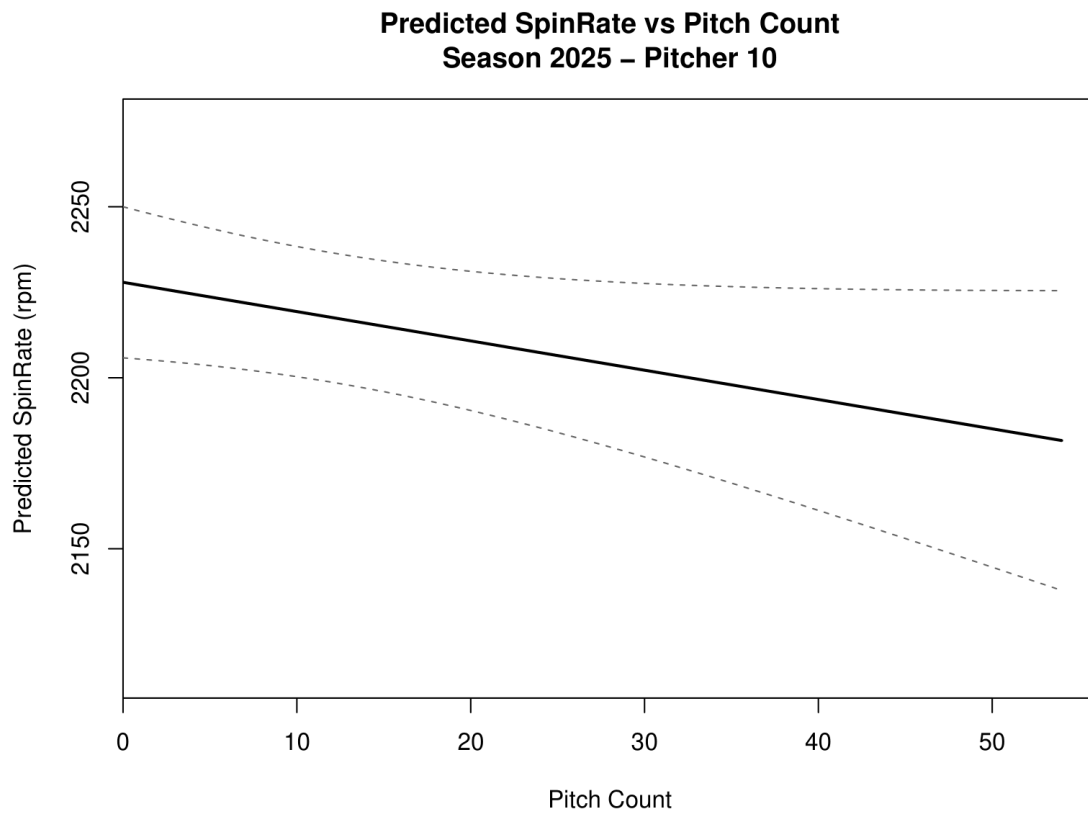


Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 7

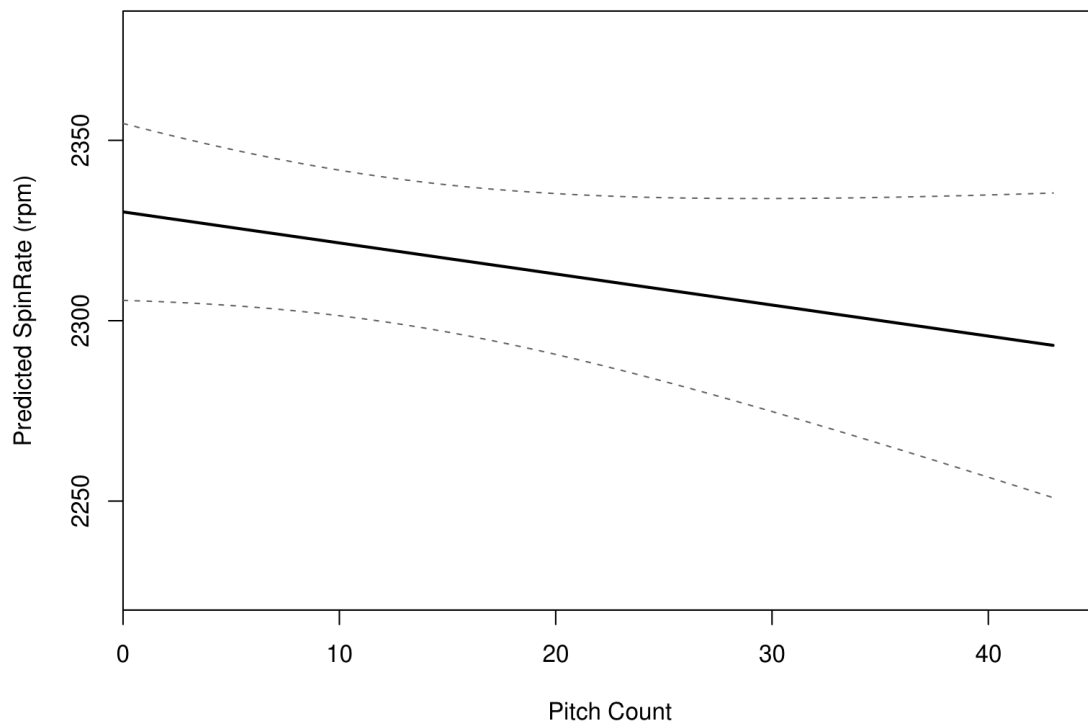


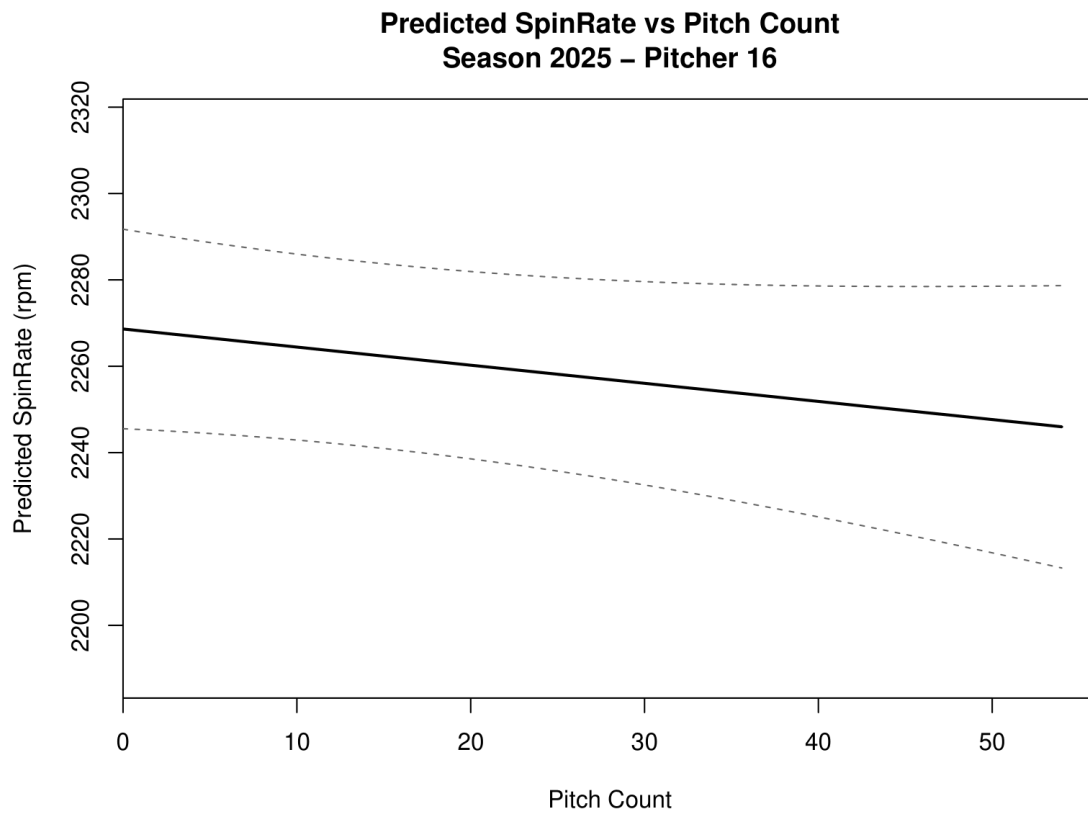
Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 9



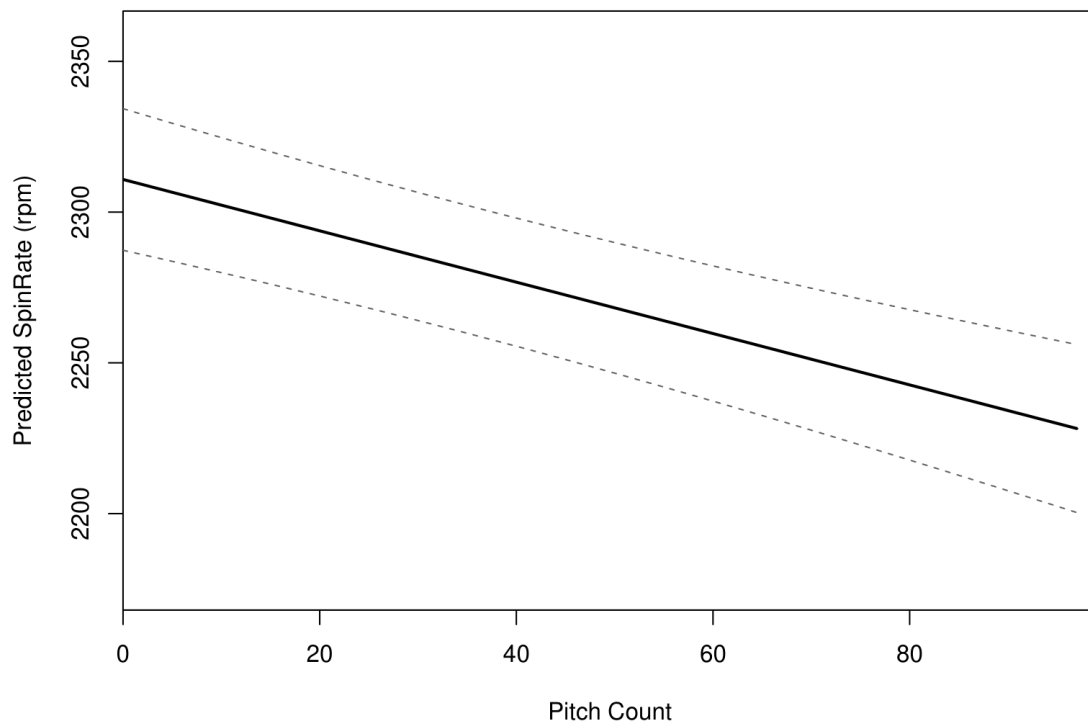


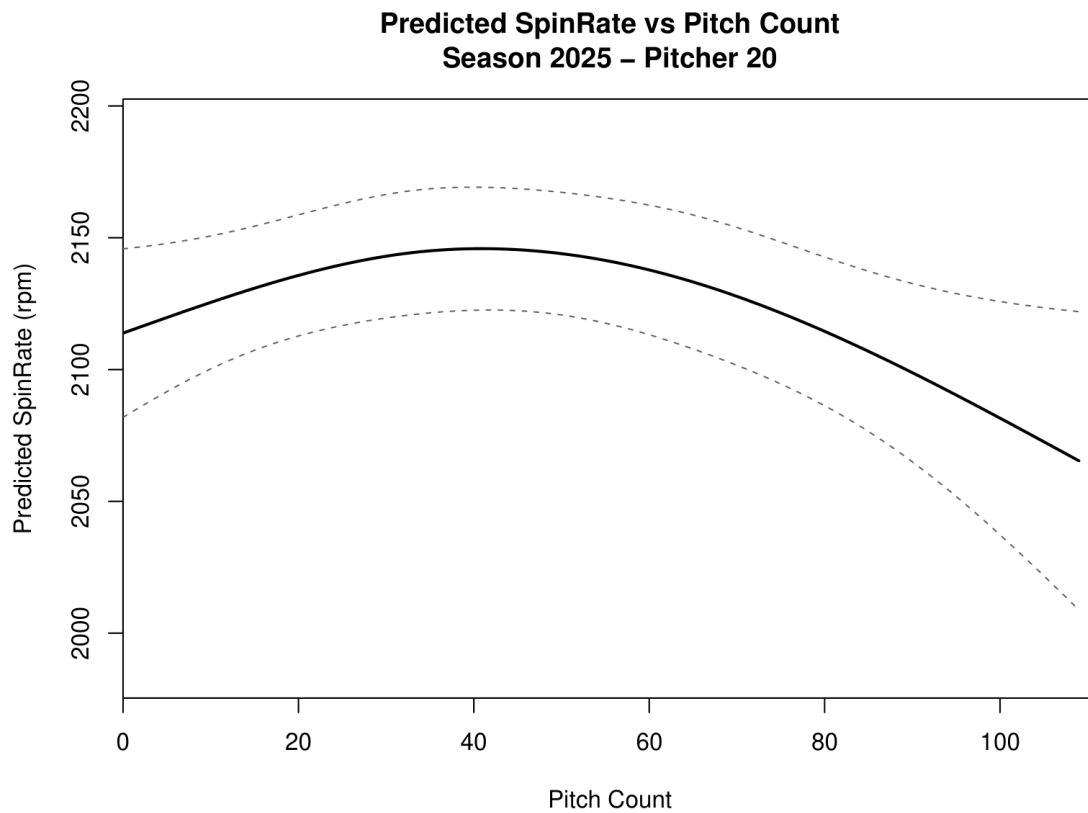
Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 13



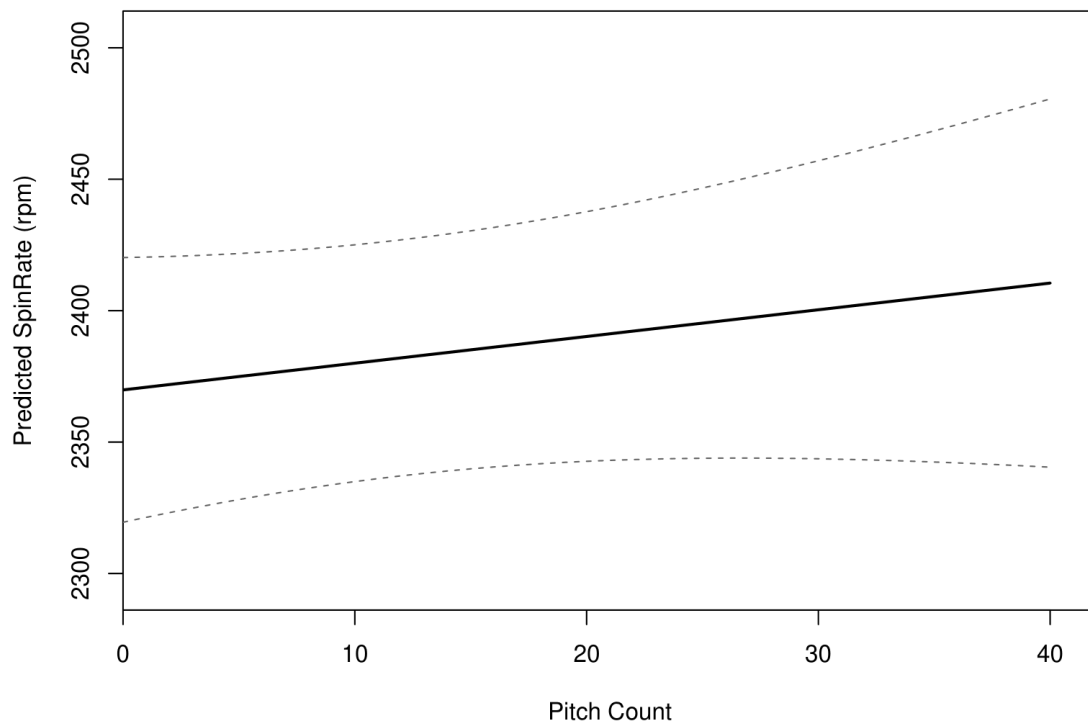


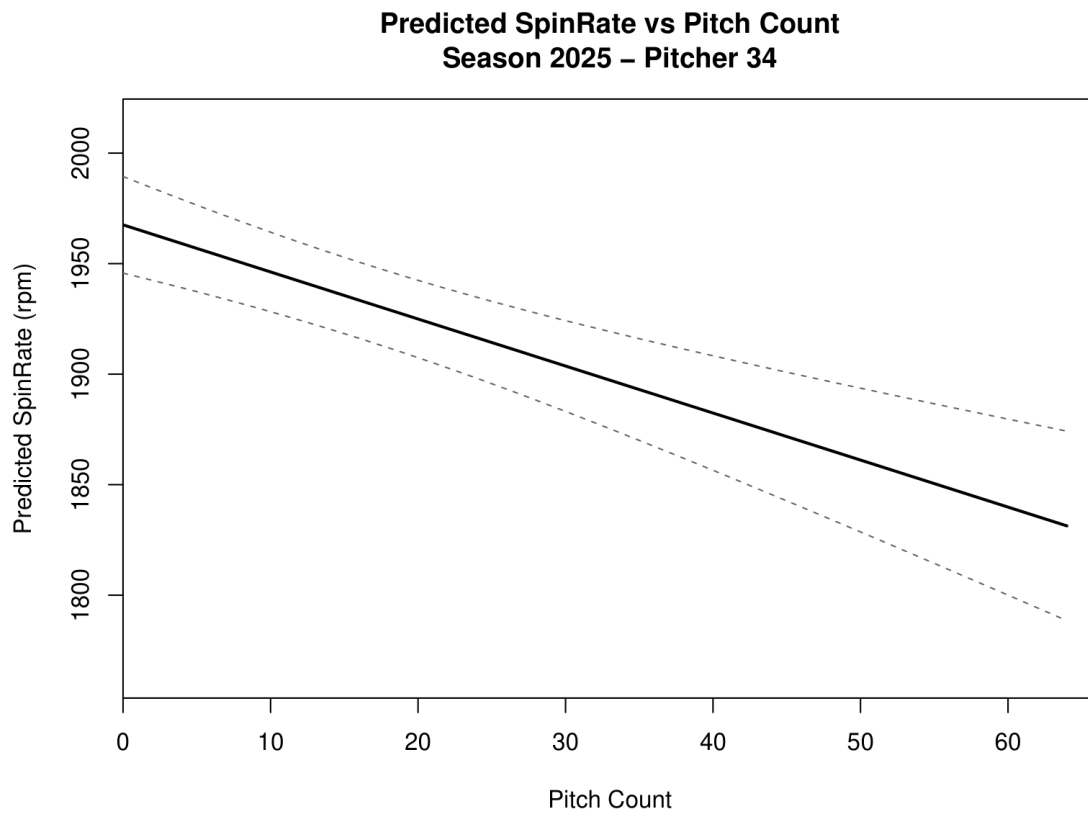
Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 18



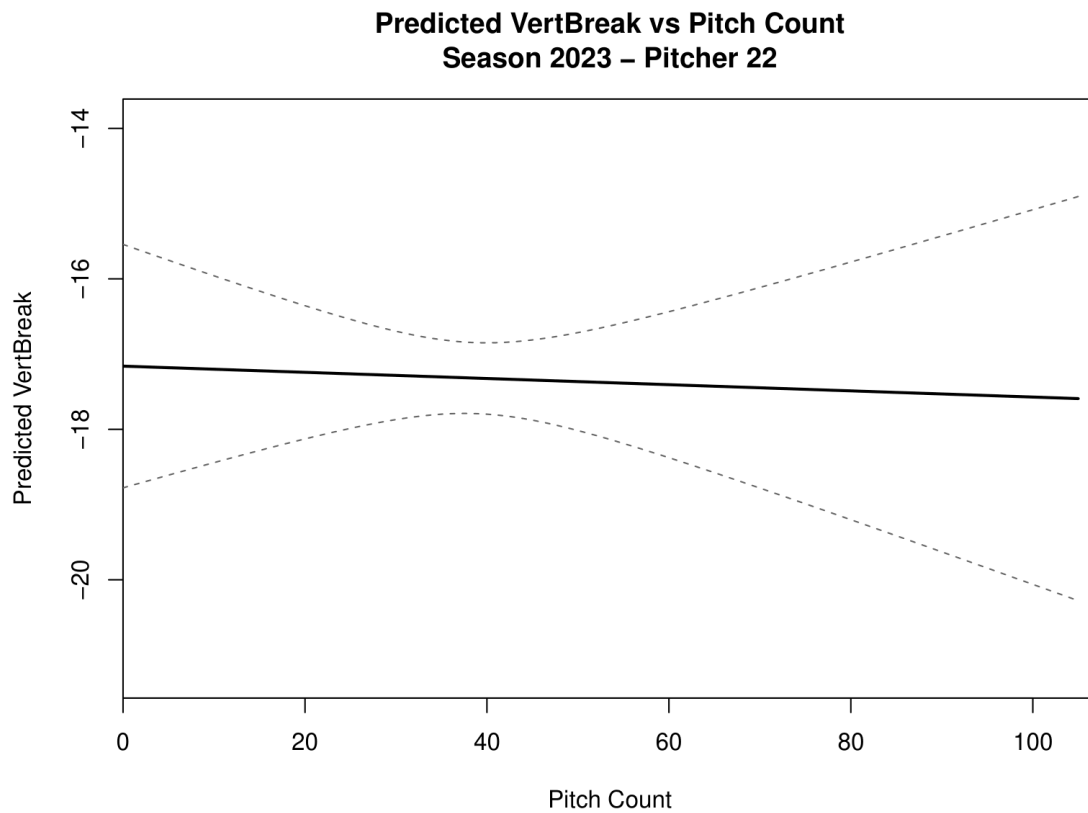


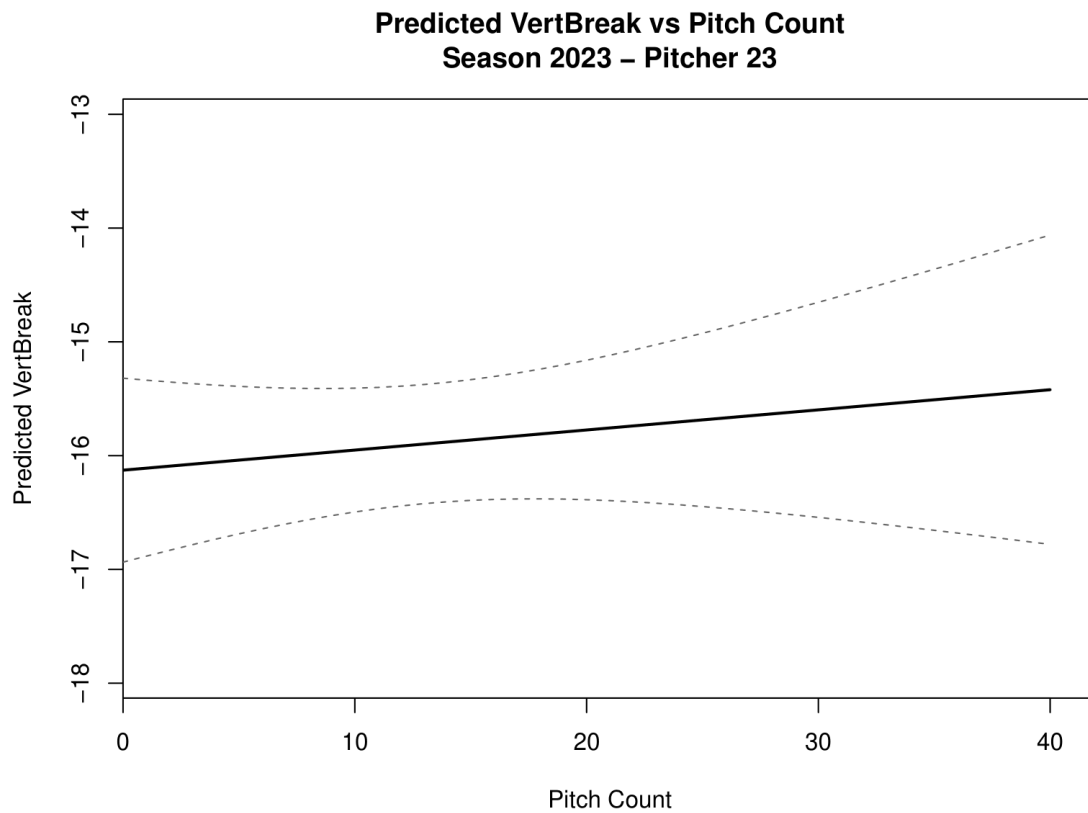
Predicted SpinRate vs Pitch Count
Season 2025 – Pitcher 32



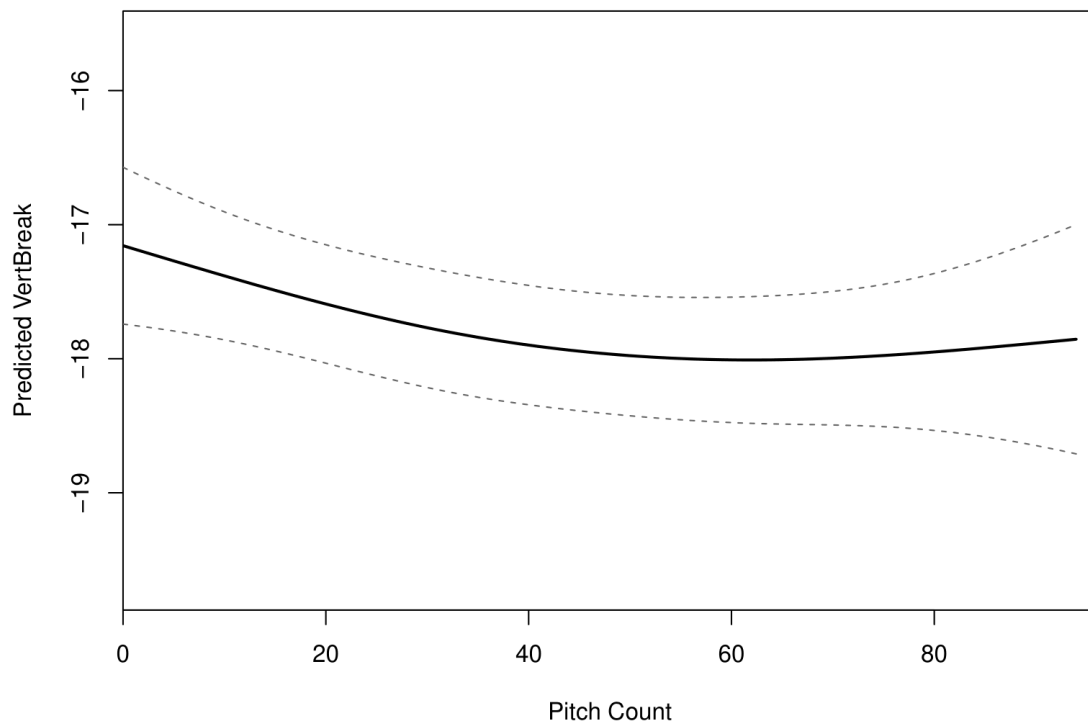


Vertical Break Prediction Curves

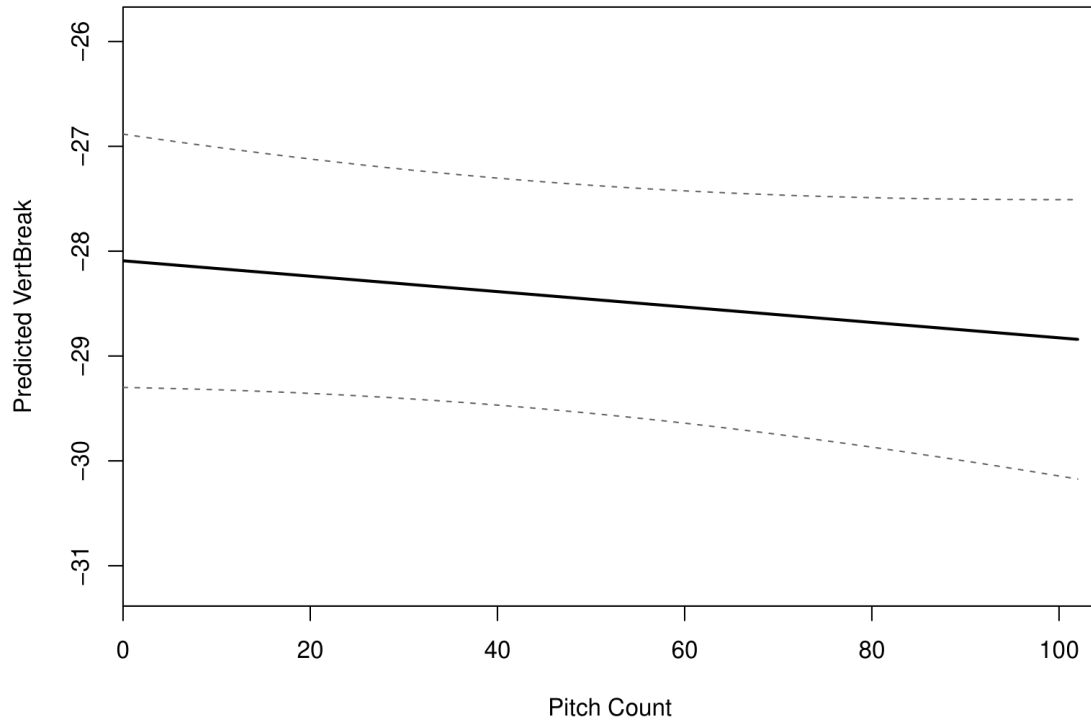




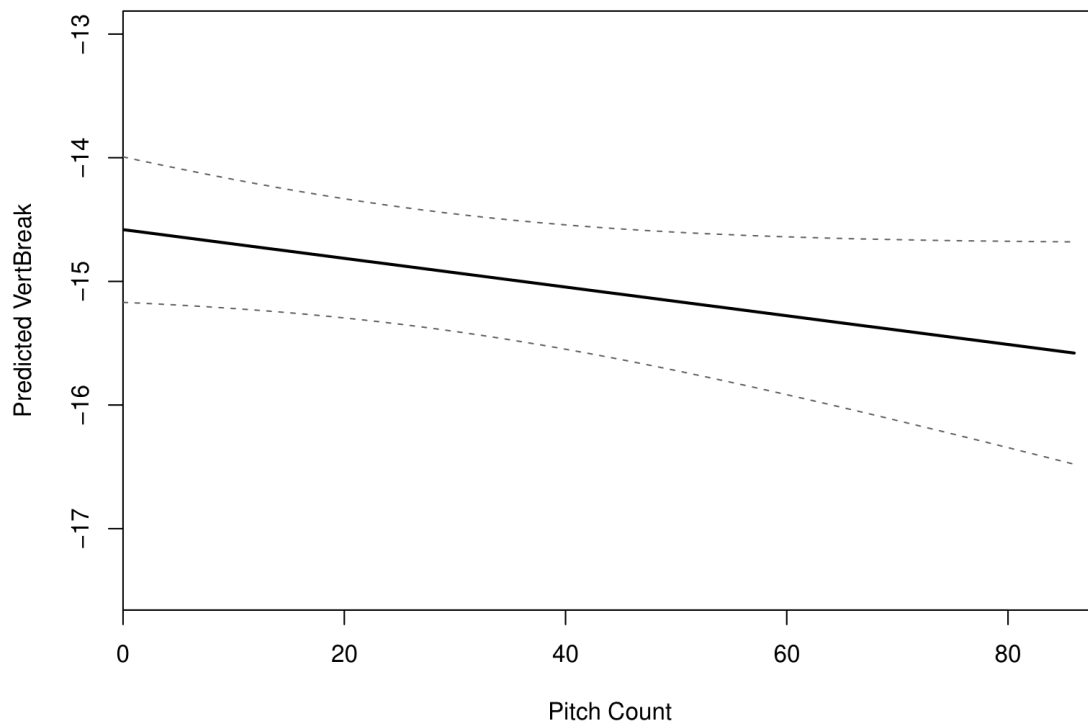
Predicted VertBreak vs Pitch Count
Season 2023 – Pitcher 24



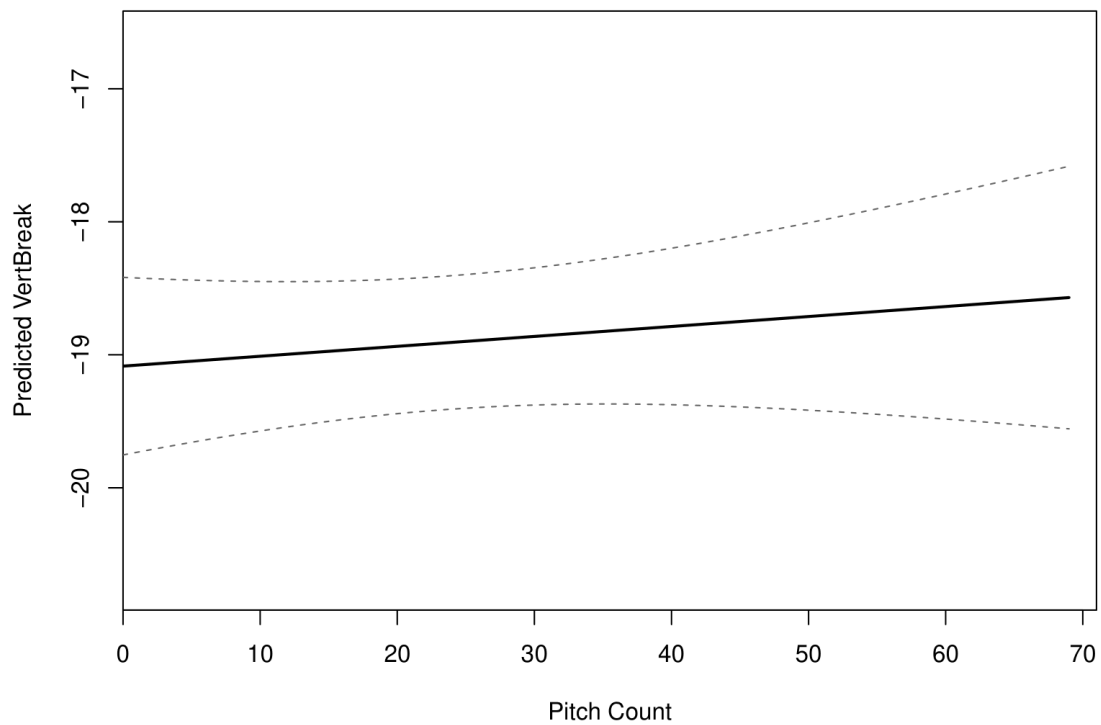
Predicted VertBreak vs Pitch Count
Season 2023 – Pitcher 25

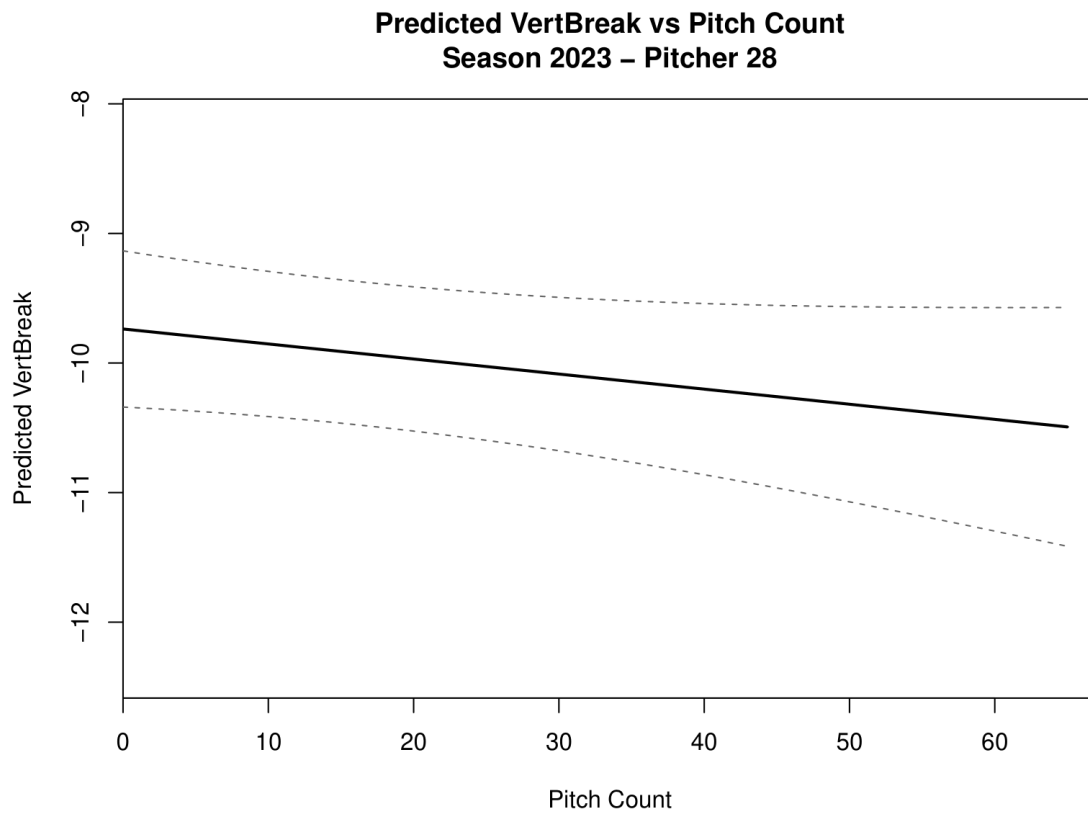


Predicted VertBreak vs Pitch Count
Season 2023 – Pitcher 26

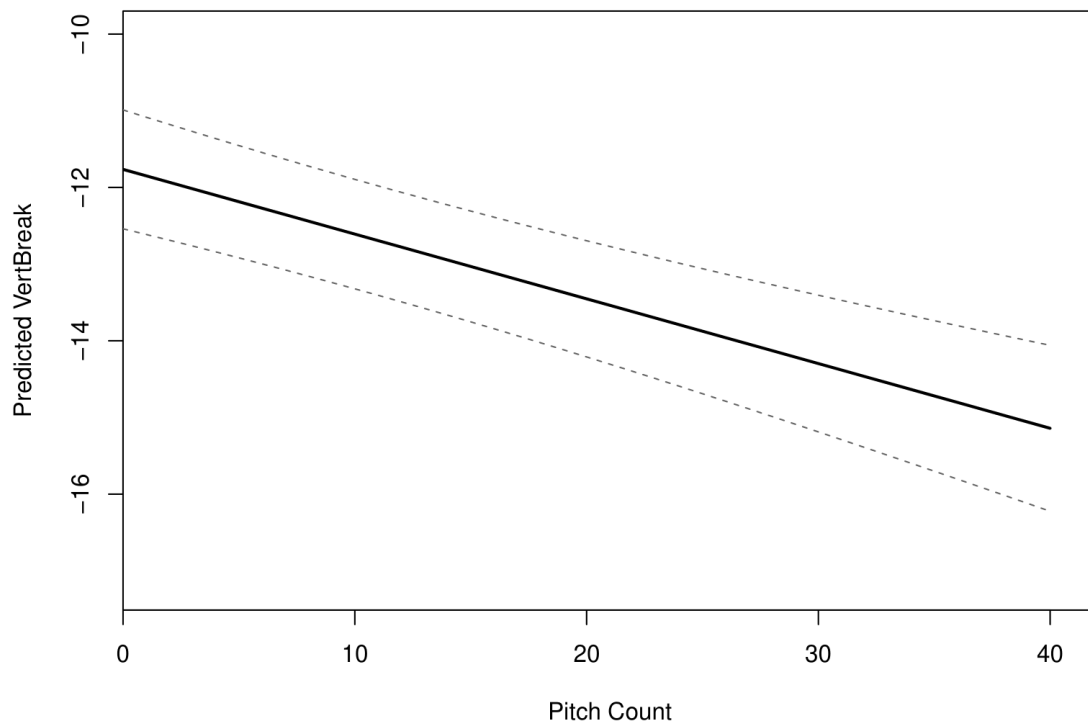


Predicted VertBreak vs Pitch Count
Season 2023 – Pitcher 27

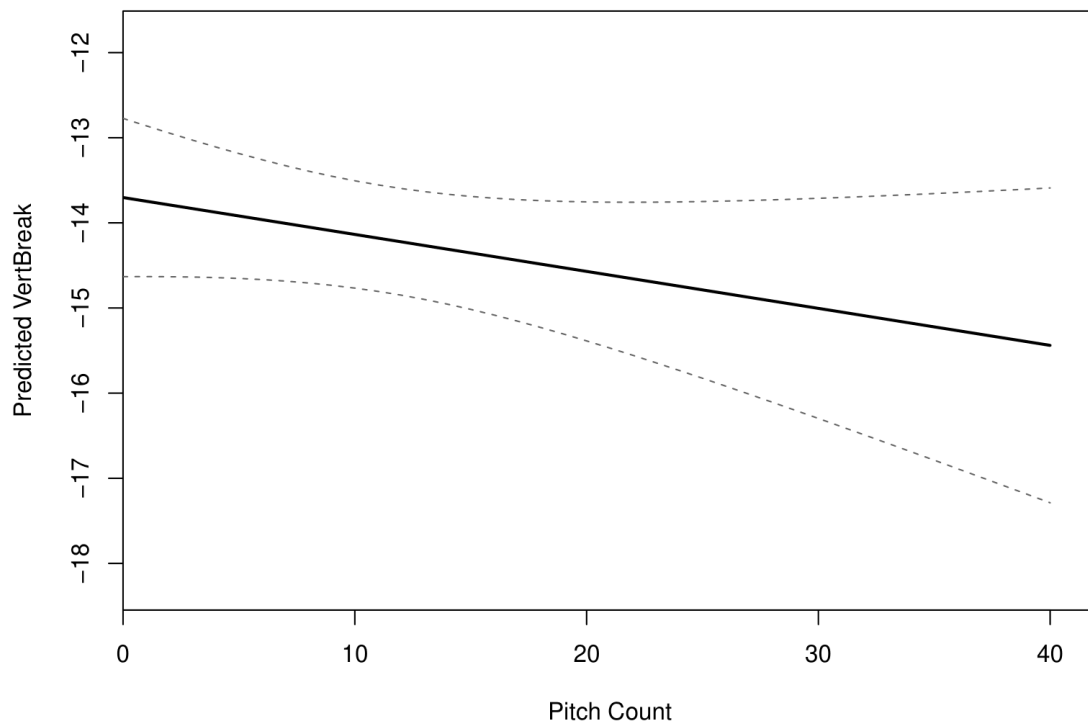


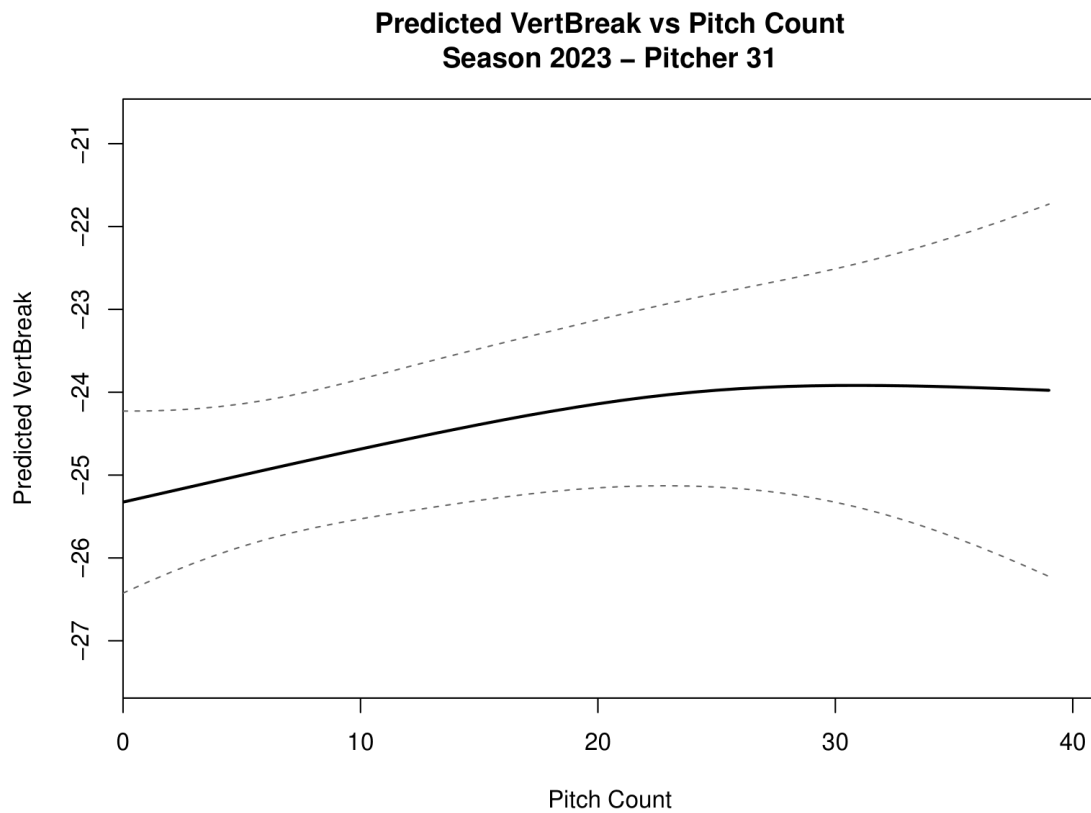


Predicted VertBreak vs Pitch Count
Season 2023 – Pitcher 29

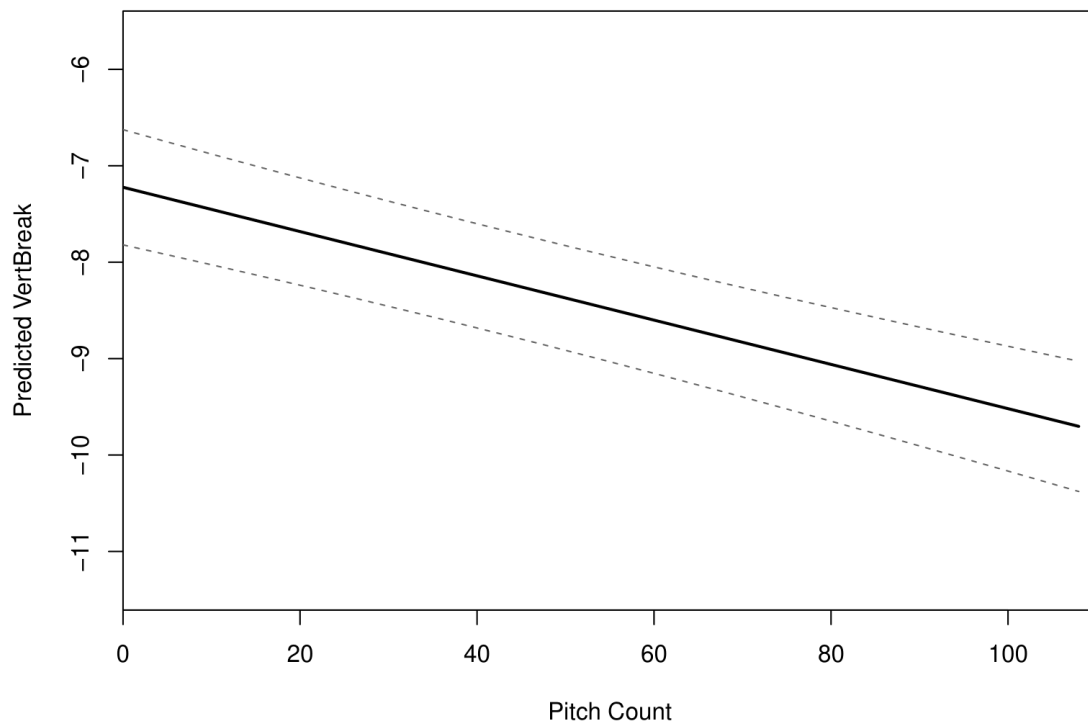


Predicted VertBreak vs Pitch Count
Season 2023 – Pitcher 30

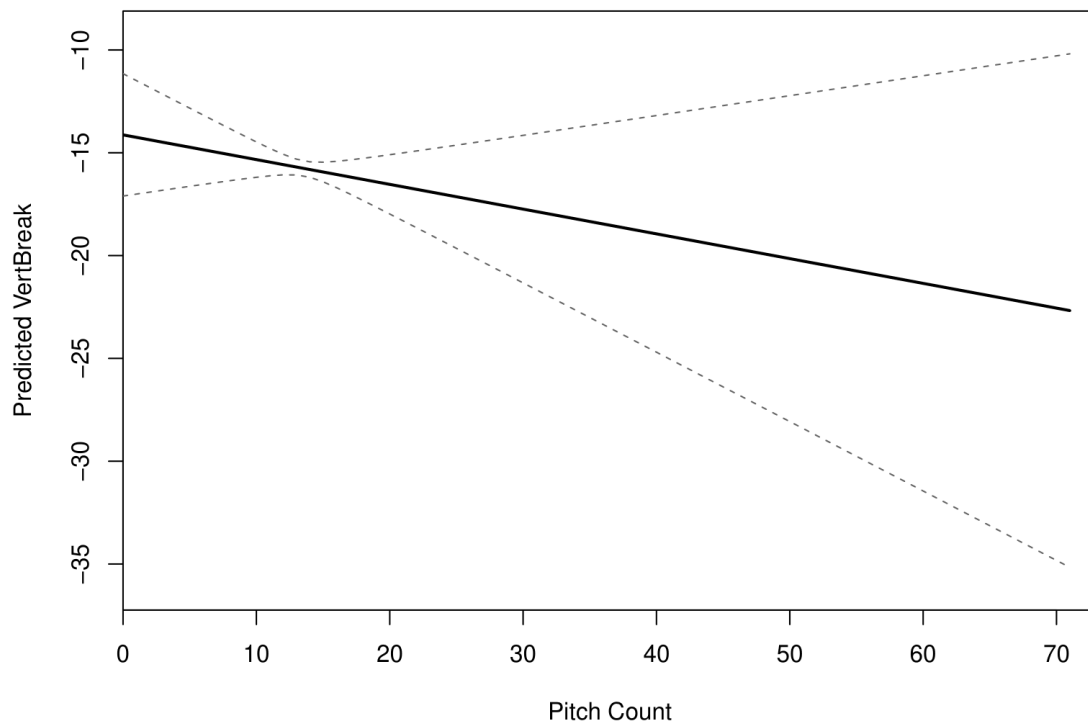




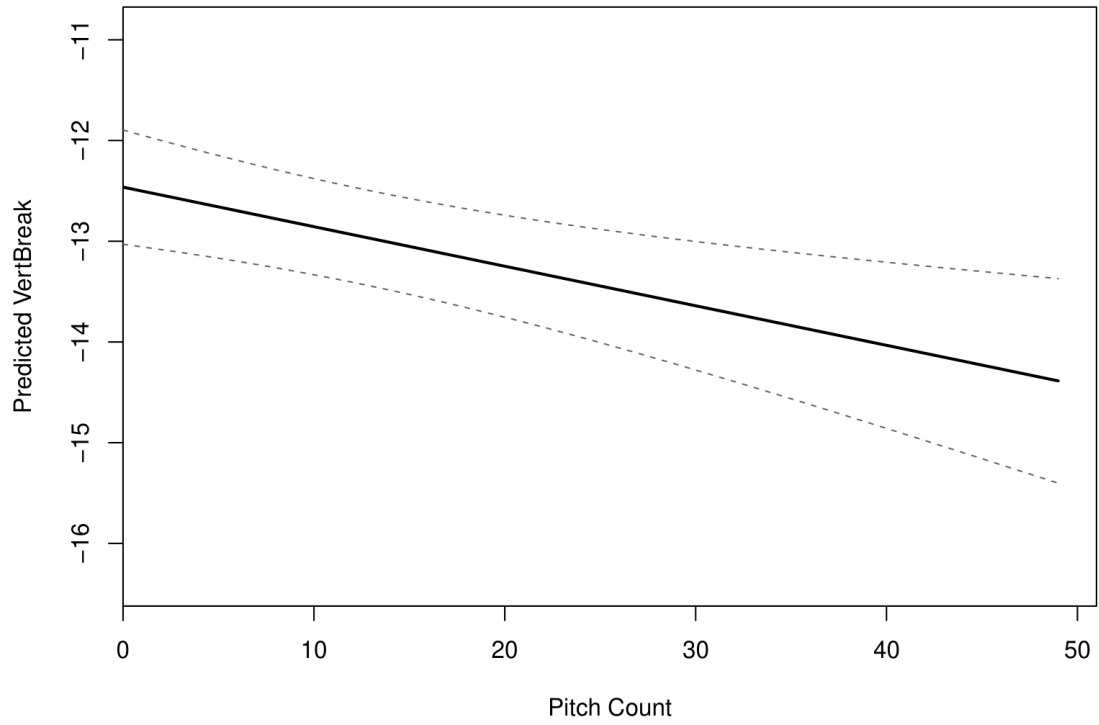
Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 5



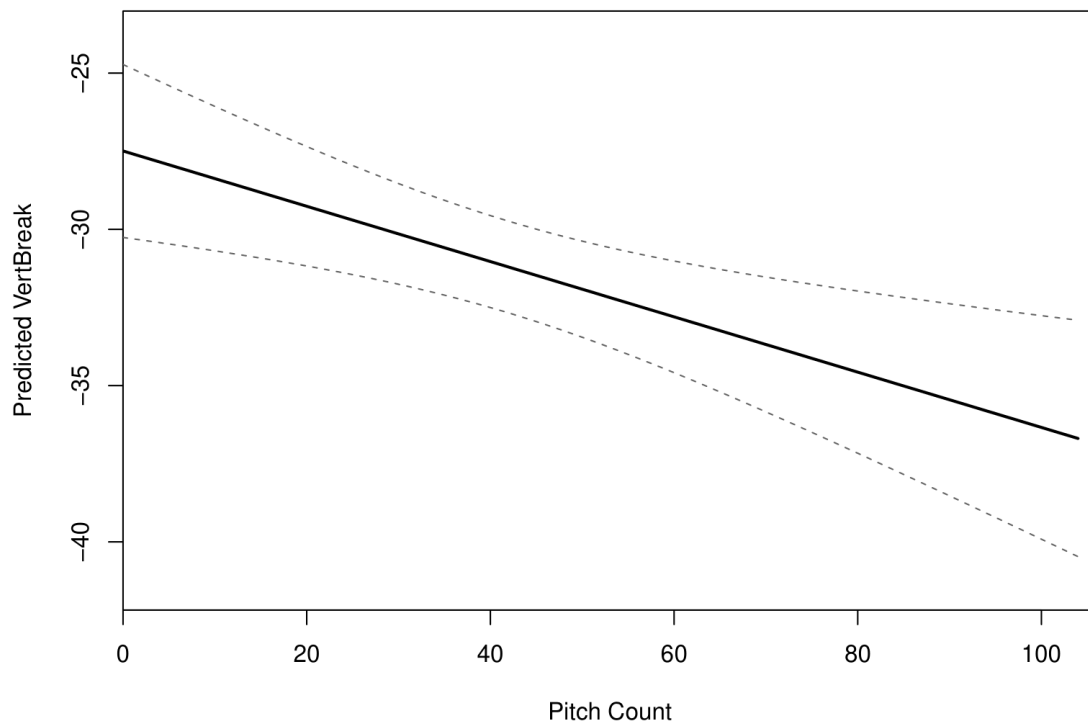
Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 8



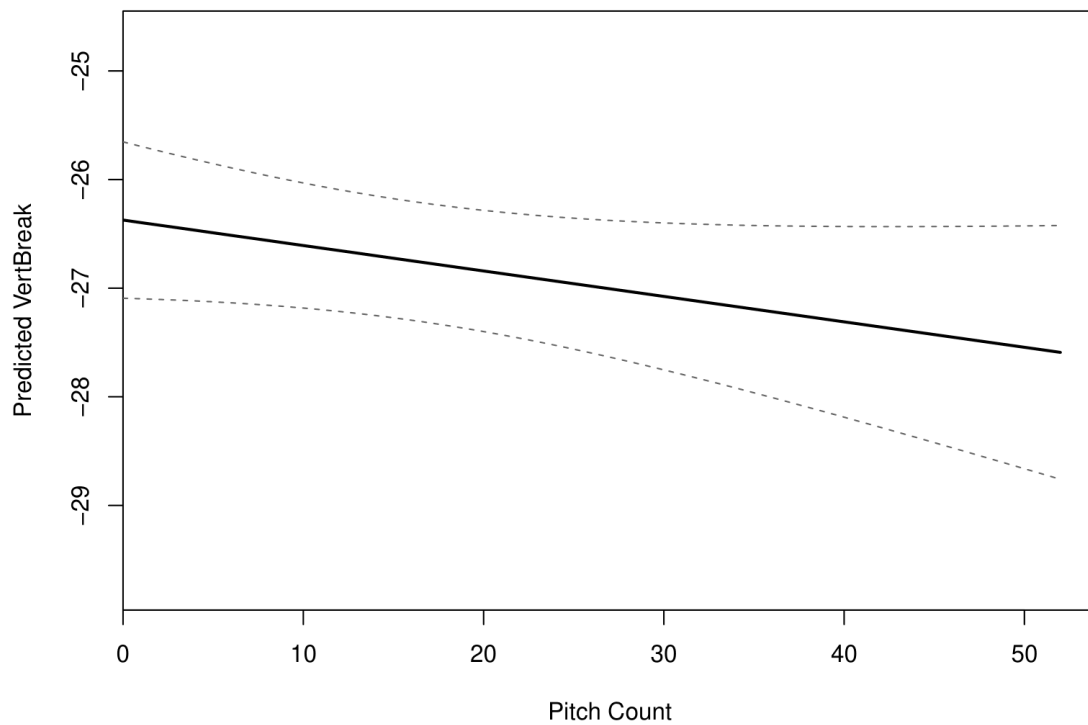
Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 11



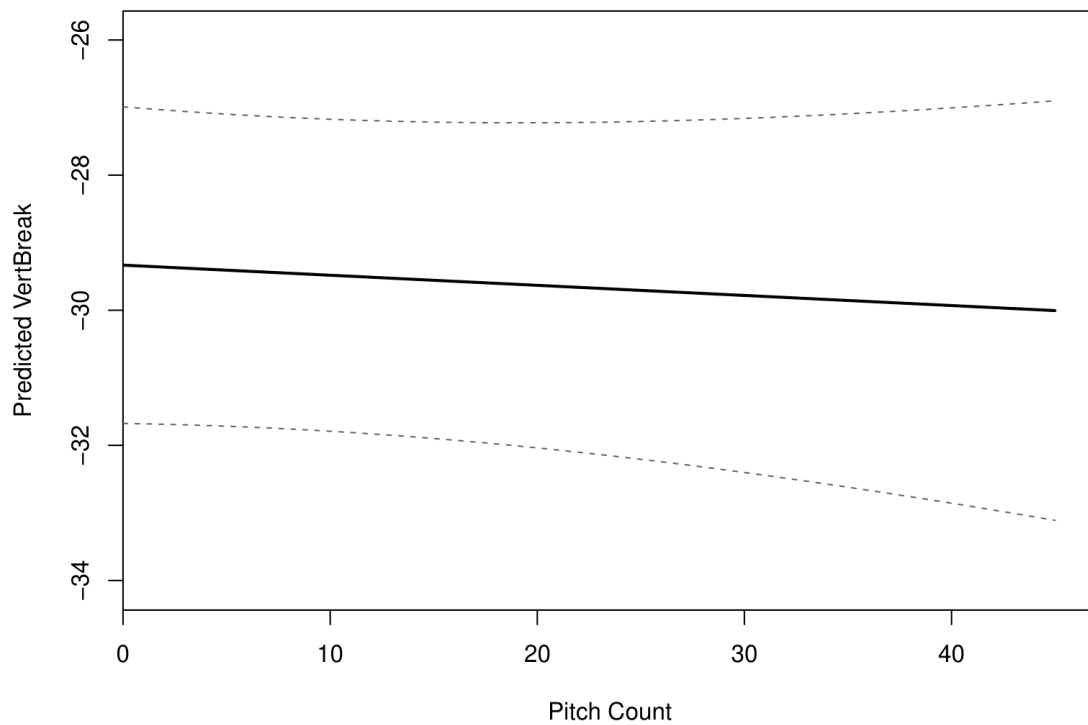
Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 12

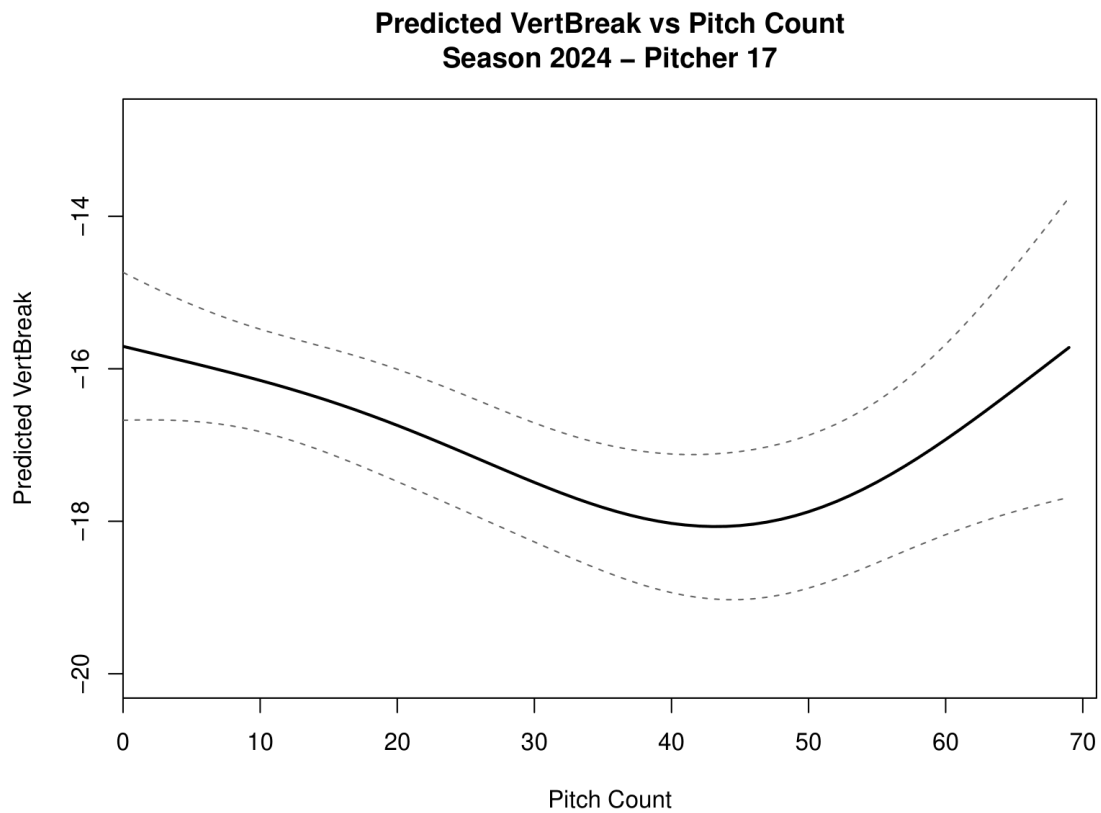


Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 14

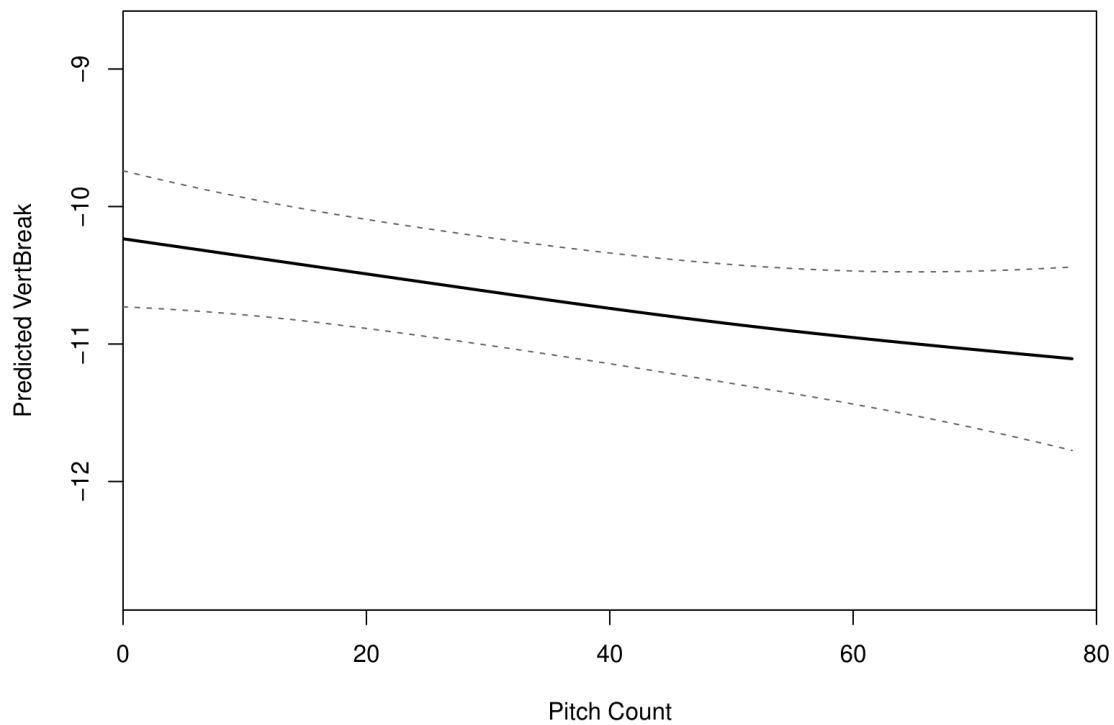


Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 15

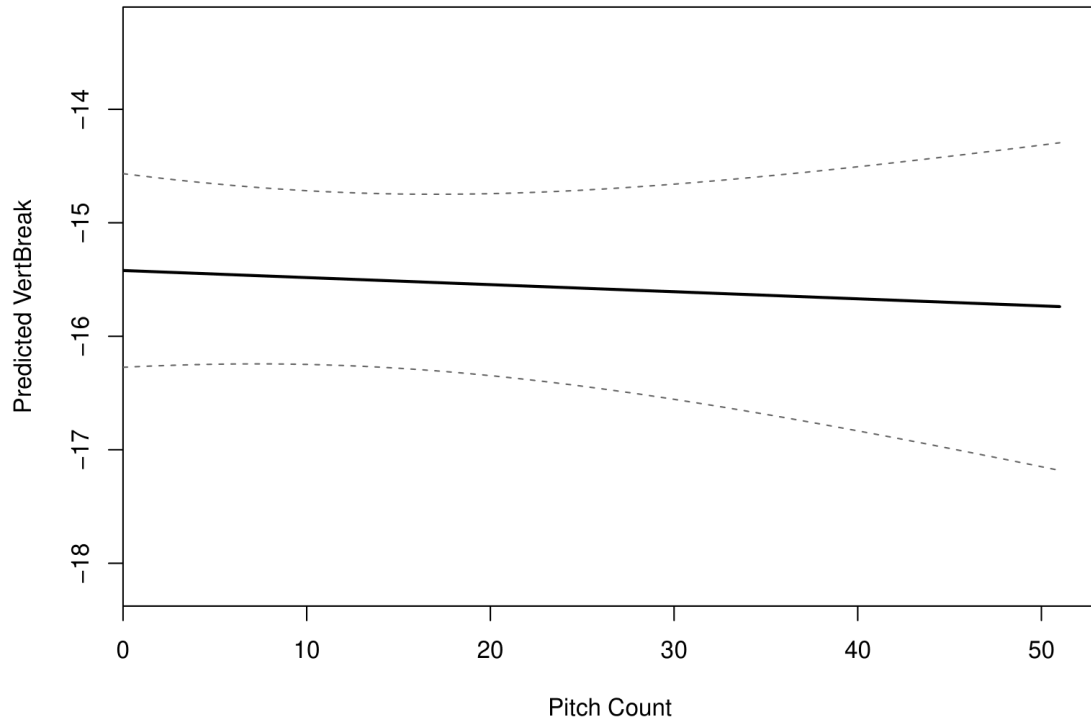




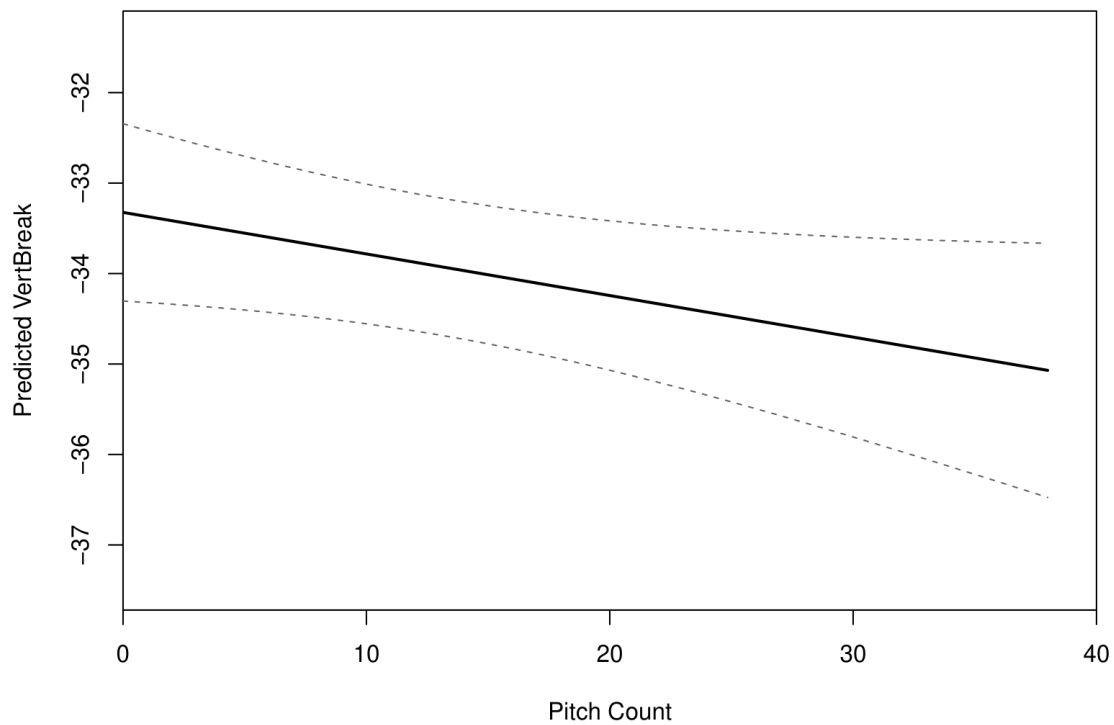
Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 19

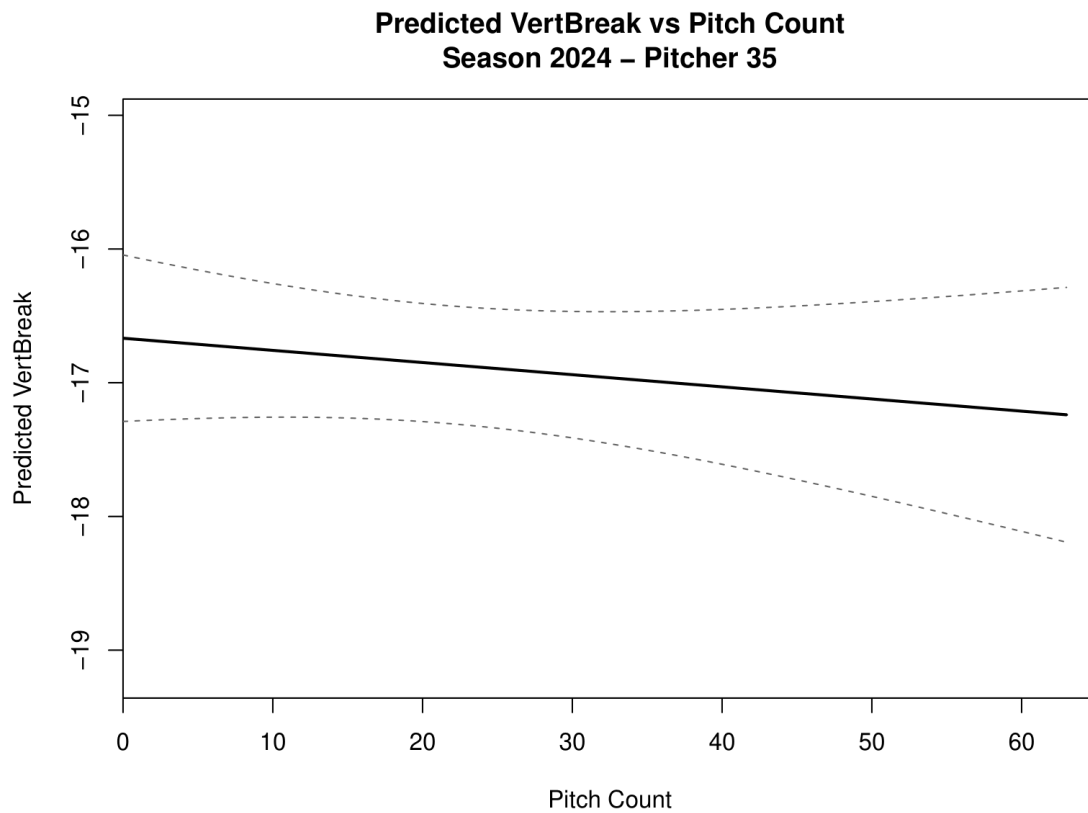


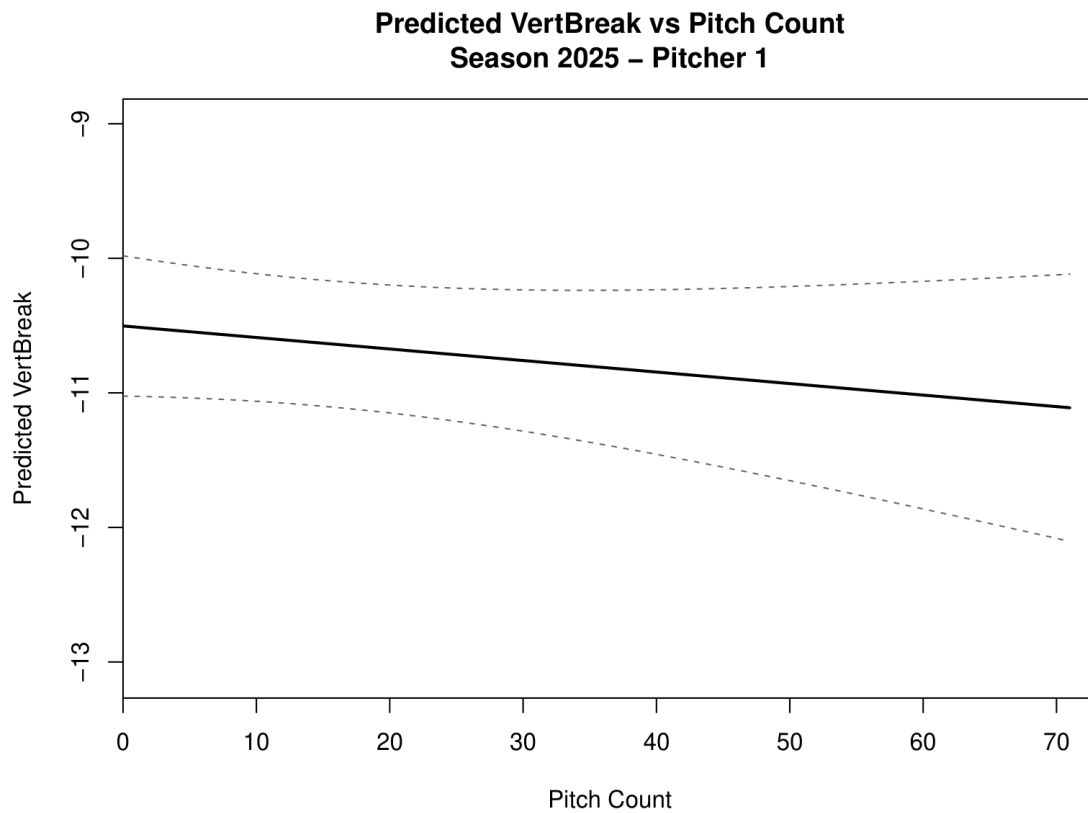
Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 21

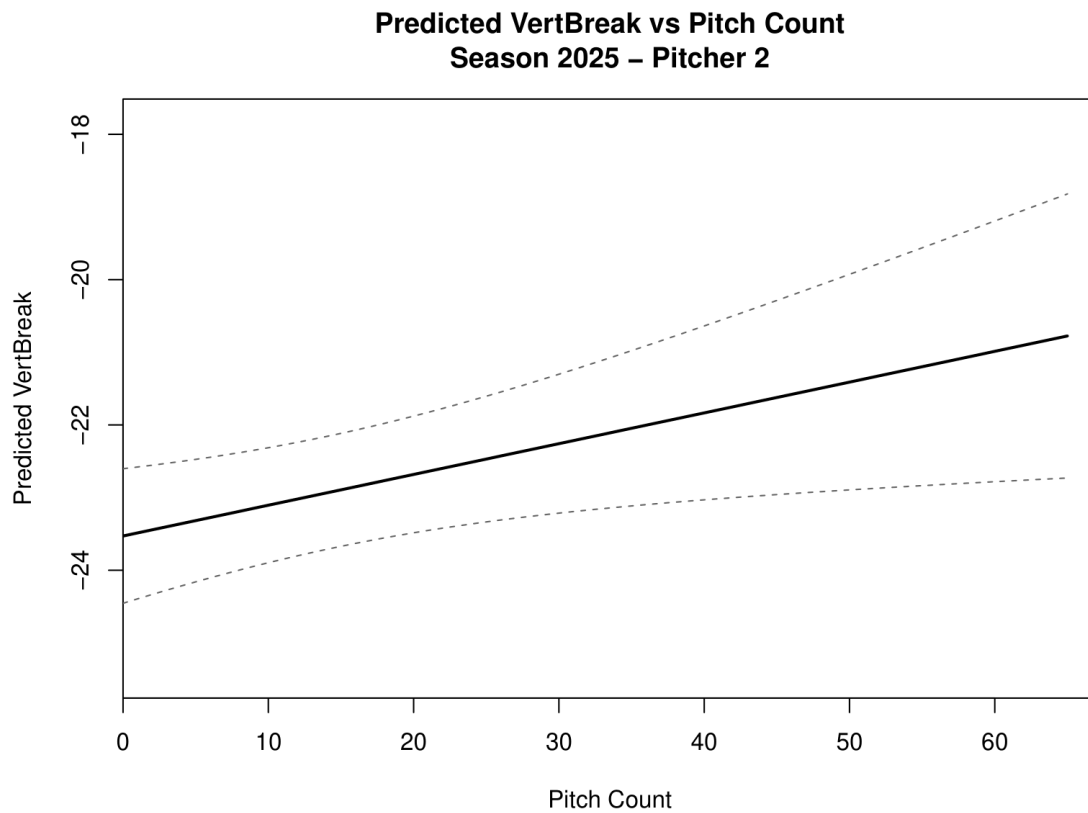


Predicted VertBreak vs Pitch Count
Season 2024 – Pitcher 33

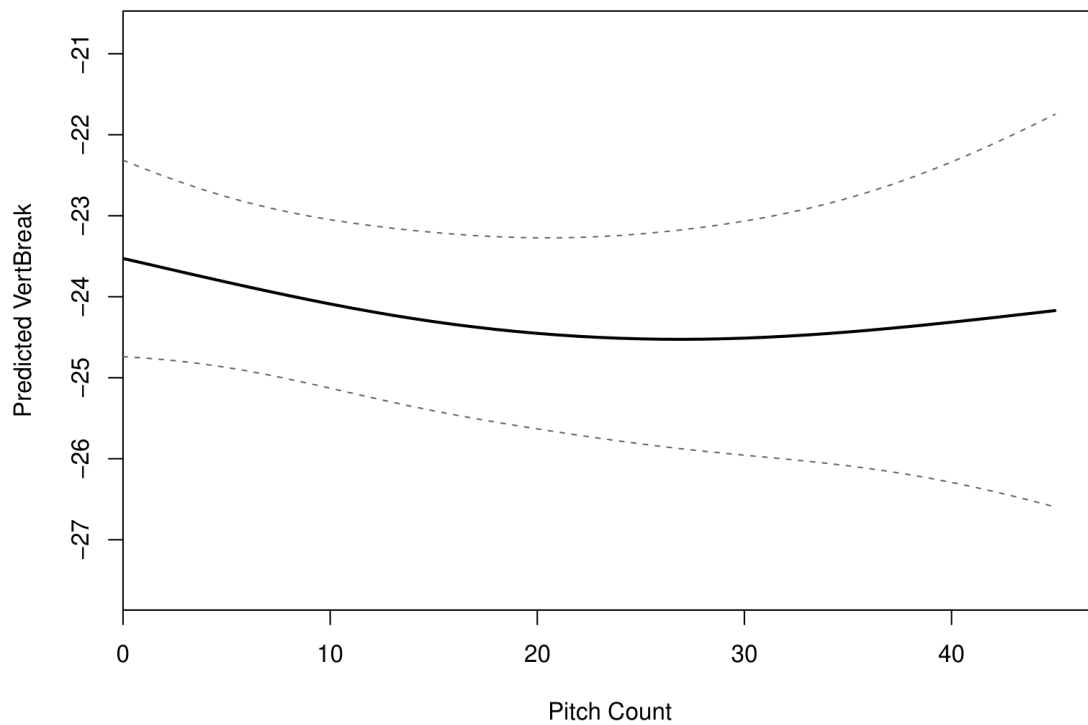




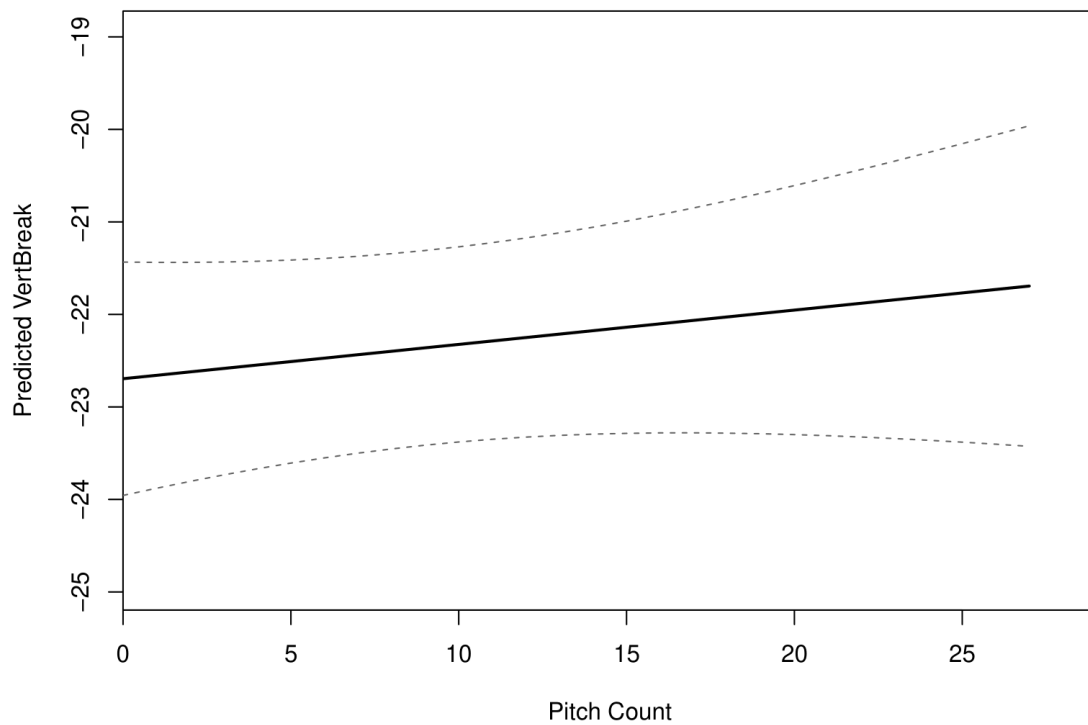


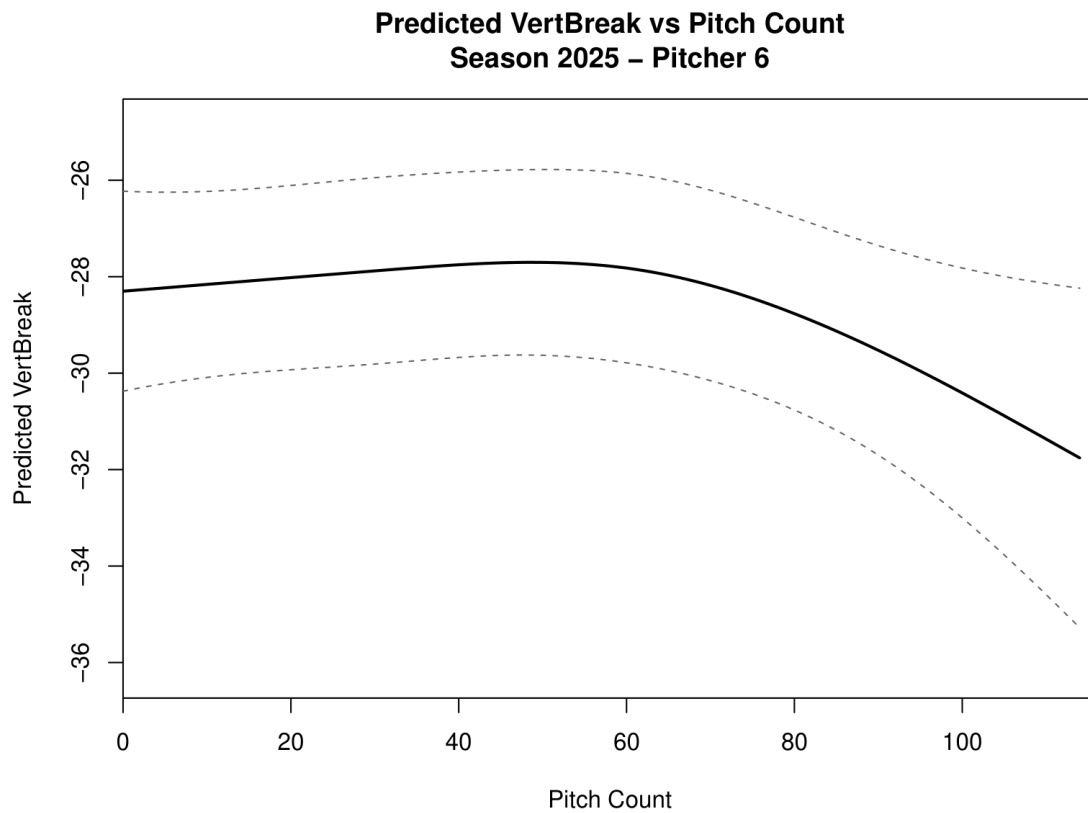


Predicted VertBreak vs Pitch Count
Season 2025 – Pitcher 3

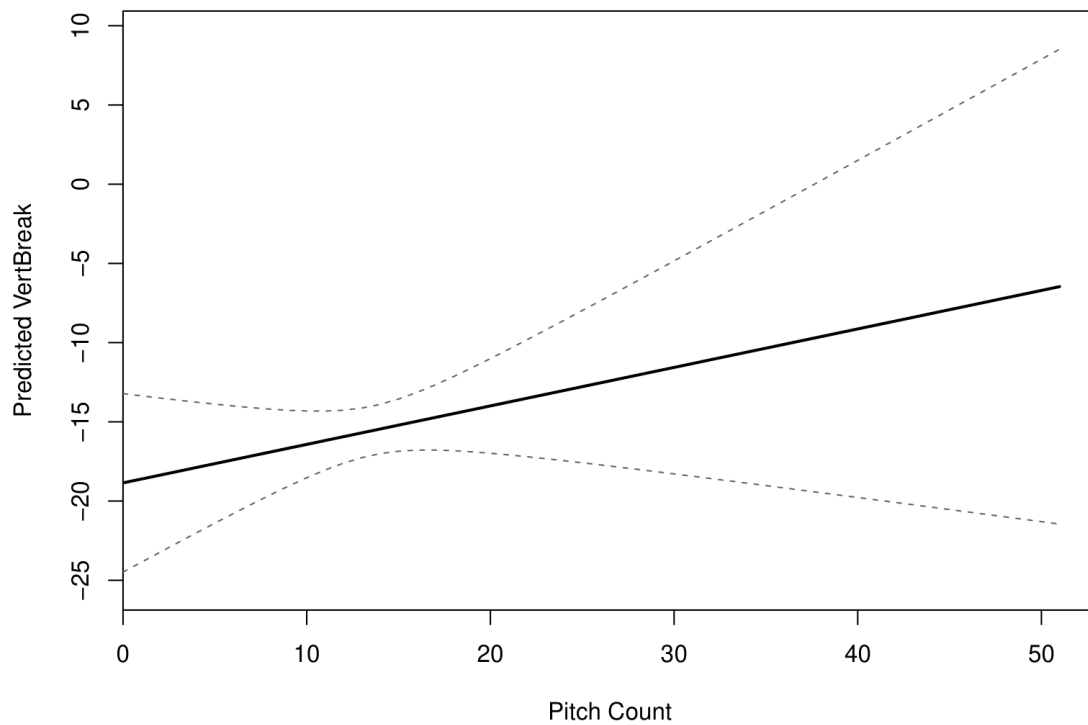


Predicted VertBreak vs Pitch Count
Season 2025 – Pitcher 4

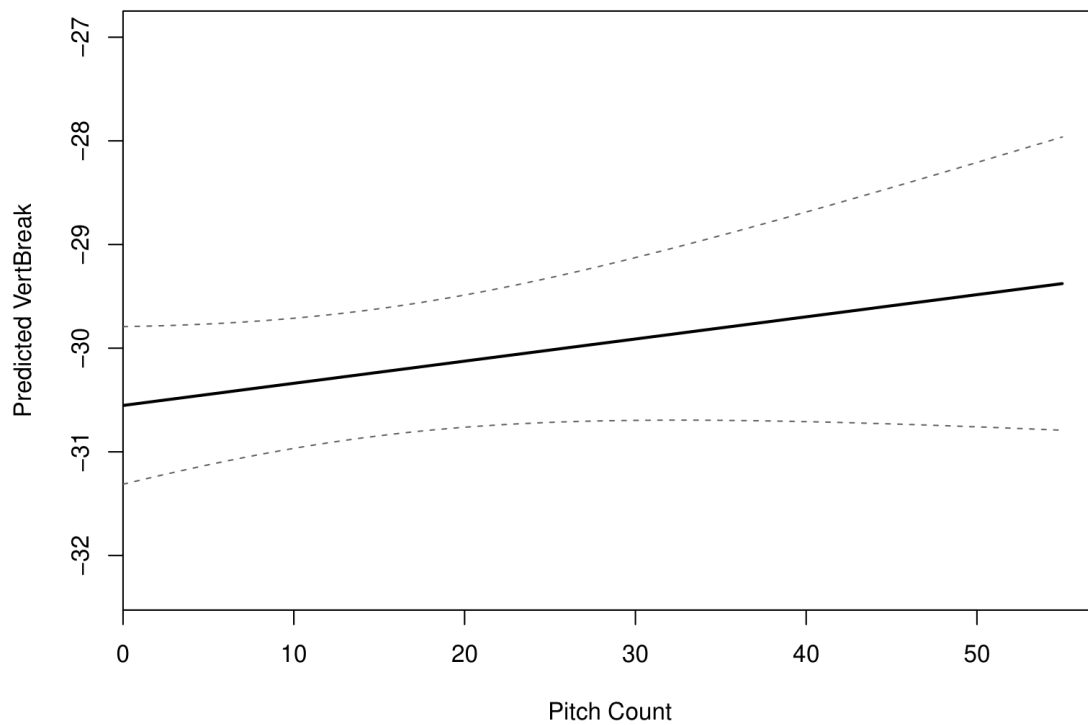


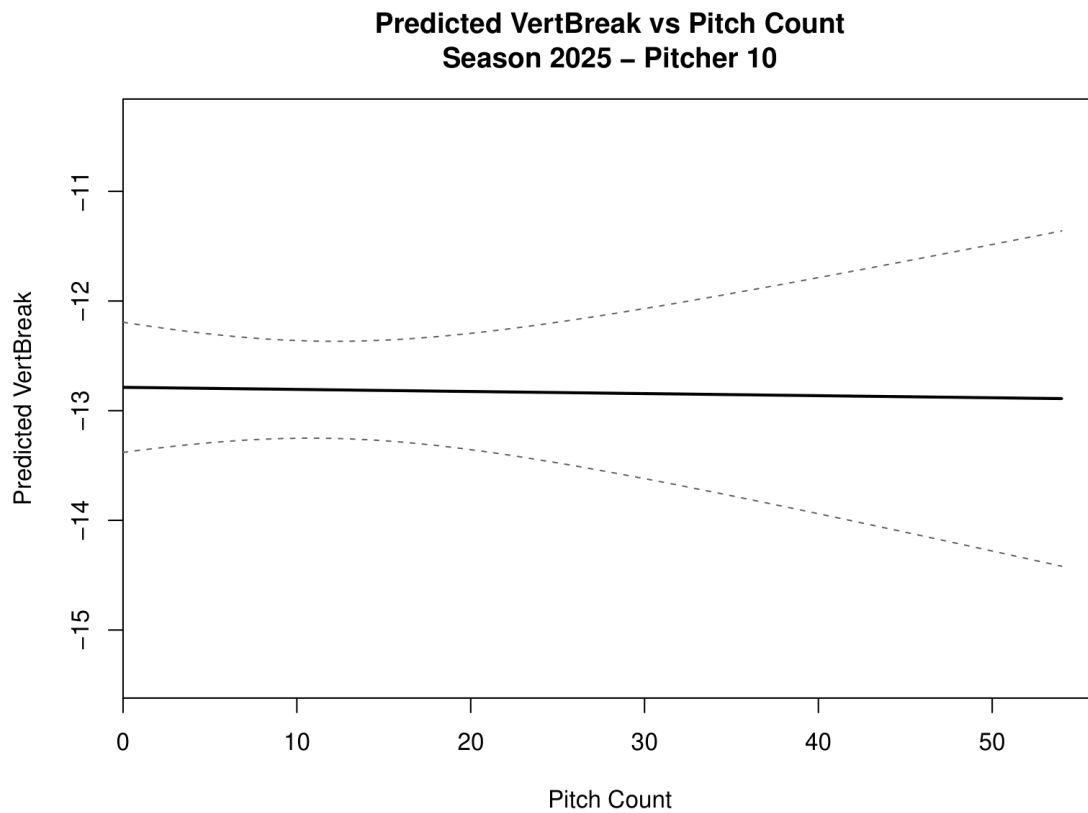


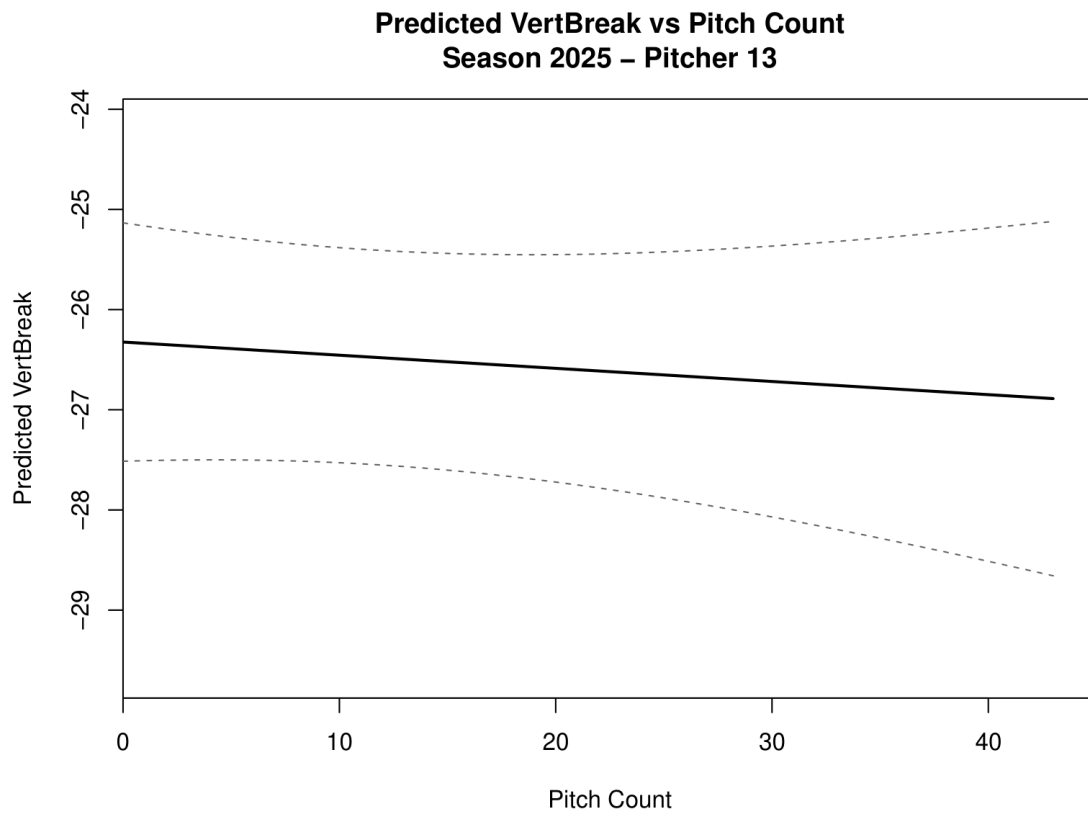
Predicted VertBreak vs Pitch Count
Season 2025 – Pitcher 7



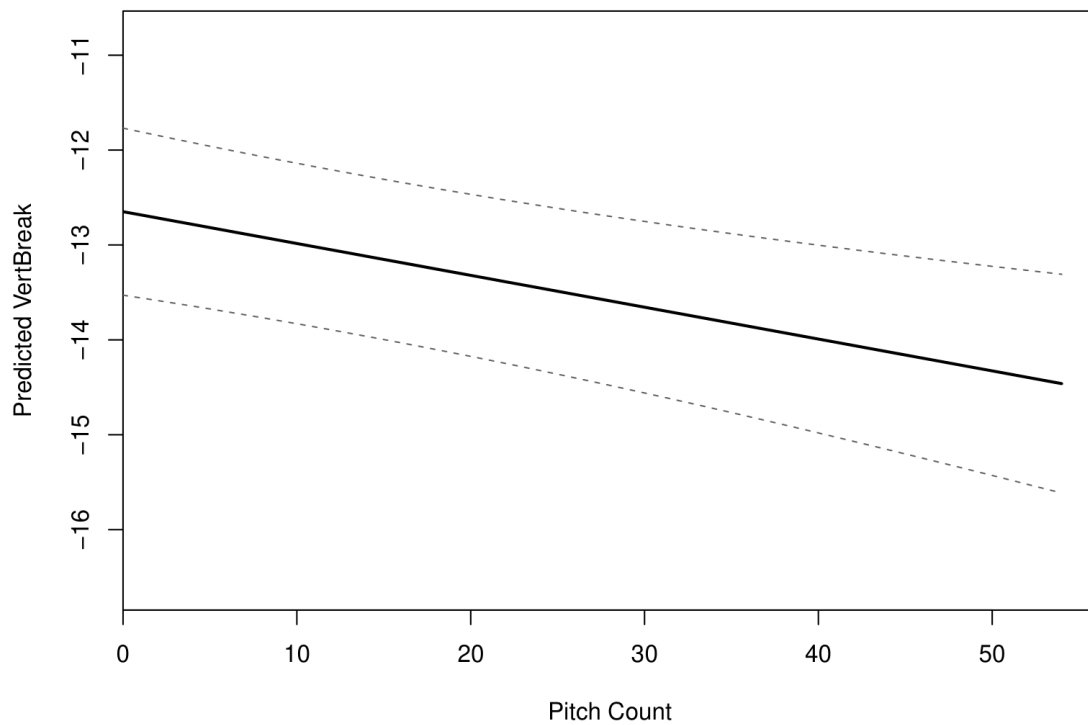
Predicted VertBreak vs Pitch Count
Season 2025 – Pitcher 9

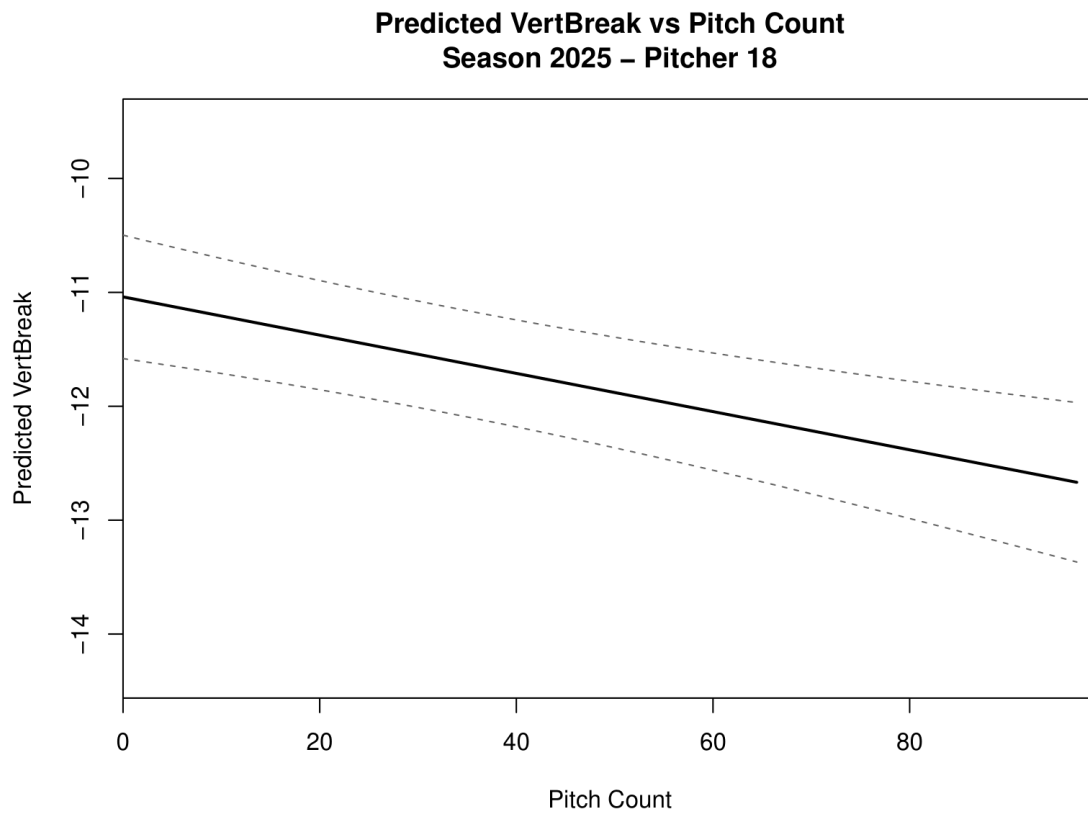




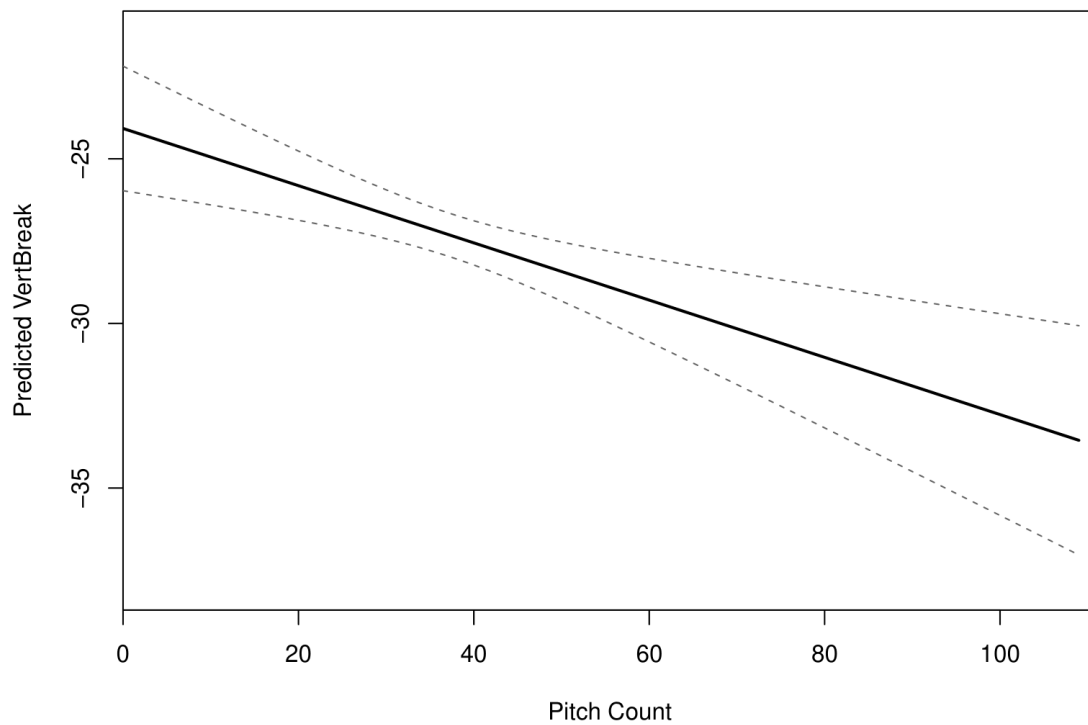


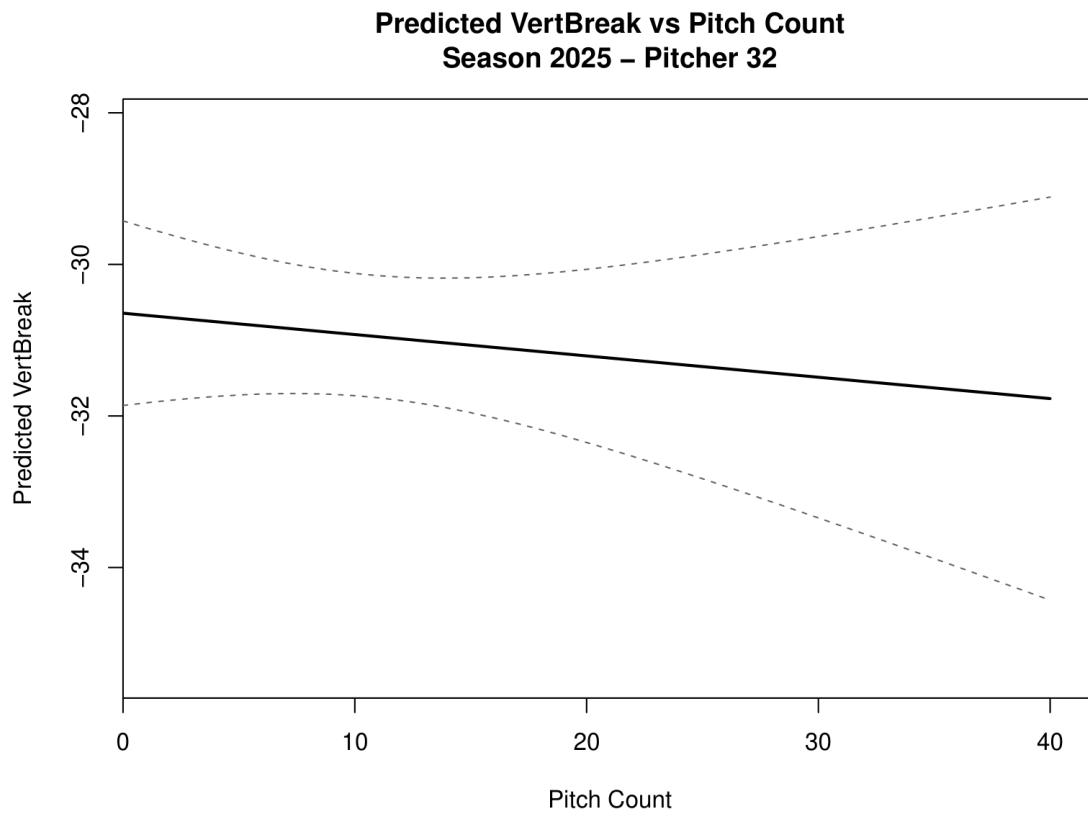
Predicted VertBreak vs Pitch Count
Season 2025 – Pitcher 16

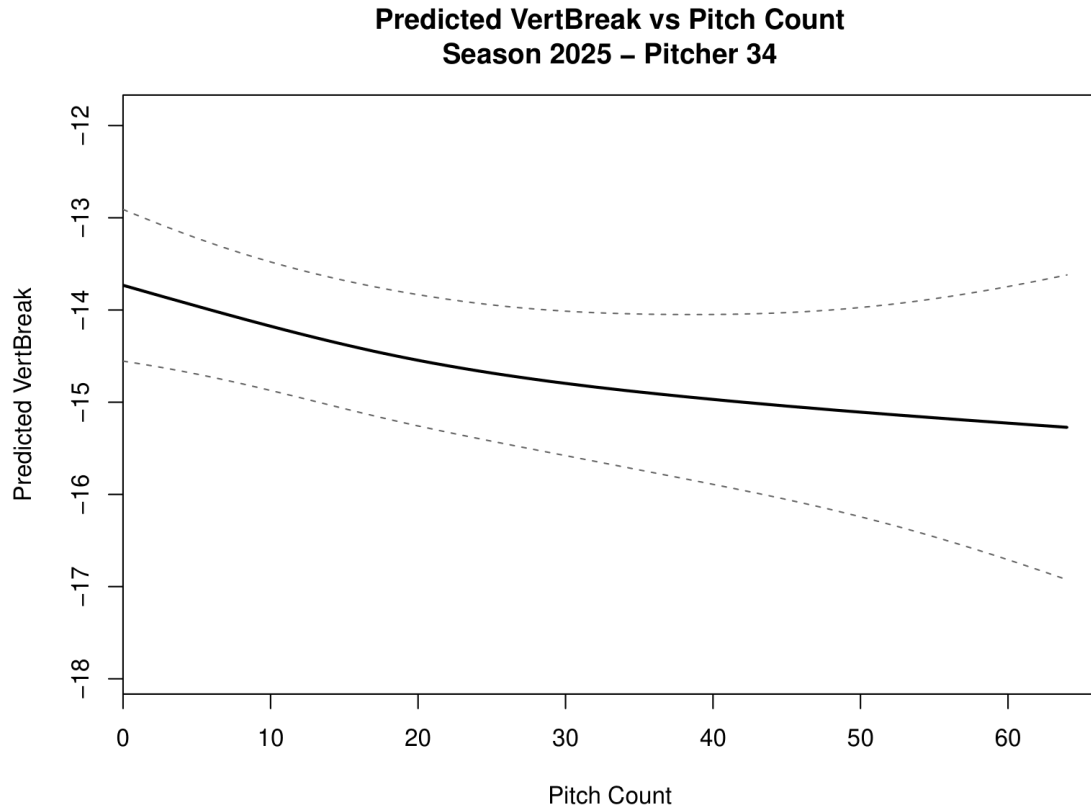




Predicted VertBreak vs Pitch Count
Season 2025 – Pitcher 20







Appendix B - Biomechanic Outcomes Descriptive Statistics

Biomechanic Outcomes Descriptive Statistics

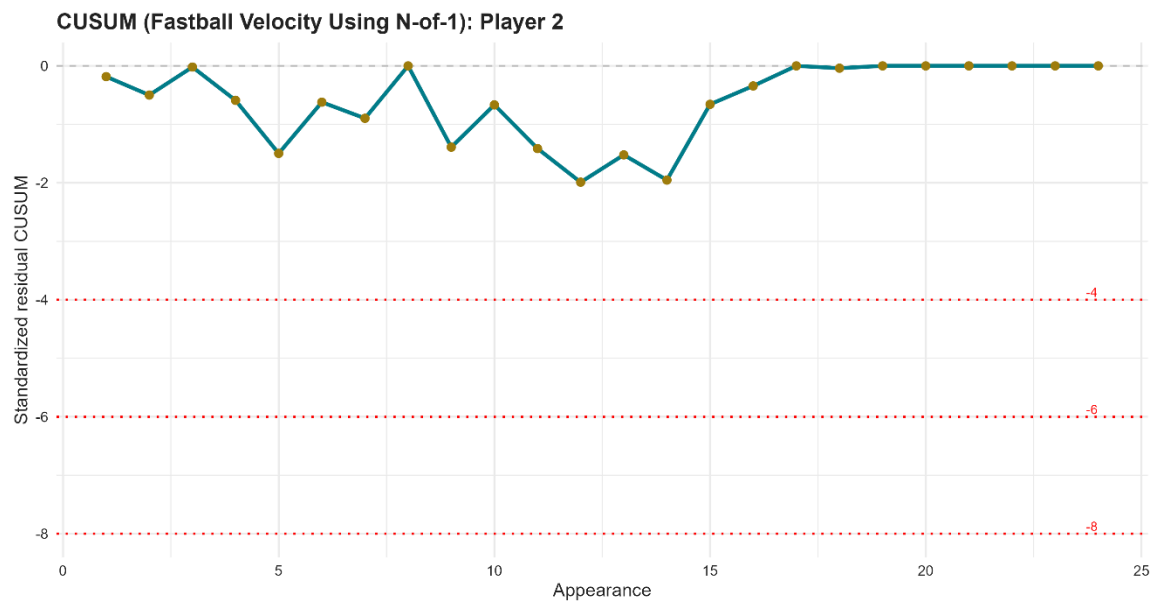
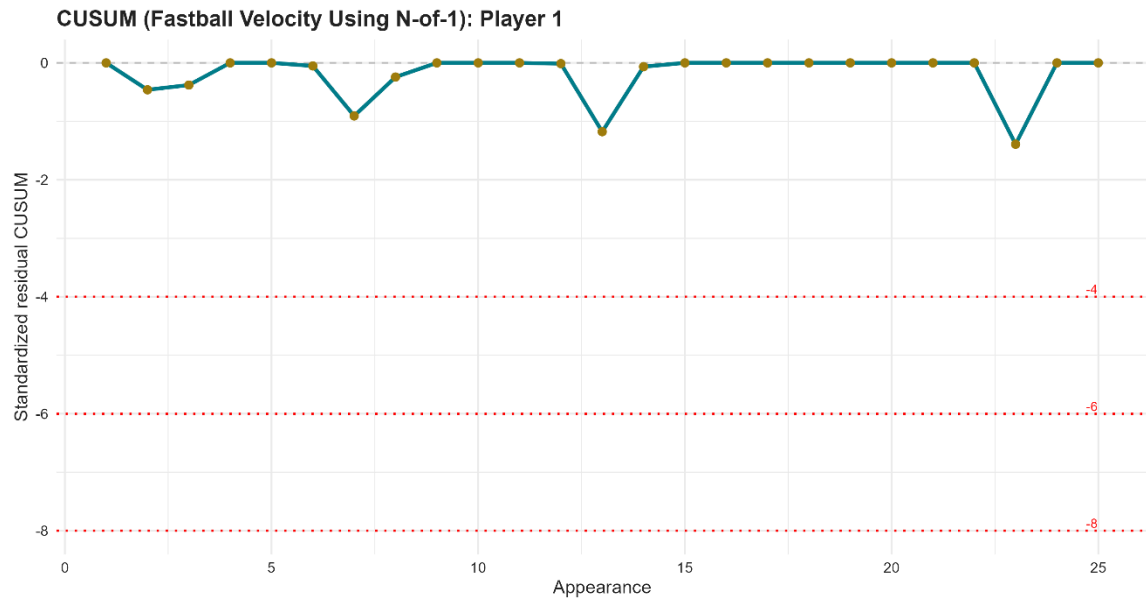
Biomechanic Outcome	2024, Mean (SD)	2025, Mean (SD)
Elbow Flexion Angle at BR	22.68 (5.36)	22.99 (4.35)
Lead Knee Flexion Angle at BR	44.20 (17.41)	42.64 (14.95)
Hip-Shoulder Separation at FC	55.50 (9.34)	58.94 (8.87)
Max Hip-Shoulder Separation	57.92 (7.47)	61.61 (9.75)

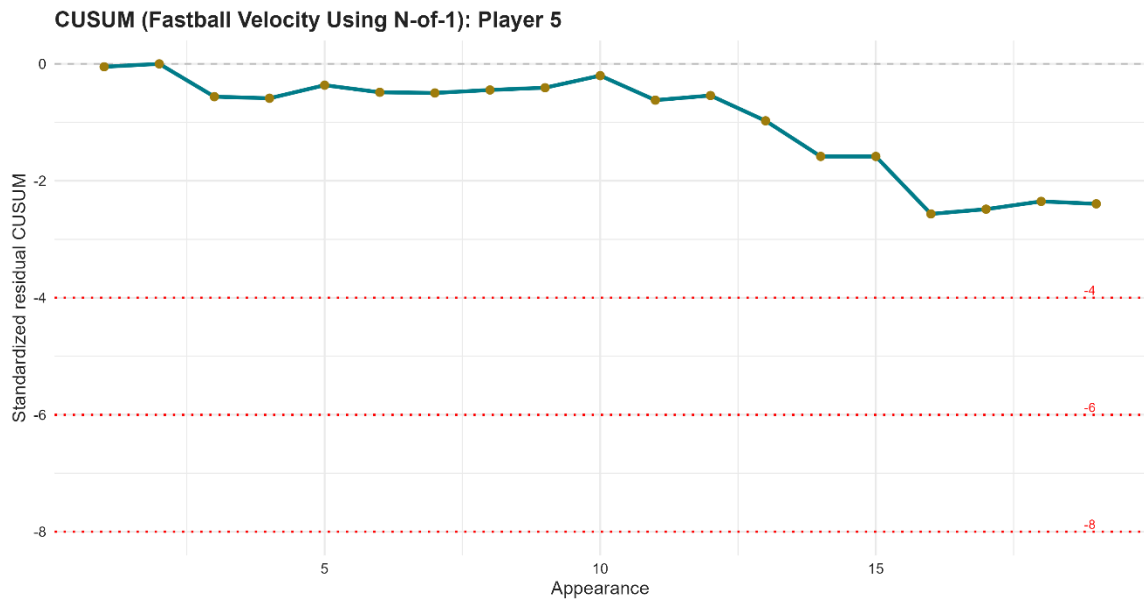
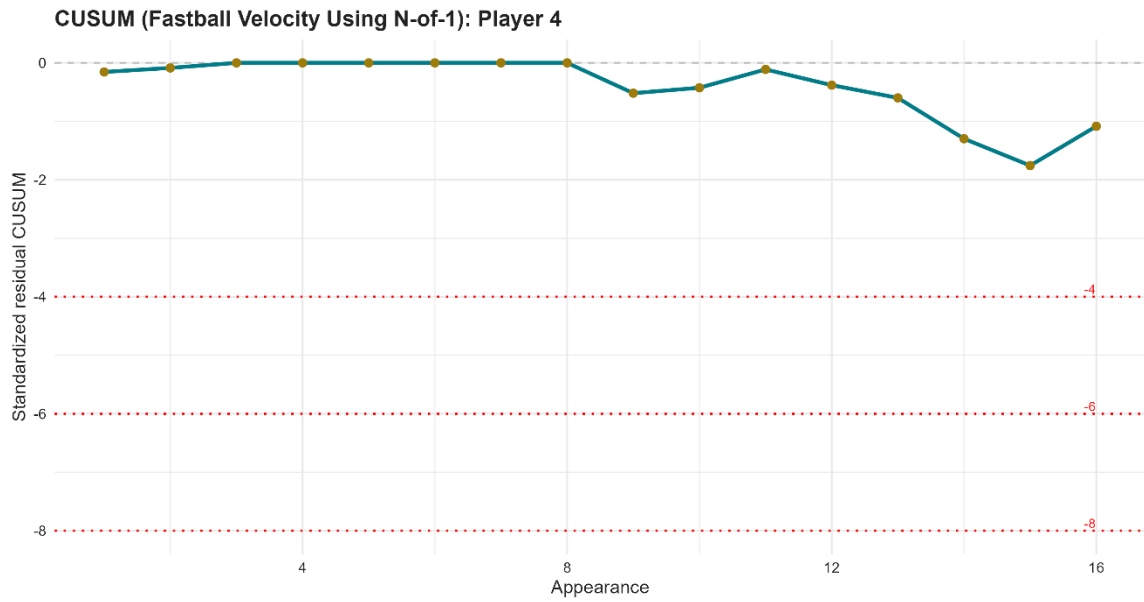
Shoulder Rotation Angle at MER	183.27 (7.32)	182.49 (5.30)
Shoulder Abduction Angle at BR	94.16 (6.96)	93.09 (5.39)
Max Resultant Shoulder Distraction Force	1402.69 (158.47)	1345.22 (159.23)
Elbow Flexion Angle at MER	81.21 (8.94)	80.04 (10.41)
Normalized Max Elbow Varus Torque	0.14 (0.03)	0.13 (0.02)
Max Shoulder Rotational Torque	110.80 (26.80)	115.23 (35.50)
Max Elbow Varus Torque	135.29 (23.99)	120.89 (18.02)

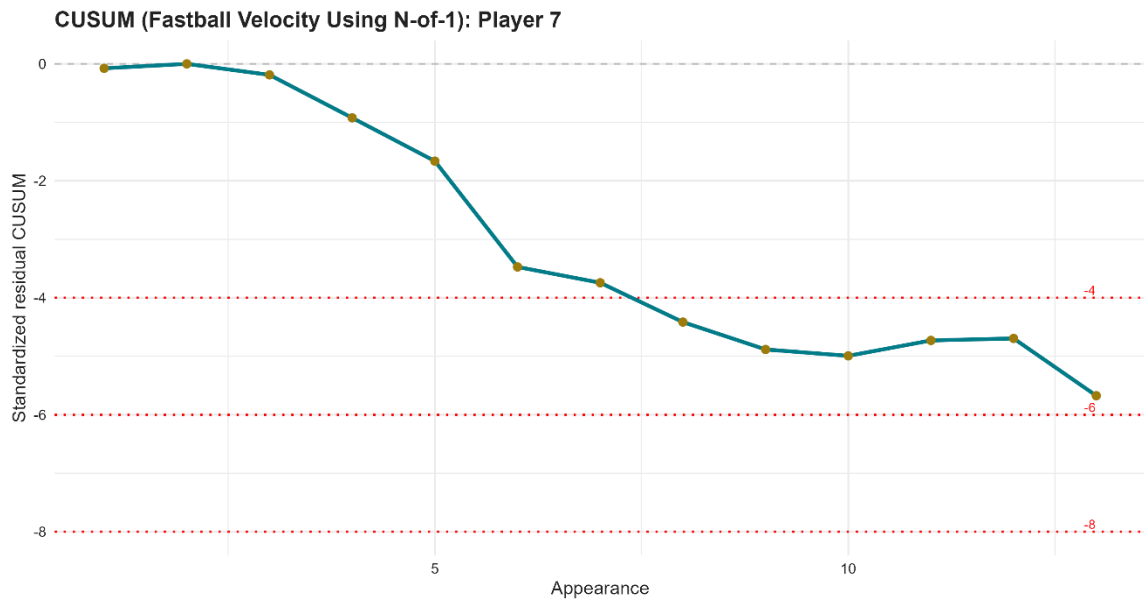
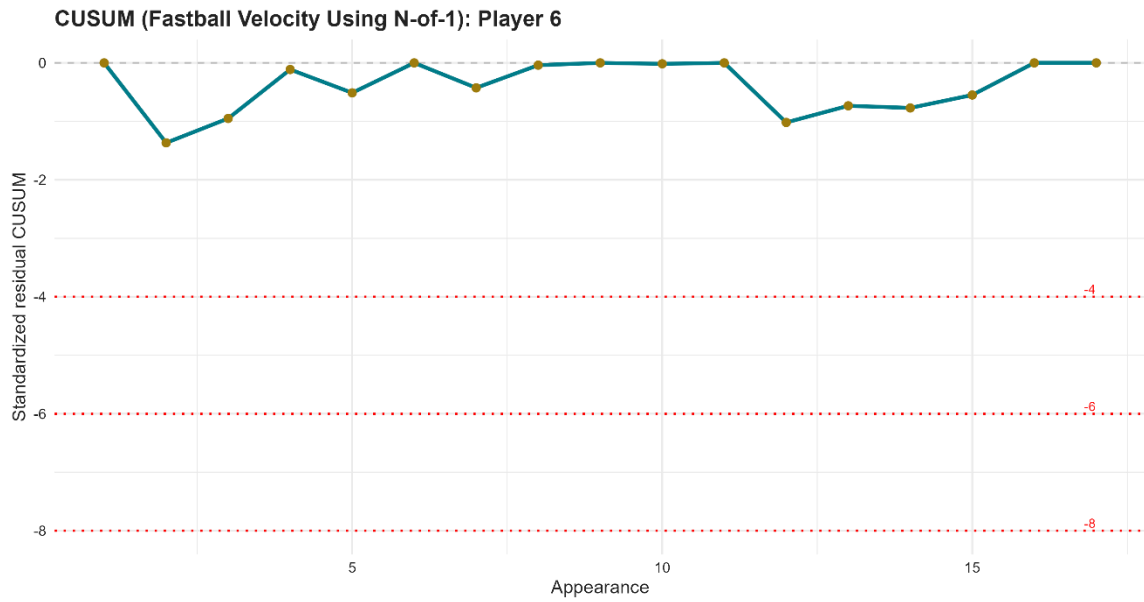
Appendix C – Additional CUSUM Plots

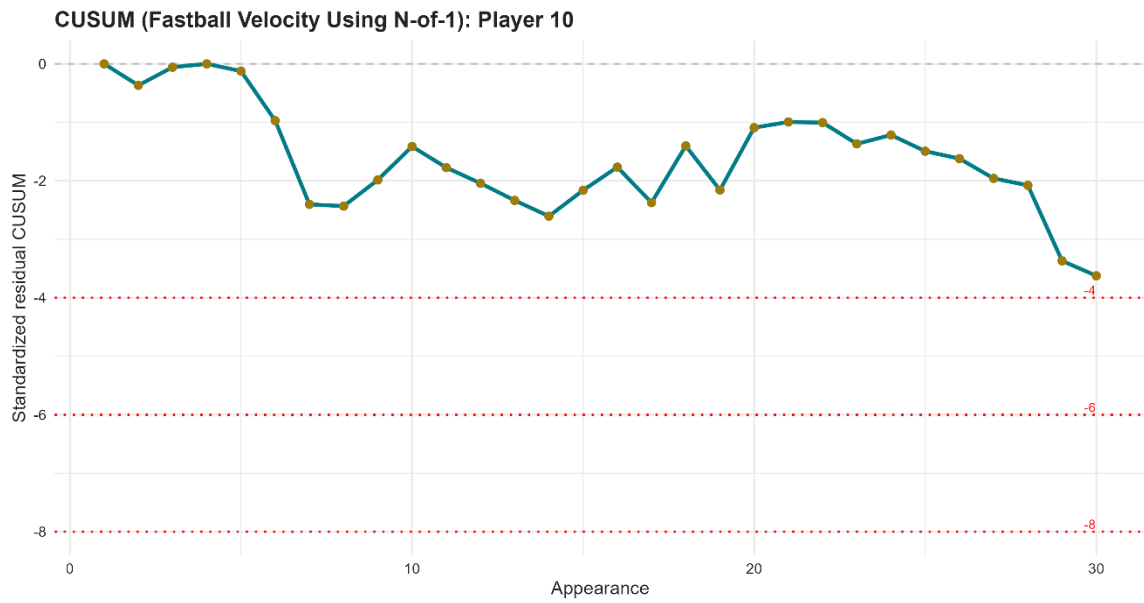
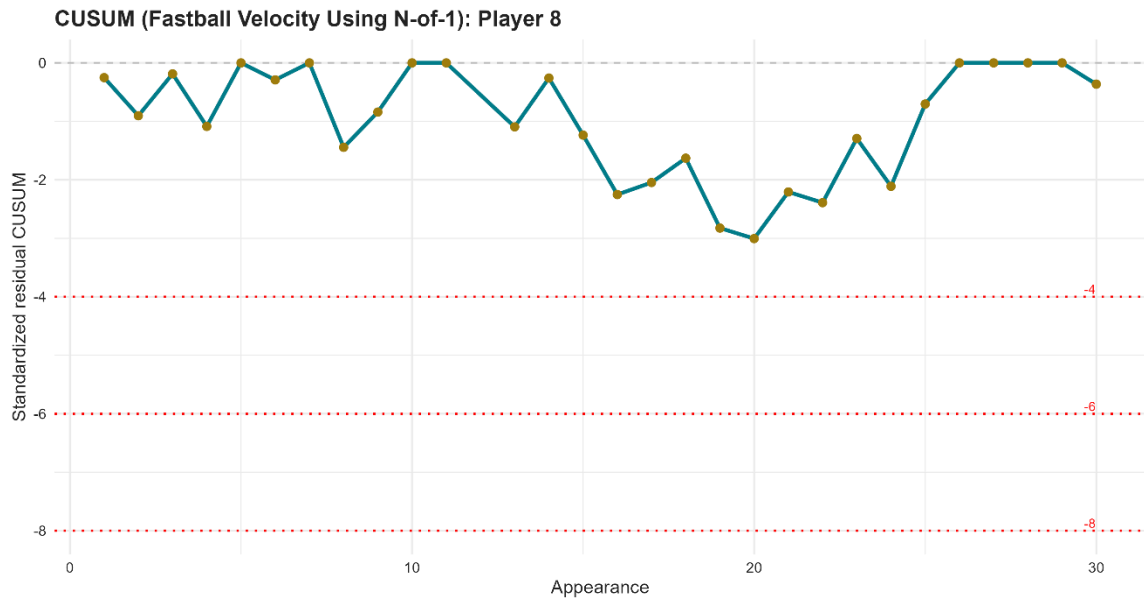
CUSUM Plots

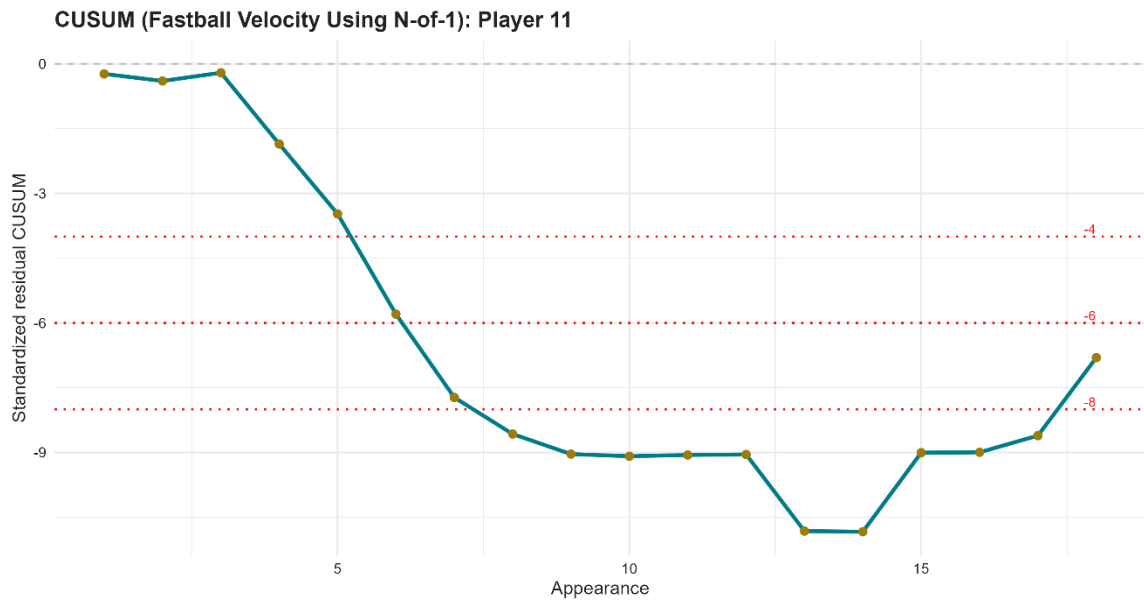
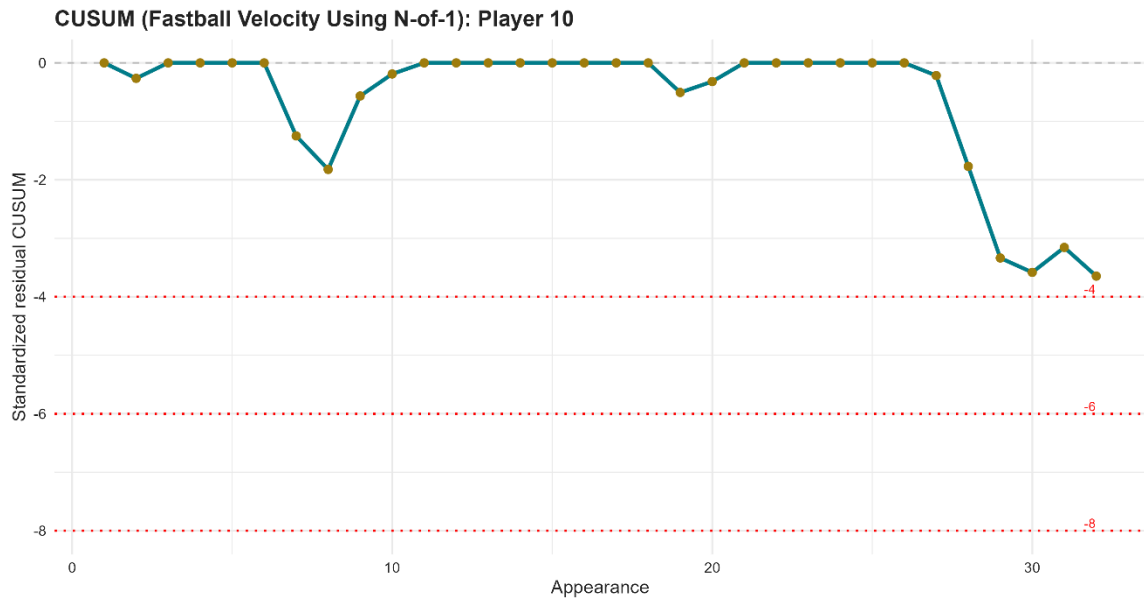
Some players have multiple plots from playing multiple seasons. CUSUM and models both treated pitcher-season as the main analytical unit.

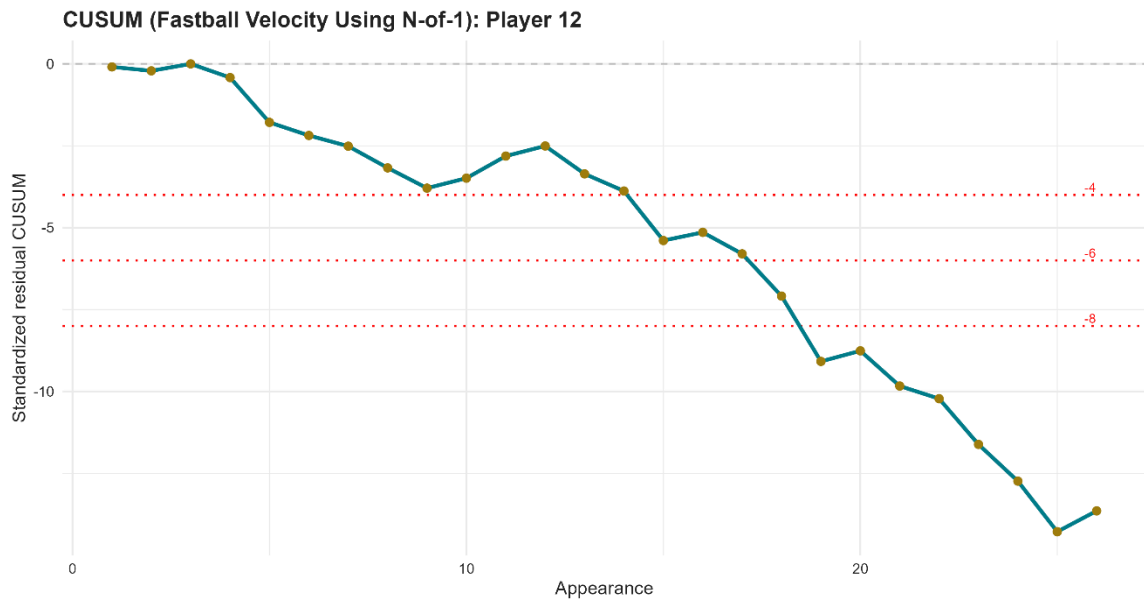
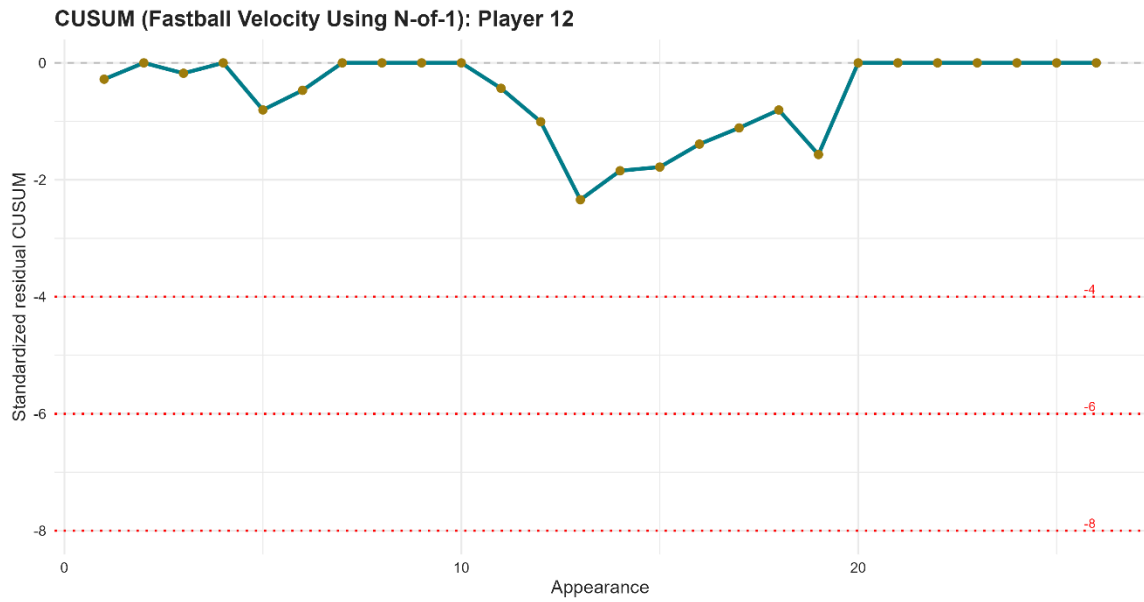


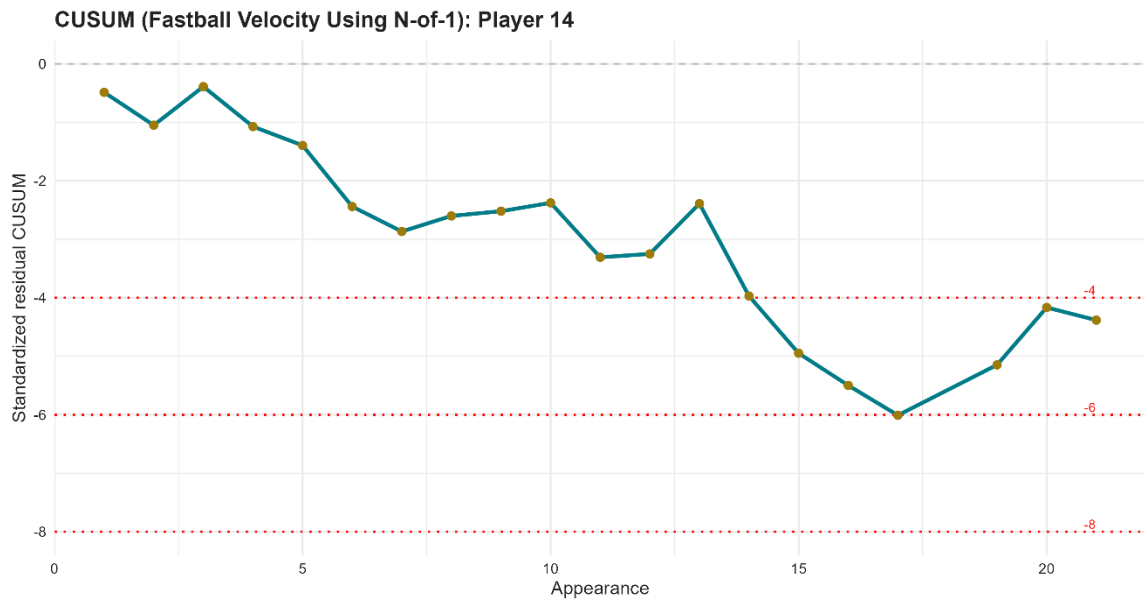
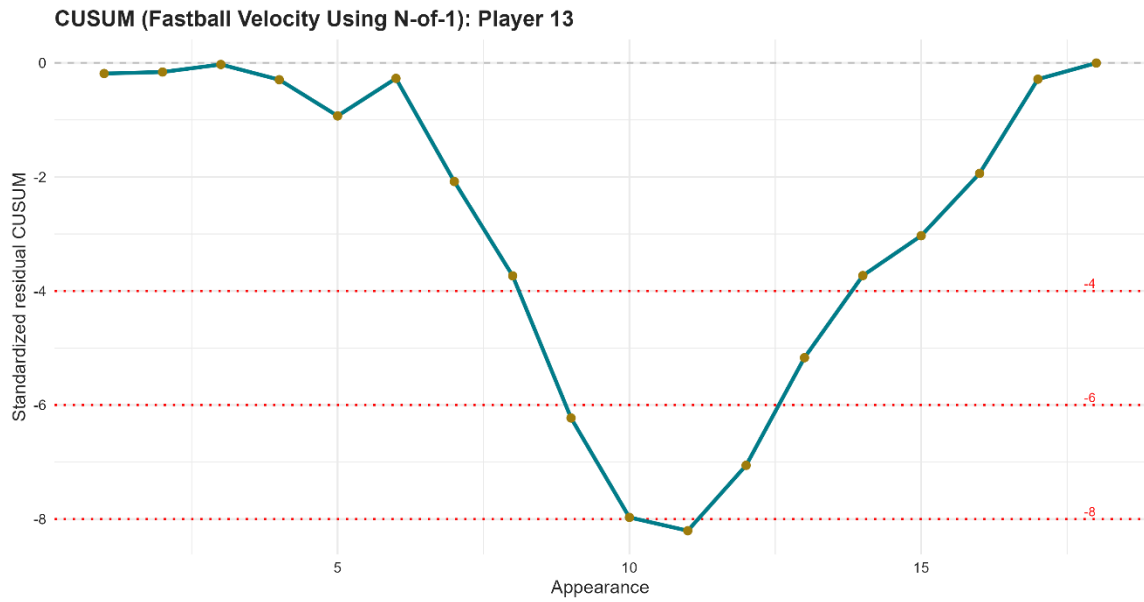


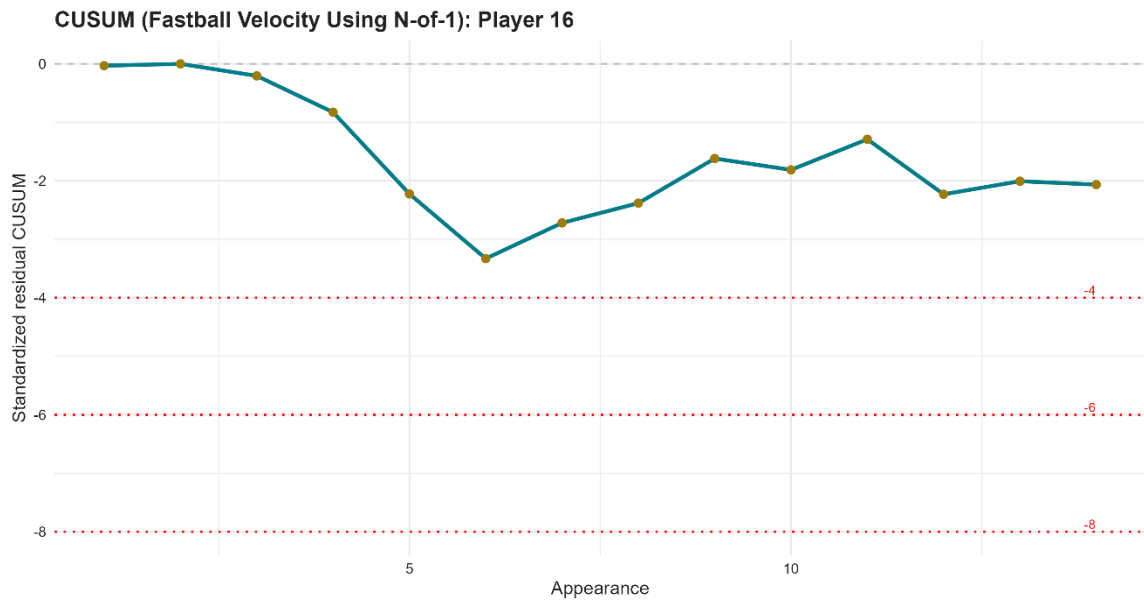
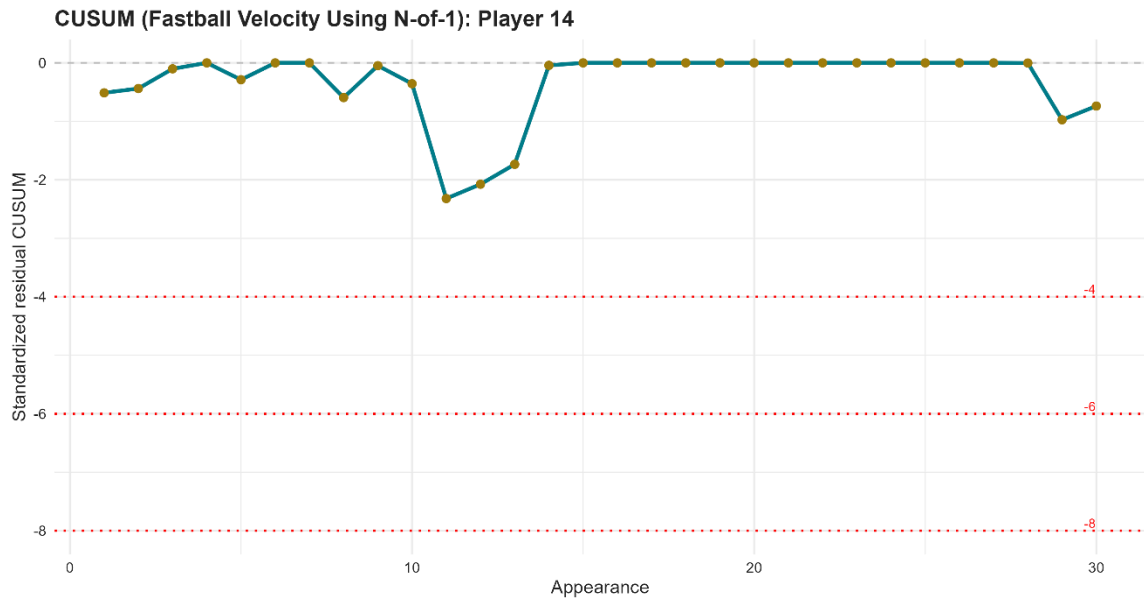


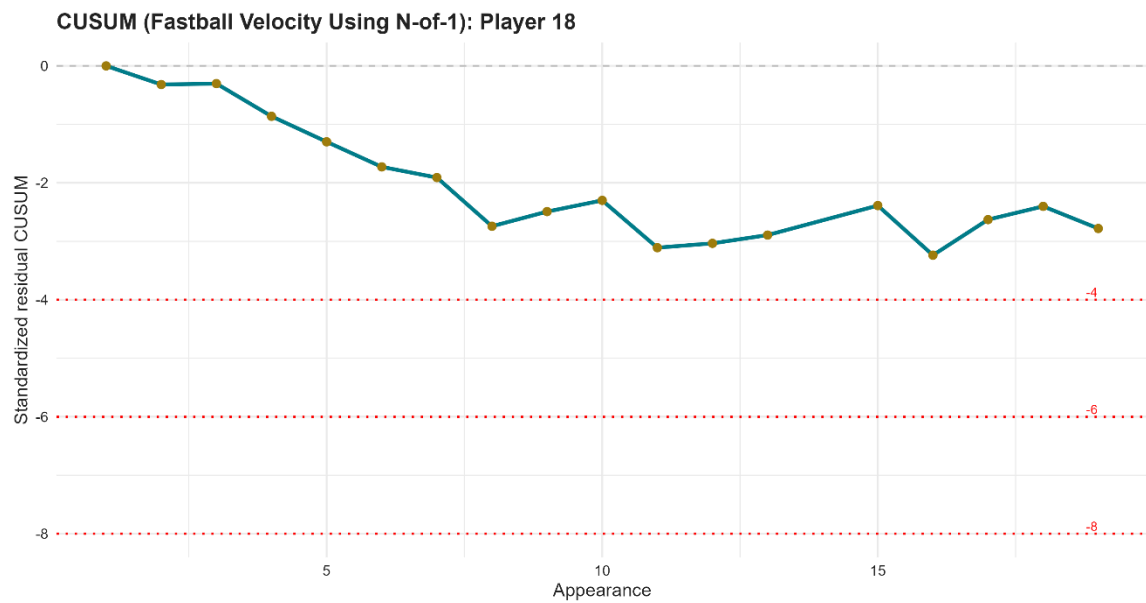
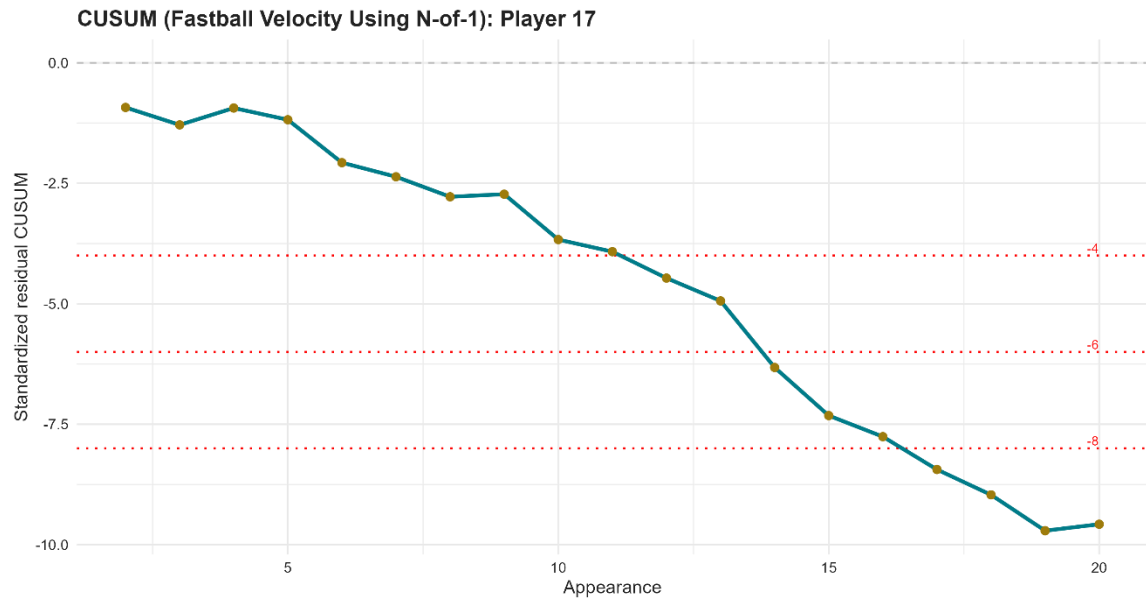


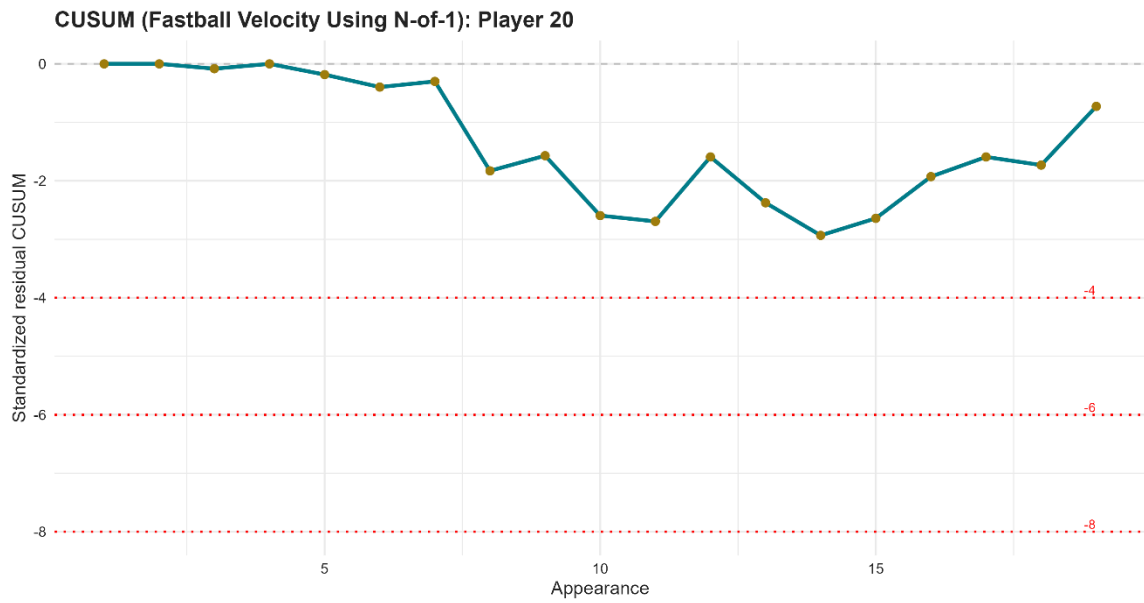
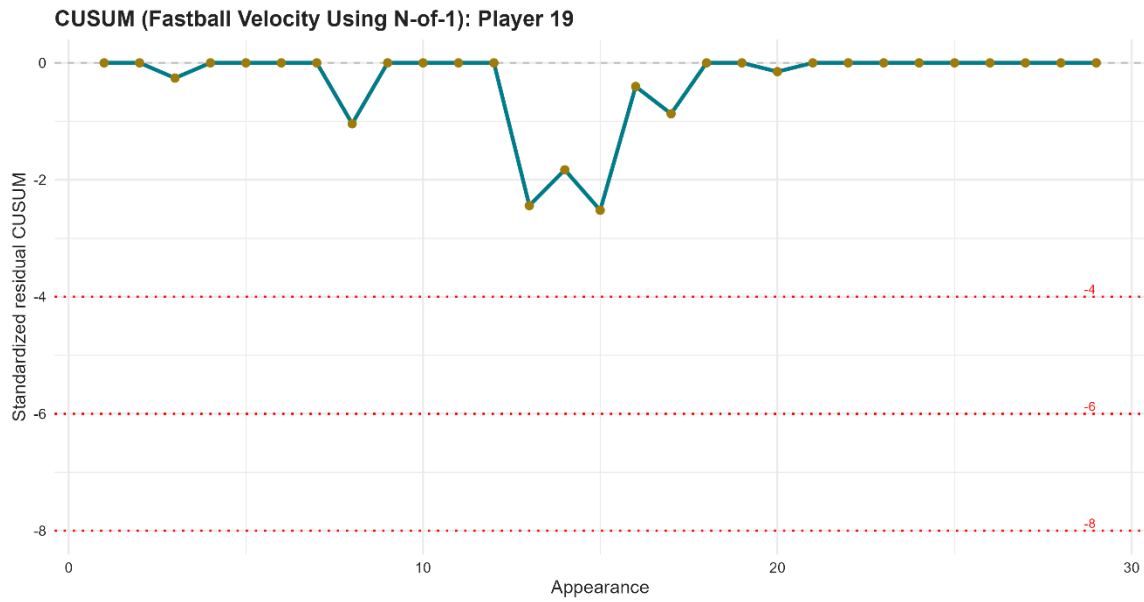


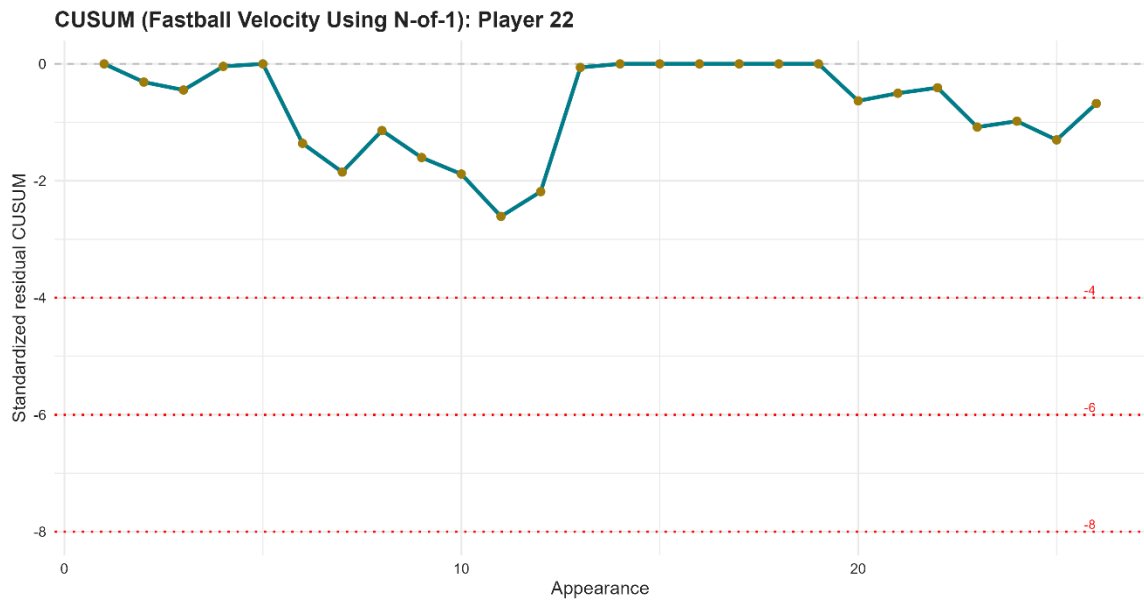
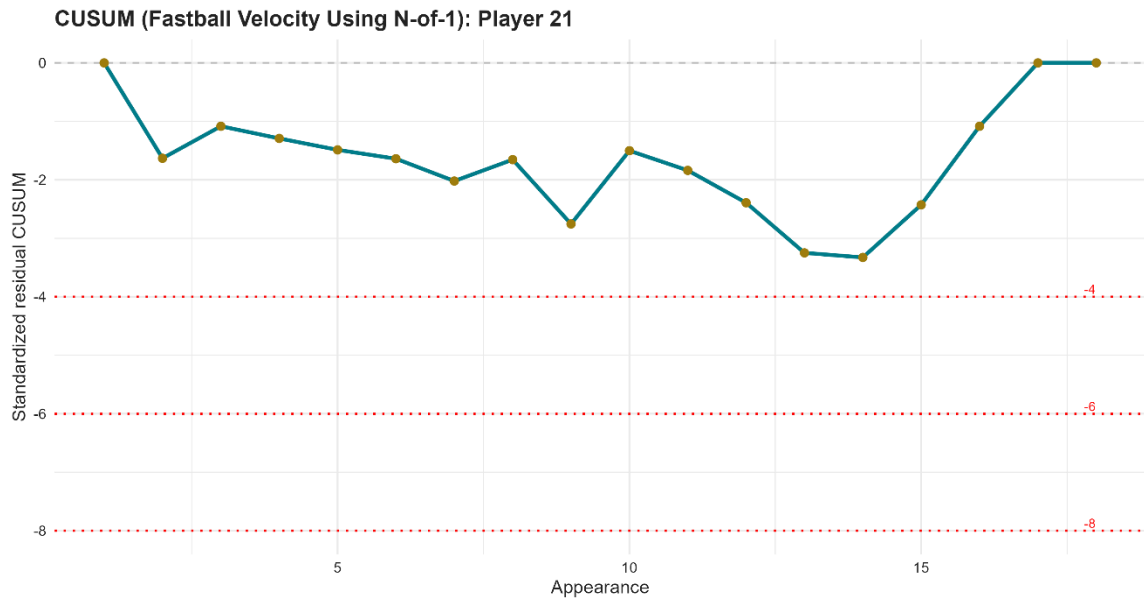


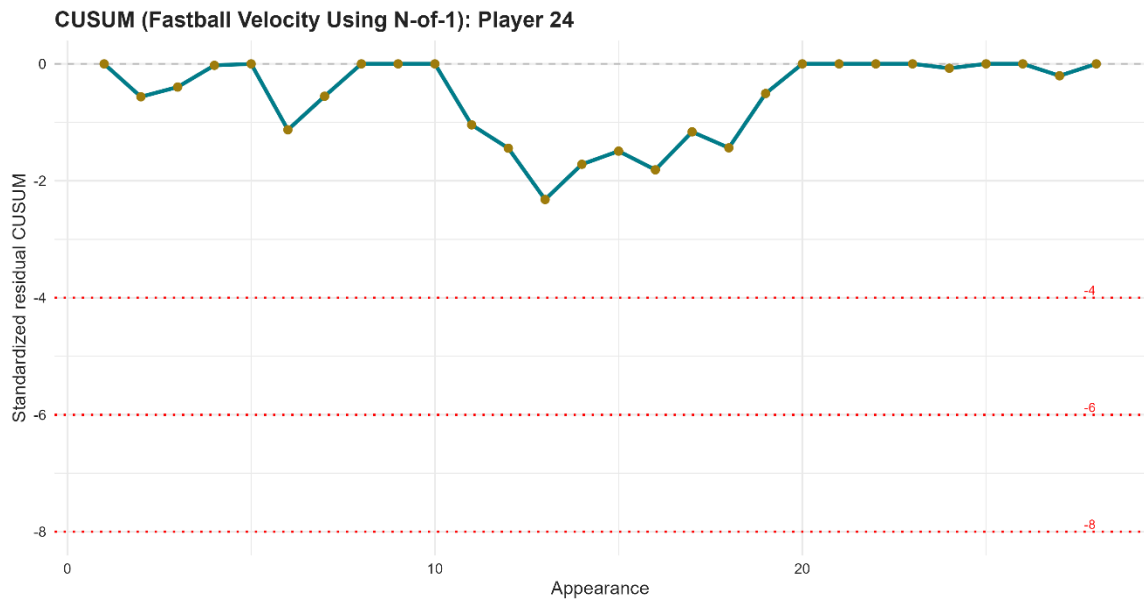
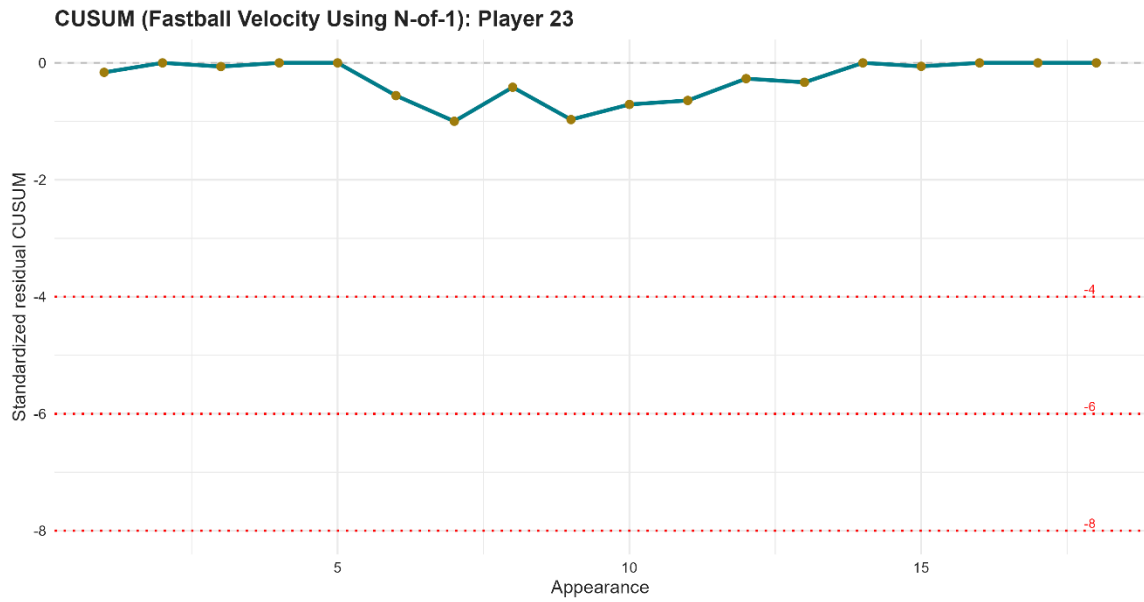


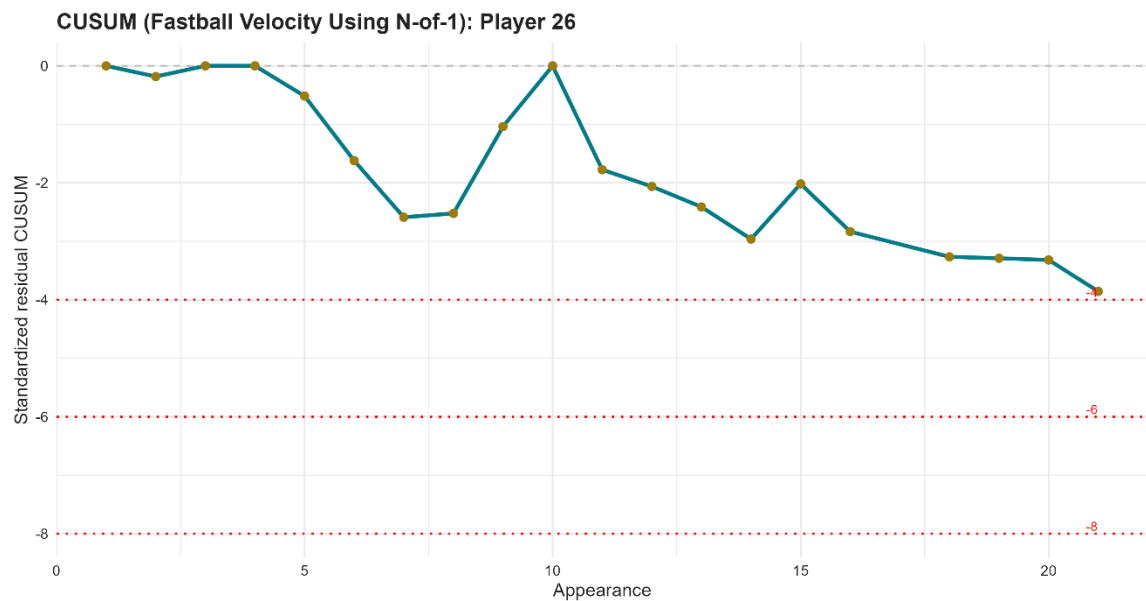
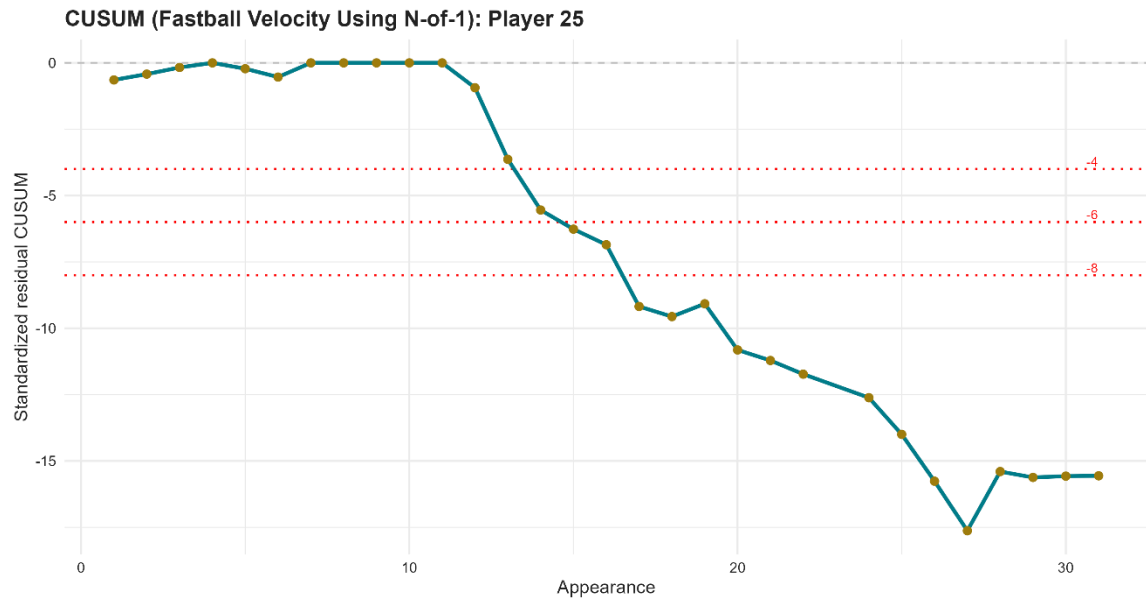












Appendix D – N-of-1 Demo Code

N-of-1 Demo Pipeline

Combined, walkthrough of unified pipeline applied to performance outcomes, biomechanics outcomes, and downstream residual-based CUSUM monitoring

What this is: A streamlined version of the master R script. It explains what each stage does, why it exists, and how the performance and biomechanics branches relate to each other.

What this is not: An original production script copied line-for-line into one file. The delivered script is a faithful demo reconstruction of the full workflow, designed for explanation and review.

How the combined pipeline is organized

Data are cleaned and inventoried first, then screened for collinearity and autocorrelation, then passed into pitcher-season specific model development, then compared against population-level models, and finally exported into a downstream residual surveillance workflow using CUSUM. Velocity is the main applied example, but the same residual-monitoring concept can be extended to selected biomechanics outcomes.

- Unit of inference for N-of-1 models: pitcher-season.
- Primary performance outcomes: Velocity, SpinRate, VertBreak.
- Primary biomechanics outcomes: a set of joint loading, segment position, and kinetic-chain metrics.
- Population-level comparison model: the same hierarchical GAMM concept is used for both performance and biomechanics outcomes so the comparison framework is aligned across outcome domains.
- Monitoring stage: downstream residual-based CUSUM shown with velocity as the primary example, but structured so the same conceptual monitoring logic can also be applied to selected biomechanics outcomes.

Scripts 01-03: data audit and modeling prerequisites

01 Descriptives

Creates a full inventory of pitchers, appearances, pitches, pitch-count distributions, missingness, and outcome distributions. This is the first quality-control stage and establishes the analytic sample before any modeling starts.

02 Multicollinearity

Checks whether candidate predictors are too strongly related to each other. Correlation matrices and VIF-style diagnostics are used as screening tools, not as final inferential models.

03 Autocorrelation

Fits simple baseline models within each pitcher-season and estimates lag-1 residual autocorrelation within appearance. This stage is important because serial dependence motivates testing AR(1) structures in the final GAMMs.

Scripts 04-05: predictor screening and individualized model development

04 LASSO screening

Runs pitcher-season specific LASSO screening with PitchCount forced into every model. It also collapses predictor groups so only one transport variable and one lag variable are carried forward when multiple related candidates compete.

05 Final N-of-1 GAMMs

Builds the individualized GAMM for each pitcher-season and outcome. Candidate formulas are created from the LASSO-selected predictors, one interaction at a time is tested with PitchCount, and no-AR versus AR(1) versions are compared primarily by AIC while retaining BIC in the reported metric set. The best overall model is retained.

Scripts 06-09: diagnostics, prediction, and comparison with group models

06 Diagnostics

Creates residual-versus-fitted, Q-Q, histogram, and related diagnostics to evaluate assumption quality and whether AR(1) reduced serial structure.

07 Prediction exports

Exports observed, fitted, and residual values and produces individualized pitch-count curves. These are the main pitcher-specific response profiles used in figures and interpretation.

08 Group models

Fits one unified population-level comparison model across both outcome domains: a hierarchical GAMM with a shared smooth pitch-count term plus pitcher-specific random intercepts and random slopes.

09 Group vs N-of-1 comparison

Compares individualized and group models using the full reported metric set of AIC, BIC, RMSE, MSE, R^2 , and within-pitcher adjusted R^2 . RMSE and within-pitcher adjusted R^2 are treated as the primary model-performance metrics for both display and interpretation, while AIC, BIC, MSE, and R^2 are retained as supporting model-comparison statistics.

Scripts 10-15: robustness, stability, and uncertainty

10 Early-season stabilization

Tests how closely early-season models reproduce the full-season curve, which speaks to when a model becomes useful for applied monitoring.

11 Hybrid parameter sensitivity

Examines how model behavior changes as more data accumulate and supports practical discussion of how many pitches are needed before curves stabilize.

12 Alternate-specification sensitivity

Refits simpler variants of the final model to quantify how much performance drops when key structural pieces are removed.

13 Trial-reduction sensitivity

Removes fixed proportions of data, refits the models, and measures how much fit degrades. This is a direct robustness check for sparse pitcher-season series.

14 Summary outputs

Collapses model results into season-level and overall summary tables for chapter writing, slides, and high-level reporting.

15 Bootstrap / conditional intervals

Uses resampling to create uncertainty bands around individualized prediction curves. These intervals are useful descriptively but should not be oversold as causal evidence.

Scripts 16-20: export and downstream velocity surveillance

16 Final exports

Packages model registries, formulas, metrics, and other objects that would typically be passed into later scripts or reporting layers.

17 CUSUM input export

Leaves the original modeling pipeline unchanged and exports residual series at pitch-level and appearance-level. Velocity is the main applied example, but the export logic can also be used for selected biomechanics outcomes. This is the bridge between model development and surveillance.

18 CUSUM backtest

Runs retrospective appearance-level lower one-sided CUSUM as the primary monitoring analysis, with secondary pitch-level and standardized-residual variants. The same residual-monitoring idea can be extended to biomechanics when a meaningful signal direction is defined.

19 Limited grid-search validation

Tunes practical surveillance settings such as baseline window, residual scale, and decision threshold using a constrained validation approach rather than an unrestricted search.

20 Final CUSUM outputs

Selects the final preferred CUSUM specification and generates final tables and figures suitable for manuscript or presentation use.

Performance branch versus biomechanics branch

Feature	Performance outcomes	Biomechanics outcomes
Seasons	2023, 2024, 2025	2024, 2025
Main signals	Velocity, SpinRate, Joint loading, joint angle, VertBreak and kinetic-chain metrics	
Model target	Pitcher-season specific performance response	Pitcher-season specific mechanical response
Comparison model	Hierarchical group GAMM comparator	Hierarchical group GAMM comparator
CUSUM use	Yes, downstream residual surveillance	Conceptually extendable to selected outcomes

```
# N-OF-1 DEMO PIPELINE

suppressPackageStartupMessages({
  library(dplyr)
  library(tidyr)
  library(readxl)
  library(readr)
  library(stringr)
  library(purrr)
  library(tibble)
  library(ggplot2)
  library(glmnet)
  library(mgcv)
  library(nlme)
```

```

library(patchwork)
})

`%||` <- function(a, b) if (!is.null(a) && length(a) > 0) a else b

# Choose outcome family for the demo.
# "performance"
# "biomechanics"
analysis_family <- "performance"

# Performance seasons
performance_seasons <- c("2023", "2024", "2025")

# Biomechanics seasons
biomech_seasons <- c("2024", "2025")

# Core directories
base_dir <- "PATH/TO/PROJECT"
perf_dir <- file.path(base_dir, "Performance")
bm_dir <- file.path(base_dir, "Biomechanics")
out_dir <- file.path(base_dir, "Master_Demo_Output")
if (!dir.exists(out_dir)) dir.create(out_dir, recursive = TRUE)

# HELPER FUNCTIONS USED ACROSS MULTIPLE STAGES

make_pitcher_map <- function(x) {
  u <- sort(unique(as.character(x)))
  tibble(
    Pitcher = u,
    Pitcher_DeID = sprintf("Player_%02d", seq_along(u))
  )
}

apply_pitcher_deid <- function(df, map_tbl, pitcher_col = "Pitcher")
{
  df %>%
    left_join(map_tbl, by = setNames("Pitcher", pitcher_col)) %>%
    mutate(
      Pitcher_Original = .data[[pitcher_col]],
      Pitcher = Pitcher_DeID
    ) %>%
    select(-Pitcher_DeID)
}

# Data harmonization helpers

```

```

find_first_existing <- function(df, candidates) {
  candidates[candidates %in% names(df)][1] %||% NA_character_
}

coerce_numeric_if_present <- function(df, cols) {
  for (nm in cols) {
    if (nm %in% names(df)) df[[nm]] <-
      suppressWarnings(as.numeric(df[[nm]]))
  }
  df
}

safe_r2 <- function(obs, pred) {
  keep <- is.finite(obs) & is.finite(pred)
  if (sum(keep) < 3) return(NA_real_)
  suppressWarnings(cor(obs[keep], pred[keep])^2)
}

safe_adj_r2 <- function(obs, pred, p_terms) {
  keep <- is.finite(obs) & is.finite(pred)
  n <- sum(keep)
  if (n < 4) return(NA_real_)
  r2 <- safe_r2(obs[keep], pred[keep])
  if (!is.finite(r2)) return(NA_real_)
  p_terms <- suppressWarnings(as.numeric(p_terms))
  if (!is.finite(p_terms) || is.na(p_terms) || p_terms < 1) p_terms
  <- 1
  if ((n - p_terms - 1) <= 0) return(NA_real_)
  1 - (1 - r2) * ((n - 1) / (n - p_terms - 1))
}

count_model_terms <- function(model_obj) {
  out <- NA_real_
  if (!is.null(model_obj$gam)) {
    out <- try(length(stats::coef(model_obj$gam)), silent = TRUE)
    if (!inherits(out, "try-error") && is.finite(out))
      return(as.numeric(out))
  }
  if (!is.null(model_obj$lme)) {
    out <- try(length(nlme::fixef(model_obj$lme)), silent = TRUE)
    if (!inherits(out, "try-error") && is.finite(out))
      return(as.numeric(out))
  }
  out <- try(length(stats::coef(model_obj)), silent = TRUE)
  if (!inherits(out, "try-error") && is.finite(out))
    return(as.numeric(out))
  NA_real_
}

standardize_transportation <- function(x) {
  out <- as.character(x)

```

```

    out[is.na(out) | !nzchar(trimws(out))] <- "None"
    factor(out)
  }

create_appearance_if_missing <- function(df) {
  if (!("Appearance" %in% names(df))) {
    if (!all(c("Pitcher", "Date") %in% names(df))) {
      stop("Need Pitcher and Date to construct Appearance.")
    }
    df$Appearance <- interaction(df$Pitcher, df$Date, drop = TRUE,
lex.order = TRUE)
  }
  df$Appearance <- as.factor(df$Appearance)
  df
}

compute_arl_for_vector <- function(x) {
  x <- x[is.finite(x)]
  if (length(x) < 3) return(NA_real_)
  acf_out <- acf(x, lag.max = 1, plot = FALSE, na.action = na.pass)
  as.numeric(acf_out$acf[2])
}

# Outcome lists

performance_outcomes <- c("Velocity", "SpinRate", "VertBreak")
biomech_outcomes <- c(
  "Elb_Flex_BR",
  "Lead_Knee_Flex_BR",
  "Hip_Sho_Sep_FC",
  "Max_Hip_Sho_Sep",
  "Sho_Rot_MER",
  "Sho_AbAd_BR",
  "Max_Resultant_Sho_Force",
  "Elb_Flex_MER",
  "Normalized_Max_Elb_Var_Torque",
  "Max_Sho_Rot_Torque",
  "Max_Elb_Var_Torque"
)

candidate_predictors <- c(
  "PitchCount",
  "PitchCount_lag1", "PitchCount_lag2", "PitchCount_lag3",
  "Rest",
  "PrevApp_MaxPitchCount",
  "Travel",
  "Distance_KM",
  "Transportation",
  "Game",
  "HRV"
)

```

```

# DATA LOADING
load_performance_data <- function() {
  season_paths <- c(
    "2023" = file.path(perf_dir, "Data_2023.xlsx"),
    "2024" = file.path(perf_dir, "Data_2024.xlsx"),
    "2025" = file.path(perf_dir, "Data_2025.xlsx")
  )

  existing <- season_paths[file.exists(season_paths)]
  if (length(existing) == 0) stop("No performance season files
found.")

  data_list <- lapply(names(existing), function(season_nm) {
    df <- read_excel(existing[[season_nm]])

    season_col <- find_first_existing(df, c("Season"))
    pitcher_col <- find_first_existing(df, c("Pitcher", "Player",
"Name"))
    velocity_col <- find_first_existing(df, c("Velocity",
"RelSpeed"))
    spin_col <- find_first_existing(df, c("SpinRate", "Spin.Rate"))
    vb_col <- find_first_existing(df, c("VertBreak",
"Vertical.Break"))

    if (!is.na(pitcher_col)) names(df)[names(df) == pitcher_col] <-
"Pitcher"
    if (!is.na(velocity_col)) names(df)[names(df) == velocity_col]
<- "Velocity"
    if (!is.na(spin_col)) names(df)[names(df) == spin_col] <-
"SpinRate"
    if (!is.na(vb_col)) names(df)[names(df) == vb_col] <-
"VertBreak"

    if (!is.na(season_col)) {
      names(df)[names(df) == season_col] <- "Season"
      df$Season <- as.character(df$Season)
    } else {
      df$Season <- season_nm
    }

    if ("Date" %in% names(df)) {
      if (inherits(df$Date, "POSIXt")) {
        df$Date <- as.Date(df$Date)
      } else if (is.numeric(df$Date)) {
        df$Date <- as.Date(df$Date, origin = "1899-12-30")
      } else {
        df$Date <- as.Date(as.character(df$Date), tryFormats =
c("%m/%d/%Y", "%Y-%m-%d"))
      }
    }

    df <- coerce_numeric_if_present(

```

```

    df,
    c("Velocity", "SpinRate", "VertBreak", "SpinAxis", "Tilt",
      "PitchCount", "PitchCount_lag1", "PitchCount_lag2",
      "PitchCount_lag3",
      "Rest", "PrevApp_MaxPitchCount", "Travel", "Distance_KM",
      "HRV", "Game")
  )

  if ("Transportation" %in% names(df)) {
    df$Transportation <-
standardize_transportation(df$Transportation)
  }

  df <- create_appearance_if_missing(df)

  df$Pitcher <- as.factor(df$Pitcher)
  df$Season <- as.factor(df$Season)
  df
})

bind_rows(data_list)
}

# Biomechanics loader

load_biomech_data <- function() {
  file_candidates <- c(
    file.path(bm_dir, "BM_Data24.xlsx"),
    file.path(bm_dir, "BM_Data25.xlsx")
  )
  existing <- file_candidates[file.exists(file_candidates)]
  if (length(existing) == 0) stop("No biomechanics files found.")

  data_list <- lapply(existing, function(path_i) {
    df <- read_excel(path_i)

    if ("Date" %in% names(df)) {
      if (inherits(df$Date, "POSIXt")) {
        df$Date <- as.Date(df$Date)
      } else if (is.numeric(df$Date)) {
        df$Date <- as.Date(df$Date, origin = "1899-12-30")
      } else {
        df$Date <- as.Date(as.character(df$Date), tryFormats =
c("%m/%d/%Y", "%Y-%m-%d"))
      }
    }

    if (!("Season" %in% names(df))) {
      df$Season <- if (grepl("BM24", basename(path_i), ignore.case =
TRUE)) 2024L else 2025L
    }
  })
}

```

```

    df <- coerce_numeric_if_present(
      df,
      c(biomech_outcomes, "PitchCount", "HRV", "Travel",
        "Distance_KM",
        "Game", "Rest", "RelSpeed", "SpinRate", "VertBreak",
        "SpinAxis")
    )

    if ("Transportation" %in% names(df)) {
      df$Transportation <-
standardize_transportation(df$Transportation)
    }

    df <- create_appearance_if_missing(df)
    df$Pitcher <- as.factor(df$Pitcher)
    df$Season <- as.factor(as.character(df$Season))
    df
  })

  bind_rows(data_list)
}

# Choose active data branch
raw_data <- if (analysis_family == "performance")
load_performance_data() else load_biomech_data()

pitcher_map <- make_pitcher_map(raw_data$Pitcher)
dat <- apply_pitcher_deid(raw_data, pitcher_map)

# For plotting and grouped operations
dat <- dat %>%
  mutate(
    Pitcher = factor(Pitcher),
    Appearance = factor(Appearance),
    Season = factor(Season)
  )

active_outcomes <- if (analysis_family == "performance") {
  intersect(performance_outcomes, names(dat))
} else {
  intersect(biomech_outcomes, names(dat))
}

# 01. DESCRIPTIVES / MISSINGNESS

dir_01 <- file.path(out_dir, "01_descriptives")
dir.create(dir_01, recursive = TRUE, showWarnings = FALSE)

overall_inventory <- tibble(
  AnalysisFamily = analysis_family,

```



```

N_Pitchers = n_distinct(dat$Pitcher),
N_Appearances = n_distinct(dat$Pitcher, dat$Appearance),
N_Pitches = nrow(dat),
FirstSeason = as.character(min(dat$Season)),
LastSeason = as.character(max(dat$Season))
)

season_inventory <- dat %>%
  group_by(Season) %>%
  summarise(
    N_Pitchers = n_distinct(Pitcher),
    N_Appearances = n_distinct(Pitcher, Appearance),
    N_Pitches = n(),
    MeanPitchCount = mean(PitchCount, na.rm = TRUE),
    MaxPitchCount = max(PitchCount, na.rm = TRUE),
    .groups = "drop"
  )

pitcher_inventory <- dat %>%
  group_by(Season, Pitcher) %>%
  summarise(
    N_Appearances = n_distinct(Appearance),
    N_Pitches = n(),
    MeanPitchCount = mean(PitchCount, na.rm = TRUE),
    MaxPitchCount = max(PitchCount, na.rm = TRUE),
    MeanRest = if ("Rest" %in% names(dat)) mean(Rest, na.rm = TRUE)
  ) else NA_real_,
  MeanHRV = if ("HRV" %in% names(dat)) mean(HRV, na.rm = TRUE)
  ) else NA_real_,
  .groups = "drop"
  )

missingness_tbl <- tibble(
  Variable = names(dat),
  N_Missing = sapply(dat, function(x) sum(is.na(x))),
  Pct_Missing = sapply(dat, function(x) mean(is.na(x)) * 100)
) %>% arrange(desc(Pct_Missing))

write_csv(overall_inventory, file.path(dir_01,
"01_overall_inventory.csv"))
write_csv(season_inventory, file.path(dir_01,
"01_season_inventory.csv"))
write_csv(pitcher_inventory, file.path(dir_01,
"01_pitcher_inventory.csv"))
write_csv(missingness_tbl, file.path(dir_01, "01_missingness.csv"))

for (y in active_outcomes) {
  p_hist <- ggplot(dat, aes(x = .data[[y]])) +
    geom_histogram(bins = 30) +
    facet_wrap(~Season, scales = "free_y") +
    theme_bw(base_size = 11) +
    labs(
      title = paste("Distribution of", y, "by season"),

```

```

      x = y, y = "Count"
    )
    ggsave(file.path(dir_01, paste0("01_hist_", y, ".png")), p_hist,
width = 8, height = 5, dpi = 300)
  }

# 02. MULTICOLLINEARITY CHECKS

dir_02 <- file.path(out_dir, "02_multicollinearity")
dir.create(dir_02, recursive = TRUE, showWarnings = FALSE)

predictors_present <- intersect(candidate_predictors, names(dat))
numeric_predictors <- predictors_present[predictors_present !=
"Transportation"]

if (length(numeric_predictors) >= 2) {
  corr_all <- cor(dat[, numeric_predictors], use =
"pairwise.complete.obs")
  write_csv(as.data.frame(as.table(corr_all)), file.path(dir_02,
"02_corr_all.csv"))

  seasons_present <- sort(unique(dat$Season))
  vif_log <- list()

  for (s in seasons_present) {
    df_s <- dat %>% filter(Season == s)
    corr_s <- cor(df_s[, numeric_predictors], use =
"pairwise.complete.obs")
    write_csv(as.data.frame(as.table(corr_s)),
              file.path(dir_02, paste0("02_corr_season_", s,
".csv"))))

    y_vif <- active_outcomes[1]
    rhs <- predictors_present[predictors_present %in% names(df_s)]
    if (length(rhs) < 2 || !(y_vif %in% names(df_s))) {
      vif_log[[length(vif_log) + 1]] <- tibble(Season =
as.character(s), Note = "Skipped VIF.")
      next
    }

    df_vif <- df_s %>% select(all_of(c(y_vif, rhs))) %>% drop_na()

    zero_var <- sapply(df_vif[, rhs, drop = FALSE], function(z) {
      if (is.numeric(z)) {
        v <- var(z, na.rm = TRUE)
        is.na(v) || v == 0
      } else {
        length(unique(z[!is.na(z)])) <= 1
      }
    })

    rhs <- rhs[!zero_var]
  }
}

```

```

    if (length(rhs) < 2) {
      vif_log[[length(vif_log) + 1]] <- tibble(Season =
as.character(s), Note = "Too few predictors after zero-variance
removal.")
      next
    }

    f_vif <- as.formula(paste(y_vif, "~", paste(rhs, collapse = " +
"))))
    fit_vif <- try(lm(f_vif, data = df_vif), silent = TRUE)

    if (inherits(fit_vif, "try-error")) {
      vif_log[[length(vif_log) + 1]] <- tibble(Season =
as.character(s), Note = "LM for VIF failed.")
      next
    }

    if ("package:car" %in% search()) {
      v <- try(car::vif(fit_vif), silent = TRUE)
      if (!inherits(v, "try-error")) {
        v_df <- if (is.matrix(v)) {
          as.data.frame(v) %>%
tibble::rownames_to_column("Predictor")
        } else {
          tibble(Predictor = names(v), VIF = as.numeric(v))
        }
        write_csv(v_df, file.path(dir_02, paste0("02_vif_season_",
s, ".csv")))
      }
    }

    if (length(vif_log) > 0) {
      write_csv(bind_rows(vif_log), file.path(dir_02,
"02_vif_log.csv"))
    }
  }
}

```

03. AUTOCORRELATION SCREENING

```

dir_03 <- file.path(out_dir, "03_autocorrelation")
dir.create(dir_03, recursive = TRUE, showWarnings = FALSE)

ar_results <- list()

for (y in active_outcomes) {
  split_keys <- dat %>%
    distinct(Season, Pitcher) %>%
    arrange(Season, Pitcher)

  for (i in seq_len(nrow(split_keys))) {
    s <- split_keys$Season[i]

```

```

p <- split_keys$Pitcher[i]

df_sp <- dat %>%
  filter(Season == s, Pitcher == p) %>%
  filter(!is.na(.data[[y]]), !is.na(PitchCount)) %>%
  arrange(Appearance, PitchCount)

if (nrow(df_sp) < 10) next

lm_fit <- try(lm(as.formula(paste(y, "~ PitchCount +
Appearance")), data = df_sp), silent = TRUE)
if (inherits(lm_fit, "try-error")) next

df_sp$resid_lm <- resid(lm_fit)

ar_tbl <- df_sp %>%
  group_by(Appearance) %>%
  summarise(
    Outcome = y,
    Season = as.character(first(Season)),
    Pitcher = as.character(first(Pitcher)),
    N_Pitches = n(),
    AR1 = compute_ar1_for_vector(resid_lm),
    .groups = "drop"
  )

ar_results[[length(ar_results) + 1]] <- ar_tbl
}
}

if (length(ar_results) > 0) {
  ar_by_app <- bind_rows(ar_results)
  ar_by_pitcher <- ar_by_app %>%
    filter(!is.na(AR1)) %>%
    group_by(Outcome, Season, Pitcher) %>%
    summarise(
      N_Appearances = n(),
      Mean_AR1 = mean(AR1, na.rm = TRUE),
      Median_AR1 = median(AR1, na.rm = TRUE),
      Prop_AR1_ge_0_3 = mean(AR1 >= 0.3, na.rm = TRUE),
      Prop_AR1_ge_0_5 = mean(AR1 >= 0.5, na.rm = TRUE),
      .groups = "drop"
    )

  write_csv(ar_by_app, file.path(dir_03, "03_ar_by_appearance.csv"))
  write_csv(ar_by_pitcher, file.path(dir_03,
"03_ar_by_pitcher.csv"))
}

# 04. LASSO PREDICTOR SCREENING

# Key dissertation logic:

```

```

# - PitchCount always retained
# - top one transport variable retained from:
#   Transportation / Travel / Distance_KM
# - top one lag retained from:
#   PitchCount_lag1 / lag2 / lag3
# - then keep top remaining predictors up to a small cap

dir_04 <- file.path(out_dir, "04_lasso")
dir.create(dir_04, recursive = TRUE, showWarnings = FALSE)

expand_design_matrix <- function(df, predictors) {
  rhs <- predictors[predictors %in% names(df)]
  if (length(rhs) == 0) return(NULL)
  mm <- model.matrix(
    as.formula(paste("~", paste(rhs, collapse = " + "))),
    data = df
  )
  mm[, colnames(mm) != "(Intercept)", drop = FALSE]
}

base_var_from_coef_name <- function(x) {
  x <- gsub("^", "", x)
  x <- gsub("^Transportation.*", "Transportation", x)
  x <- gsub("^s\\(.*$", "PitchCount", x)
  x <- gsub(":.*$", "", x)
  x
}

lasso_results <- list()

for (y in active_outcomes) {
  split_keys <- dat %>%
    distinct(Season, Pitcher) %>%
    arrange(Season, Pitcher)

  for (i in seq_len(nrow(split_keys))) {
    s <- split_keys$Season[i]
    p <- split_keys$Pitcher[i]

    df_sp <- dat %>%
      filter(Season == s, Pitcher == p) %>%
      filter(!is.na(.data[[y]]), !is.na(PitchCount))

    if (nrow(df_sp) < 20) next

    rhs <- intersect(candidate_predictors, names(df_sp))
    x <- expand_design_matrix(df_sp, rhs)
    if (is.null(x) || ncol(x) == 0) next

    y_vec <- df_sp[[y]]

    penalty_factor <- rep(1, ncol(x))
    penalty_factor[grepl("^PitchCount$", colnames(x))] <- 0
  }
}

```

```

cv_fit <- try(
  cv.glmnet(
    x = x,
    y = y_vec,
    alpha = 1,
    family = "gaussian",
    penalty.factor = penalty_factor
  ),
  silent = TRUE
)
if (inherits(cv_fit, "try-error")) next

coef_min <- as.matrix(coef(cv_fit, s = "lambda.min"))
coef_tbl <- tibble(
  Coefficient = rownames(coef_min),
  Value = as.numeric(coef_min[, 1])
) %>%
  filter(Coefficient != "(Intercept)", Value != 0)

if (nrow(coef_tbl) == 0) {
  selected <- "PitchCount"
  selected_travel <- NA_character_
  selected_lag <- NA_character_
} else {
  coef_tbl <- coef_tbl %>%
    mutate(BaseVar = sapply(Coefficient,
      base_var_from_coef_name,
      AbsValue = abs(Value))

    agg <- coef_tbl %>%
      group_by(BaseVar) %>%
      summarise(Score = max(AbsValue), .groups = "drop")

    if (!("PitchCount" %in% agg$BaseVar)) {
      agg <- bind_rows(agg, tibble(BaseVar = "PitchCount", Score =
0))
    }

    travel_group <- intersect(c("Transportation", "Travel",
"Distance_KM"), agg$BaseVar)
    selected_travel <- NA_character_
    if (length(travel_group) > 0) {
      best_travel <- agg %>% filter(BaseVar %in% travel_group) %>%
slice_max(Score, n = 1)
      selected_travel <- best_travel$BaseVar[1]
      agg <- agg %>% filter(!(BaseVar %in% setdiff(travel_group,
selected_travel)))
    }

    lag_group <- intersect(c("PitchCount_lag1", "PitchCount_lag2",
"PitchCount_lag3"), agg$BaseVar)
    selected_lag <- NA_character_

```

```

    if (length(lag_group) > 0) {
      best_lag <- agg %>% filter(BaseVar %in% lag_group) %>%
slice_max(Score, n = 1)
      selected_lag <- best_lag$BaseVar[1]
      agg <- agg %>% filter(!(BaseVar %in% setdiff(lag_group,
selected_lag)))
    }

    agg_sorted <- agg %>% arrange(desc(Score))
    extras <- agg_sorted %>%
      filter(BaseVar != "PitchCount") %>%
      slice_head(n = 4) %>%
      pull(BaseVar)

    selected <- c("PitchCount", extras) %>% unique()
  }

  lasso_results[[length(lasso_results) + 1]] <- tibble(
    Outcome = y,
    Season = as.character(s),
    Pitcher = as.character(p),
    N_Obs = nrow(df_sp),
    Lambda_Min = cv_fit$lambda.min,
    Lambda_1SE = cv_fit$lambda.1se,
    SelectedPredictors = paste(selected, collapse = ";"),
    TravelGroupChoice = selected_travel,
    LagGroupChoice = selected_lag
  )
}
}

lasso_tbl <- bind_rows(lasso_results)
write_csv(lasso_tbl, file.path(dir_04, "04_lasso_results.csv"))

# 05. N-OF-1 GAMM MODEL FITTING

# Model development logic:
# - exposure: rolling PitchCount
# - smooth: s(PitchCount, k = 4)
# - random intercept: Appearance
# - AR(1): optional, nested within Appearance
# - max one interaction term retained
# - interactions:
#   numeric X -> te(PitchCount, X)
#   Transportation -> s(PitchCount, by = Transportation) +
Transportation

dir_05 <- file.path(out_dir, "05_n1_models")
dir.create(dir_05, recursive = TRUE, showWarnings = FALSE)

parse_selected_predictors <- function(df_lasso, outcome, season,
pitcher) {

```

```

hit <- df_lasso %>%
  filter(Outcome == outcome, Season == season, Pitcher == pitcher)
if (nrow(hit) == 0) return("PitchCount")
str_split(hit$SelectedPredictors[1], ";")[[1]]
}

safe_aic <- function(model_obj) {
  out <- try(AIC(model_obj$lme %||% model_obj), silent = TRUE)
  if (inherits(out, "try-error")) return(Inf)
  as.numeric(out)
}

build_candidate_formulas <- function(response, selected_predictors,
df) {
  selected_predictors <- selected_predictors[selected_predictors
%in% names(df)]
  selected_predictors <- unique(c("PitchCount",
selected_predictors))

  linear_terms <- setdiff(selected_predictors, c("PitchCount",
"Transportation"))
  base_terms <- c("s(PitchCount, k = 4)", linear_terms)

  if ("Transportation" %in% selected_predictors && "Transportation"
%in% names(df)) {
    base_terms <- c(base_terms, "Transportation")
  }

  forms <- list()
  base_formula <- as.formula(paste(response, "~", paste(base_terms,
collapse = " + ")))
  forms[["base"]] <- base_formula

  interaction_candidates <- setdiff(selected_predictors,
"PitchCount")

  for (x in interaction_candidates) {
    if (!(x %in% names(df))) next

    if (x == "Transportation" && is.factor(df$Transportation)) {
      form_i <- as.formula(
        paste(response, "~",
          paste(c(setdiff(base_terms, "Transportation"),
"Transportation",
"s(PitchCount, by = Transportation, k = 4)"),
collapse = " + ")
        )
      forms[[paste0("interaction_", x)]] <- form_i
    } else if (is.numeric(df[[x]])) {
      non_x_terms <- setdiff(base_terms, x)
      form_i <- as.formula(
        paste(response, "~",

```



```

      paste(c(non_x_terms, paste0("te(PitchCount, ", x, ", ", k
= c(4, 4))")),
      collapse = " + "))
    )
    forms[[paste0("interaction_", x)]] <- form_i
  }
}

forms
}

fit_gamm_pair <- function(formula_obj, df) {
  fit_no_ar <- try(
    gamm(
      formula = formula_obj,
      random = list(Appearance = ~1),
      data = df,
      method = "REML"
    ),
    silent = TRUE
  )

  fit_ar1 <- try(
    gamm(
      formula = formula_obj,
      random = list(Appearance = ~1),
      correlation = corAR1(form = ~ PitchCount | Appearance),
      data = df,
      method = "REML"
    ),
    silent = TRUE
  )

  list(no_ar = fit_no_ar, ar1 = fit_ar1)
}

n1_model_index <- list()
n1_summary <- list()

for (y in active_outcomes) {
  split_keys <- dat %>% distinct(Season, Pitcher) %>%
  arrange(Season, Pitcher)

  for (i in seq_len(nrow(split_keys))) {
    s <- as.character(split_keys$Season[i])
    p <- as.character(split_keys$Pitcher[i])

    df_sp <- dat %>%
      filter(Season == s, Pitcher == p) %>%
      filter(!is.na(.data[[y]]), !is.na(PitchCount)) %>%
      arrange(Appearance, PitchCount)

    if (nrow(df_sp) < 20) next
  }
}

```

```

    selected_predictors <- parse_selected_predictors(lasso_tbl, y,
s, p)
    candidate_forms <- build_candidate_formulas(y,
selected_predictors, df_sp)

    best_obj <- NULL
    best_label <- NA_character_
    best_model_type <- NA_character_
    best_aic <- Inf

    for (nm in names(candidate_forms)) {
      fits <- fit_gamm_pair(candidate_forms[[nm]], df_sp)

      if (!inherits(fits$no_ar, "try-error")) {
        aic_no_ar <- safe_aic(fits$no_ar)
        if (is.finite(aic_no_ar) && aic_no_ar < best_aic) {
          best_obj <- fits$no_ar
          best_label <- nm
          best_model_type <- "no_ar"
          best_aic <- aic_no_ar
        }
      }

      if (!inherits(fits$ar1, "try-error")) {
        aic_ar1 <- safe_aic(fits$ar1)
        if (is.finite(aic_ar1) && aic_ar1 < best_aic) {
          best_obj <- fits$ar1
          best_label <- nm
          best_model_type <- "ar1"
          best_aic <- aic_ar1
        }
      }
    }

    if (is.null(best_obj)) next

    key <- paste(y, s, p, sep = "__")
    n1_model_index[[key]] <- best_obj

    pred <- try(predict(best_obj$gam, newdata = df_sp), silent =
TRUE)
    resid <- if (!inherits(pred, "try-error")) df_sp[[y]] -
as.numeric(pred) else rep(NA_real_, nrow(df_sp))

    n1_summary[[length(n1_summary) + 1]] <- tibble(
      Outcome = y,
      Season = s,
      Pitcher = p,
      N_Obs = nrow(df_sp),
      SelectedBasePredictors = paste(selected_predictors, collapse =
";"),
      BestFormulaLabel = best_label,

```

```

      BestModelType = best_model_type,
      AIC = best_aic,
      BIC = try(BIC(best_model$lme %||% best_model), silent = TRUE)
    }>% as.numeric(),
      RMSE = sqrt(mean(resid^2, na.rm = TRUE)),
      MSE = mean(resid^2, na.rm = TRUE),
      R2 = safe_r2(df_sp[[y]], pred),
      WithinPitcher_Adj_R2 = safe_adj_r2(df_sp[[y]], pred,
count_model_terms(best_model))
    )
  }
}

nl_summary_tbl <- bind_rows(nl_summary)
write_csv(nl_summary_tbl, file.path(dir_05,
"05_nl_model_summary.csv"))

# 06. DIAGNOSTICS
# - residual vs fitted
# - Q-Q plot
# - histogram of residuals
# - ACF of residuals
# - used to check distributional assumptions and whether
#   residual autocorrelation has been meaningfully reduced

dir_06 <- file.path(out_dir, "06_diagnostics")
dir.create(dir_06, recursive = TRUE, showWarnings = FALSE)

make_diagnostic_plots <- function(obs, fit, outcome, season,
pitcher) {
  res <- obs - fit
  dfp <- tibble(Fitted = fit, Residual = res)

  p1 <- ggplot(dfp, aes(Fitted, Residual)) +
    geom_point(alpha = 0.7) +
    geom_hline(yintercept = 0, linetype = 2) +
    theme_bw(base_size = 11) +
    labs(title = paste(outcome, "-", pitcher, "-", season, "-
Residual vs fitted"))

  p2 <- ggplot(dfp, aes(sample = Residual)) +
    stat_qq() + stat_qq_line() +
    theme_bw(base_size = 11) +
    labs(title = paste(outcome, "-", pitcher, "-", season, "- Q-Q
plot"))

  p3 <- ggplot(dfp, aes(Residual)) +
    geom_histogram(bins = 30) +
    theme_bw(base_size = 11) +
    labs(title = paste(outcome, "-", pitcher, "-", season, "-
Residual histogram"))

```

```

    list(p1 = p1, p2 = p2, p3 = p3)
  }

  for (key in names(n1_model_index)) {
    parts <- str_split(key, "_")[[1]]
    y <- parts[1]; s <- parts[2]; p <- parts[3]

    df_sp <- dat %>%
      filter(Season == s, Pitcher == p) %>%
      filter(!is.na(.data[[y]]), !is.na(PitchCount)) %>%
      arrange(Appearance, PitchCount)

    fit_obj <- n1_model_index[[key]]
    pred <- try(predict(fit_obj$gam, newdata = df_sp), silent = TRUE)
    if (inherits(pred, "try-error")) next

    plots <- make_diagnostic_plots(df_sp[[y]], pred, y, s, p)
    ggsave(file.path(dir_06, paste0("06_diag_", y, "_", s, "_", p,
"_residfit.png")), plots$p1, width = 7, height = 5, dpi = 300)
    ggsave(file.path(dir_06, paste0("06_diag_", y, "_", s, "_", p,
"_qq.png")), plots$p2, width = 7, height = 5, dpi = 300)
    ggsave(file.path(dir_06, paste0("06_diag_", y, "_", s, "_", p,
"_hist.png")), plots$p3, width = 7, height = 5, dpi = 300)
  }

```

07. PREDICTION EXPORTS / CURVE VISUALIZATION

```

dir_07 <- file.path(out_dir, "07_predictions")
dir.create(dir_07, recursive = TRUE, showWarnings = FALSE)

prediction_exports <- list()

for (key in names(n1_model_index)) {
  parts <- str_split(key, "_")[[1]]
  y <- parts[1]; s <- parts[2]; p <- parts[3]

  df_sp <- dat %>%
    filter(Season == s, Pitcher == p) %>%
    filter(!is.na(.data[[y]]), !is.na(PitchCount))

  fit_obj <- n1_model_index[[key]]
  pred <- try(predict(fit_obj$gam, newdata = df_sp), silent = TRUE)
  if (inherits(pred, "try-error")) next

  out_df <- df_sp %>%
    mutate(
      Outcome = y,
      Observed = .data[[y]],
      Predicted = as.numeric(pred),
      Residual = Observed - Predicted
    ) %>%

```

```

    select(Outcome, Season, Pitcher, Appearance, PitchCount,
Observed, Predicted, Residual)

prediction_exports[[length(prediction_exports) + 1]] <- out_df

p_curve <- ggplot(out_df, aes(PitchCount, Observed)) +
  geom_point(alpha = 0.3) +
  geom_line(aes(y = Predicted), linewidth = 1) +
  theme_bw(base_size = 11) +
  labs(title = paste("Observed vs N-of-1 fitted:", y, "-", p, "-",
s))
  ggsave(file.path(dir_07, paste0("07_curve_", y, "_", s, "_", p,
".png")), p_curve, width = 7, height = 5, dpi = 300)
}

pred_tbl <- bind_rows(prediction_exports)
write_csv(pred_tbl, file.path(dir_07,
"07_pitch_level_predictions.csv"))

# 08. UNIFIED GROUP MODEL FITTING

# - this comparator is a hierarchical GAMM with:
#   * one shared smooth term for PitchCount
#   * pitcher-specific random intercepts
#   * pitcher-specific random slopes for PitchCount
# - it is fit separately by season and, when useful, on pooled
data

dir_08 <- file.path(out_dir, "08_group_models")
dir.create(dir_08, recursive = TRUE, showWarnings = FALSE)

group_model_index <- list()
group_summary <- list()

for (y in active_outcomes) {
  seasons_to_fit <- c(levels(dat$Season), "ALL")

  for (s in seasons_to_fit) {
    df_s <- if (s == "ALL") dat else dat %>% filter(Season == s)
    df_s <- df_s %>% filter(!is.na(.data[[y]]), !is.na(PitchCount))
    if (nrow(df_s) < 30) next

    fit_group <- try(
      gamm(
        formula = as.formula(paste(y, "~ s(PitchCount, k = 4)")),
        random = list(Pitcher = ~1 + PitchCount),
        data = df_s,
        method = "REML"
      ),
      silent = TRUE
    )
  }
}

```

```

    if (!inherits(fit_group, "try-error")) {
      key <- paste("HierarchicalGAMM", y, s, sep = "__")
      group_model_index[[key]] <- fit_group
      pred <- try(predict(fit_group$gam, newdata = df_s), silent =
TRUE)
      rmse <- if (!inherits(pred, "try-error")) sqrt(mean((df_s[[y]]
- pred)^2, na.rm = TRUE)) else NA_real_
      group_summary[[length(group_summary) + 1]] <- tibble(
        GroupStructure = "HierarchicalGAMM",
        Outcome = y,
        Season = as.character(s),
        N_Obs = nrow(df_s),
        RMSE = rmse,
        MSE = mean((df_s[[y]] - pred_group)^2, na.rm = TRUE),
        R2 = safe_r2(df_s[[y]], pred_group),
        WithinPitcher_Adj_R2 = safe_adj_r2(df_s[[y]], pred_group,
count_model_terms(fit_group)),
        AIC = try(AIC(fit_group$lme), silent = TRUE) %>%
as.numeric(),
        BIC = try(BIC(fit_group$lme), silent = TRUE) %>%
as.numeric()
      )
    }
  }
}

```

```

group_summary_tbl <- bind_rows(group_summary)
write_csv(group_summary_tbl, file.path(dir_08,
"08_group_model_summary.csv"))

```

09. GROUP VS N-OF-1 COMPARISON

```

# What is being done:
#   - compares explanatory performance at the pitcher-season level
#   - main statistics:
#       AIC, BIC, RMSE, MSE, R2, and within-pitcher adjusted R2
#   - interpretation emphasis:
#       RMSE and within-pitcher adjusted R2

```

```

dir_09 <- file.path(out_dir, "09_group_vs_n1")
dir.create(dir_09, recursive = TRUE, showWarnings = FALSE)

```

```

comparison_rows <- list()

```

```

for (y in active_outcomes) {
  keys_sp <- dat %>% distinct(Season, Pitcher) %>% arrange(Season,
Pitcher)

```

```

  for (i in seq_len(nrow(keys_sp))) {
    s <- as.character(keys_sp$Season[i])

```

```

p <- as.character(keys_sp$Pitcher[i])

df_sp <- dat %>% filter(Season == s, Pitcher == p,
!is.na(.data[[y]]), !is.na(PitchCount))
if (nrow(df_sp) < 20) next

n1_key <- paste(y, s, p, sep = "__")
if (!n1_key %in% names(n1_model_index)) next
pred_n1 <- try(predict(n1_model_index[[n1_key]]$gam, newdata =
df_sp), silent = TRUE)
if (inherits(pred_n1, "try-error")) next

g_key <- paste("HierarchicalGAMM", y, s, sep = "__")
if (!(g_key %in% names(group_model_index))) next

g_obj <- group_model_index[[g_key]]
pred_group <- try({
  if (!is.null(g_obj$gam)) predict(g_obj$gam, newdata = df_sp)
else predict(g_obj, newdata = df_sp)
}, silent = TRUE)
if (inherits(pred_group, "try-error")) next

obs <- df_sp[[y]]
model_n1 <- n1_model_index[[n1_key]]
model_group <- g_obj
aic_n1 <- try(AIC(model_n1$lme %||% model_n1), silent = TRUE)
%>% as.numeric()
aic_group <- try(AIC(model_group$lme %||% model_group), silent =
TRUE) %>% as.numeric()
bic_n1 <- try(BIC(model_n1$lme %||% model_n1), silent = TRUE)
%>% as.numeric()
bic_group <- try(BIC(model_group$lme %||% model_group), silent =
TRUE) %>% as.numeric()
mse_n1 <- mean((obs - pred_n1)^2, na.rm = TRUE)
mse_group <- mean((obs - pred_group)^2, na.rm = TRUE)
rmse_n1 <- sqrt(mse_n1)
rmse_group <- sqrt(mse_group)
r2_n1 <- safe_r2(obs, pred_n1)
r2_group <- safe_r2(obs, pred_group)
adj_r2_n1 <- safe_adj_r2(obs, pred_n1,
count_model_terms(model_n1))
adj_r2_group <- safe_adj_r2(obs, pred_group,
count_model_terms(model_group))
comparison_rows[[length(comparison_rows) + 1]] <- tibble(
  Outcome = y,
  Season = s,
  Pitcher = p,
  GroupStructure = "HierarchicalGAMM",
  N_Obs = nrow(df_sp),
  AIC_N1 = aic_n1,
  AIC_Group = aic_group,
  BIC_N1 = bic_n1,
  BIC_Group = bic_group,

```

```

      MSE_N1 = mse_n1,
      MSE_Group = mse_group,
      RMSE_N1 = rmse_n1,
      RMSE_Group = rmse_group,
      R2_N1 = r2_n1,
      R2_Group = r2_group,
      WithinPitcher_Adj_R2_N1 = adj_r2_n1,
      WithinPitcher_Adj_R2_Group = adj_r2_group,
      Delta_AIC = aic_group - aic_n1,
      Delta_BIC = bic_group - bic_n1,
      Delta_MSE = mse_group - mse_n1,
      Delta_RMSE = rmse_group - rmse_n1,
      Delta_R2 = r2_n1 - r2_group,
      Delta_WithinPitcher_Adj_R2 = adj_r2_n1 - adj_r2_group
    )
  }
}

comparison_tbl <- bind_rows(comparison_rows)
write_csv(comparison_tbl, file.path(dir_09,
"09_group_vs_n1_metrics.csv"))

if (nrow(comparison_tbl) > 0) {
  p_box <- ggplot(comparison_tbl, aes(x = Season, y =
Delta_WithinPitcher_Adj_R2)) +
    geom_boxplot() +
    theme_bw(base_size = 11) +
    labs(
      title = paste("Within-pitcher Delta Adjusted R2:",
analysis_family),
      y = "N-of-1 adjusted R2 minus Group adjusted R2"
    )
  ggsave(file.path(dir_09,
"09_delta_within_pitcher_adj_r2_boxplot.png"), p_box, width = 7,
height = 5, dpi = 300)
}

```

10. EARLY-SEASON STABILIZATION ANALYSIS

```

# - asks how much the fitted model changes when only early-season
#   data are used
# - evaluates how close early curves are to the full-season curve
# - used as a practical check on when models become stable enough
#   to support monitoring

```

```

dir_10 <- file.path(out_dir, "10_stabilization")
dir.create(dir_10, recursive = TRUE, showWarnings = FALSE)

```

```

stabilization_rows <- list()

```

```

for (key in names(n1_model_index)) {
  parts <- str_split(key, "__")[[1]]

```



```

y <- parts[1]; s <- parts[2]; p <- parts[3]

df_sp <- dat %>%
  filter(Season == s, Pitcher == p, !is.na(.data[[y]]),
!is.na(PitchCount)) %>%
  arrange(Appearance, PitchCount)

if (nrow(df_sp) < 40) next

appearance_order <- df_sp %>%
  distinct(Appearance) %>%
  mutate(AppOrder = row_number())

df_sp <- df_sp %>% left_join(appearance_order, by = "Appearance")

cut_grid <- c(10, 15, 20, 25, 30)

full_fit <- nl_model_index[[key]]
pred_full <- try(predict(full_fit$gam, newdata = df_sp), silent =
TRUE)
if (inherits(pred_full, "try-error")) next

for (k_cut in cut_grid) {
  df_early <- df_sp %>% filter(AppOrder <= k_cut)
  if (nrow(df_early) < 20) next

  selected_predictors <- parse_selected_predictors(lasso_tbl, y,
s, p)
  candidate_forms <- build_candidate_formulas(y,
selected_predictors, df_early)
  if (length(candidate_forms) == 0) next

  early_fit <- try(
    gamm(candidate_forms[[1]], random = list(Appearance = ~1),
data = df_early, method = "REML"),
    silent = TRUE
  )
  if (inherits(early_fit, "try-error")) next

  pred_early_on_full <- try(predict(early_fit$gam, newdata =
df_sp), silent = TRUE)
  if (inherits(pred_early_on_full, "try-error")) next

  stabilization_rows[[length(stabilization_rows) + 1]] <- tibble(
    Outcome = y,
    Season = s,
    Pitcher = p,
    EarlyAppearanceCut = k_cut,
    RMSE_EarlyVsObserved = sqrt(mean((df_sp[[y]] -
pred_early_on_full)^2, na.rm = TRUE)),
    RMSE_FullVsObserved = sqrt(mean((df_sp[[y]] - pred_full)^2,
na.rm = TRUE)),

```

```

      Delta_RMSE = sqrt(mean((df_sp[[y]] - pred_early_on_full)^2,
na.rm = TRUE)) -
      sqrt(mean((df_sp[[y]] - pred_full)^2, na.rm = TRUE))
    )
  }
}

stabilization_tbl <- bind_rows(stabilization_rows)
write_csv(stabilization_tbl, file.path(dir_10,
"10_stabilization_summary.csv"))

# 11. HYBRID PARAMETER SENSITIVITY / K-THRESHOLD

# What is being done:
# - checks how smoothness / effective curve stability behaves as
#   information accumulates
# - intended to identify practical thresholds for reliable curve
#   fitting rather than claiming one universal required sample
size
#
# In the original dissertation workflow, this stage explored
# into "how many pitches are enough?"

dir_11 <- file.path(out_dir, "11_hybrid_sensitivity")
dir.create(dir_11, recursive = TRUE, showWarnings = FALSE)

hybrid_rows <- list()

for (key in names(n1_model_index)) {
  parts <- str_split(key, "_")[[1]]
  y <- parts[1]; s <- parts[2]; p <- parts[3]

  df_sp <- dat %>%
    filter(Season == s, Pitcher == p, !is.na(.data[[y]]),
!is.na(PitchCount)) %>%
    arrange(Appearance, PitchCount)

  if (nrow(df_sp) < 40) next

  k_grid <- seq(15, min(120, nrow(df_sp)), by = 5)

  selected_predictors <- parse_selected_predictors(lasso_tbl, y, s,
p)

  for (k_n in k_grid) {
    df_sub <- df_sp %>% slice_head(n = k_n)
    if (nrow(df_sub) < 20) next

    forms <- build_candidate_formulas(y, selected_predictors,
df_sub)
    if (length(forms) == 0) next
  }
}

```

```

    fit_sub <- try(
      gamm(forms[[1]], random = list(Appearance = ~1), data =
df_sub, method = "REML"),
      silent = TRUE
    )
    if (inherits(fit_sub, "try-error")) next

    pred_sub <- try(predict(fit_sub$gam, newdata = df_sp), silent =
TRUE)
    pred_full <- try(predict(nl_model_index[[key]]$gam, newdata =
df_sp), silent = TRUE)
    if (inherits(pred_sub, "try-error") || inherits(pred_full, "try-
error")) next

    hybrid_rows[[length(hybrid_rows) + 1]] <- tibble(
      Outcome = y,
      Season = s,
      Pitcher = p,
      N_Used = k_n,
      RMSE_ToObserved = sqrt(mean((df_sp[[y]] - pred_sub)^2, na.rm =
TRUE)),
      RMSE_ToFullCurve = sqrt(mean((pred_full - pred_sub)^2, na.rm =
TRUE))
    )
  }
}

hybrid_tbl <- bind_rows(hybrid_rows)
write_csv(hybrid_tbl, file.path(dir_11,
"11_hybrid_parameter_sensitivity.csv"))

# 12. ALTERNATE-SPECIFICATION SENSITIVITY

# What is being done:
# - removes selected components from the final specification
# - examples:
#   remove AR(1)
#   remove interaction
#   simplify predictor set
# - quantifies how much fit degrades under simpler models

dir_12 <- file.path(out_dir, "12_spec_sensitivity")
dir.create(dir_12, recursive = TRUE, showWarnings = FALSE)

spec_rows <- list()

for (key in names(nl_model_index)) {
  parts <- str_split(key, "_")[[1]]
  y <- parts[1]; s <- parts[2]; p <- parts[3]

  df_sp <- dat %>%

```

```

    filter(Season == s, Pitcher == p, !is.na(.data[[y]]),
!is.na(PitchCount))

    if (nrow(df_sp) < 20) next

    full_fit <- nl_model_index[[key]]
    pred_full <- try(predict(full_fit$gam, newdata = df_sp), silent =
TRUE)
    if (inherits(pred_full, "try-error")) next

    alt_fit <- try(
      gamm(
        as.formula(paste(y, "~ s(PitchCount, k = 4)")),
        random = list(Appearance = ~1),
        data = df_sp,
        method = "REML"
      ),
      silent = TRUE
    )
    if (inherits(alt_fit, "try-error")) next

    pred_alt <- try(predict(alt_fit$gam, newdata = df_sp), silent =
TRUE)
    if (inherits(pred_alt, "try-error")) next

    spec_rows[[length(spec_rows) + 1]] <- tibble(
      Outcome = y,
      Season = s,
      Pitcher = p,
      RMSE_Full = sqrt(mean((df_sp[[y]] - pred_full)^2, na.rm =
TRUE)),
      RMSE_Alt = sqrt(mean((df_sp[[y]] - pred_alt)^2, na.rm = TRUE)),
      Delta_RMSE = sqrt(mean((df_sp[[y]] - pred_alt)^2, na.rm = TRUE))
    )
  }

spec_tbl <- bind_rows(spec_rows)
write_csv(spec_tbl, file.path(dir_12,
"12_specification_sensitivity.csv"))

```

13. TRIAL-REDUCTION SENSITIVITY

```

# What is being done:
# - removes a fixed percentage of observations
# - refits models on the reduced data
# - compares reduced-data fit to full-data fit

```

```

dir_13 <- file.path(out_dir, "13_trial_reduction")
dir.create(dir_13, recursive = TRUE, showWarnings = FALSE)

```

```

set.seed(123)
trial_rows <- list()

for (key in names(nl_model_index)) {
  parts <- str_split(key, "_")[[1]]
  y <- parts[1]; s <- parts[2]; p <- parts[3]

  df_sp <- dat %>%
    filter(Season == s, Pitcher == p, !is.na(.data[[y]]),
           !is.na(PitchCount))

  if (nrow(df_sp) < 40) next

  pred_full <- try(predict(nl_model_index[[key]]$gam, newdata =
df_sp), silent = TRUE)
  if (inherits(pred_full, "try-error")) next

  for (drop_frac in c(0.10, 0.20, 0.30)) {
    keep_n <- floor(nrow(df_sp) * (1 - drop_frac))
    idx_keep <- sort(sample(seq_len(nrow(df_sp)), size = keep_n,
replace = FALSE))
    df_red <- df_sp[idx_keep, , drop = FALSE]
    if (nrow(df_red) < 20) next

    selected_predictors <- parse_selected_predictors(lasso_tbl, y,
s, p)
    forms <- build_candidate_formulas(y, selected_predictors,
df_red)
    if (length(forms) == 0) next

    fit_red <- try(
      gamm(forms[[1]], random = list(Appearance = ~1), data =
df_red, method = "REML"),
      silent = TRUE
    )
    if (inherits(fit_red, "try-error")) next

    pred_red_on_full <- try(predict(fit_red$gam, newdata = df_sp),
silent = TRUE)
    if (inherits(pred_red_on_full, "try-error")) next

    trial_rows[[length(trial_rows) + 1]] <- tibble(
      Outcome = y,
      Season = s,
      Pitcher = p,
      DropFraction = drop_frac,
      RMSE_Full = sqrt(mean((df_sp[[y]] - pred_full)^2, na.rm =
TRUE)),
      RMSE_Reduced = sqrt(mean((df_sp[[y]] - pred_red_on_full)^2,
na.rm = TRUE)),
      Delta_RMSE = sqrt(mean((df_sp[[y]] - pred_red_on_full)^2,
na.rm = TRUE)) -

```

```

      sqrt(mean((df_sp[[y]] - pred_full)^2, na.rm = TRUE))
    )
  }
}

trial_tbl <- bind_rows(trial_rows)
write_csv(trial_tbl, file.path(dir_13,
"13_trial_reduction_sensitivity.csv"))

# 14. SUMMARY TABLES

dir_14 <- file.path(out_dir, "14_summaries")
dir.create(dir_14, recursive = TRUE, showWarnings = FALSE)

summary_season <- n1_summary_tbl %>%
  group_by(Outcome, Season) %>%
  summarise(
    N_Models = n(),
    Mean_RMSE = mean(RMSE, na.rm = TRUE),
    Median_RMSE = median(RMSE, na.rm = TRUE),
    Mean_R2 = mean(R2, na.rm = TRUE),
    Median_R2 = median(R2, na.rm = TRUE),
    Mean_WithinPitcher_Adj_R2 = mean(WithinPitcher_Adj_R2, na.rm =
TRUE),
    Median_WithinPitcher_Adj_R2 = median(WithinPitcher_Adj_R2, na.rm
= TRUE),
    .groups = "drop"
  )

summary_overall <- n1_summary_tbl %>%
  group_by(Outcome) %>%
  summarise(
    N_Models = n(),
    Mean_RMSE = mean(RMSE, na.rm = TRUE),
    Mean_R2 = mean(R2, na.rm = TRUE),
    Mean_WithinPitcher_Adj_R2 = mean(WithinPitcher_Adj_R2, na.rm =
TRUE),
    Prop_AR1 = mean(BestModelType == "ar1", na.rm = TRUE),
    .groups = "drop"
  )

write_csv(summary_season, file.path(dir_14,
"14_summary_by_season.csv"))
write_csv(summary_overall, file.path(dir_14,
"14_summary_overall.csv"))

# 15. BOOTSTRAP / CONDITIONAL CONFIDENCE INTERVALS

# What is being done:

```

```

# - approximates uncertainty around individualized prediction
curves
# - not for causal claims
# - used to show plausible range around fitted N-of-1 curves

dir_15 <- file.path(out_dir, "15_bootstrap_ci")
dir.create(dir_15, recursive = TRUE, showWarnings = FALSE)

set.seed(456)
boot_rows <- list()
boot_summary_rows <- list()

for (key in names(n1_model_index)) {
  parts <- str_split(key, "_")[[1]]
  y <- parts[1]; s <- parts[2]; p <- parts[3]

  df_sp <- dat %>%
    filter(Season == s, Pitcher == p, !is.na(.data[[y]]),
!is.na(PitchCount)) %>%
    arrange(PitchCount)

  if (nrow(df_sp) < 40) next

  selected_predictors <- parse_selected_predictors(lasso_tbl, y, s,
p)
  forms <- build_candidate_formulas(y, selected_predictors, df_sp)
  if (length(forms) == 0) next

  x_grid <- seq(min(df_sp$PitchCount, na.rm = TRUE),
max(df_sp$PitchCount, na.rm = TRUE), length.out = 100)
  newdata_base <- df_sp[rep(1, length(x_grid)), , drop = FALSE]
  newdata_base$PitchCount <- x_grid

  B_boot <- 100
  boot_mat <- matrix(NA_real_, nrow = length(x_grid), ncol = B_boot)

  for (b in seq_len(B_boot)) {
    idx_b <- sample(seq_len(nrow(df_sp)), size = nrow(df_sp),
replace = TRUE)
    df_b <- df_sp[idx_b, , drop = FALSE]

    fit_b <- try(
      gamm(forms[[1]], random = list(Appearance = ~1), data = df_b,
method = "REML"),
      silent = TRUE
    )
    if (inherits(fit_b, "try-error")) next

    pred_b <- try(predict(fit_b$gam, newdata = newdata_base), silent
= TRUE)
    if (inherits(pred_b, "try-error")) next

    boot_mat[, b] <- as.numeric(pred_b)
  }
}

```

```

}

q_low <- apply(boot_mat, 1, quantile, probs = 0.025, na.rm = TRUE)
q_mid <- apply(boot_mat, 1, quantile, probs = 0.500, na.rm = TRUE)
q_hi <- apply(boot_mat, 1, quantile, probs = 0.975, na.rm = TRUE)

boot_df <- tibble(
  Outcome = y,
  Season = s,
  Pitcher = p,
  PitchCount = x_grid,
  Pred_q2.5 = q_low,
  Pred_q50 = q_mid,
  Pred_q97.5 = q_hi
)
boot_rows[[length(boot_rows) + 1]] <- boot_df

boot_summary_rows[[length(boot_summary_rows) + 1]] <- tibble(
  Outcome = y,
  Season = s,
  Pitcher = p,
  Mean_CI_Width = mean(q_hi - q_low, na.rm = TRUE)
)
}

boot_tbl <- bind_rows(boot_rows)
boot_sum_tbl <- bind_rows(boot_summary_rows)
write_csv(boot_tbl, file.path(dir_15, "15_bootstrap_curves.csv"))
write_csv(boot_sum_tbl, file.path(dir_15,
"15_bootstrap_summary.csv"))

=
# 16. FINAL EXPORTS

dir_16 <- file.path(out_dir, "16_exports")
dir.create(dir_16, recursive = TRUE, showWarnings = FALSE)

model_registry <- nl_summary_tbl %>%
  mutate(
    ModelKey = paste(Outcome, Season, Pitcher, sep = "__")
  )

write_csv(model_registry, file.path(dir_16,
"16_model_registry.csv"))

# saveRDS(nl_model_index, file.path(dir_16,
"16_nl_model_objects.rds"))
# saveRDS(group_model_index, file.path(dir_16,
"16_group_model_objects.rds"))

```



```

# 17. EXPORT CUSUM INPUTS

dir_17 <- file.path(out_dir, "17_cusum_inputs")
dir.create(dir_17, recursive = TRUE, showWarnings = FALSE)

export_cusum_inputs <- function(pred_df, outcome_name, file_stub) {
  if (nrow(pred_df) == 0) return(NULL)

  pitch_df <- pred_df %>%
    arrange(Season, Pitcher, Appearance, PitchCount) %>%
    group_by(Season, Pitcher, Appearance) %>%
    mutate(
      PitchIndexWithinAppearance = row_number(),
      NPitchesInAppearance = n()
    ) %>%
    ungroup() %>%
    group_by(Season, Pitcher) %>%
    mutate(
      AppearanceOrder = dense_rank(Appearance),
      RowIndexWithinPitcherSeason = row_number(),
      Residual_SD_Training = sd(Residual, na.rm = TRUE),
      Residual_Z = Residual / Residual_SD_Training
    ) %>%
    ungroup()

  app_df <- pitch_df %>%
    group_by(Season, Pitcher, Appearance, AppearanceOrder) %>%
    summarise(
      Outcome = outcome_name,
      NPitches = n(),
      MinPitchCount = min(PitchCount, na.rm = TRUE),
      MaxPitchCount = max(PitchCount, na.rm = TRUE),
      MeanObserved = mean(Observed, na.rm = TRUE),
      MeanPredicted = mean(Predicted, na.rm = TRUE),
      MeanResidual = mean(Residual, na.rm = TRUE),
      SDResidual = sd(Residual, na.rm = TRUE),
      MeanResidual_Z = mean(Residual_Z, na.rm = TRUE),
      .groups = "drop"
    )

  write_csv(pitch_df, file.path(dir_17, paste0("17_cusum_inputs_",
file_stub, "_pitch_level.csv")))
  write_csv(app_df, file.path(dir_17, paste0("17_cusum_inputs_",
file_stub, "_appearance_level.csv")))

  list(pitch = pitch_df, appearance = app_df)
}

# Primary applied example: velocity
velocity_pred_tbl <- pred_tbl %>% filter(Outcome == "Velocity")
cusum_velocity_exports <- export_cusum_inputs(velocity_pred_tbl,
"Velocity", "velocity")

```

```

velocity_pitch <- if (!is.null(cusum_velocity_exports))
  cusum_velocity_exports$pitch else tibble()
velocity_app <- if (!is.null(cusum_velocity_exports))
  cusum_velocity_exports$appearance else tibble()

# Conceptual biomechanics extension:
# The same residual-based export can be created for any biomechanics
# outcome. A user would typically monitor one biomechanics target at
# a
# time (for example, Max_Elb_Var_Torque) rather than combine them.
biomech_example_outcomes <- intersect(active_outcomes,
  biomech_outcomes)
cusum_biomech_exports <- list()

for (y in biomech_example_outcomes) {
  bm_pred_tbl <- pred_tbl %>% filter(Outcome == y)
  stub <- paste0("biomech_", tolower(gsub("[^A-Za-z0-9]+", "_", y)))
  cusum_biomech_exports[[y]] <- export_cusum_inputs(bm_pred_tbl, y,
  stub)
}

# 18. RETROSPECTIVE CUSUM BACKTEST

# What is being done:
# - runs retrospective monitoring using residuals
# - primary:
#   appearance-level, one-sided lower CUSUM
# - secondary:
#   pitch-level, two-sided comparisons
# - tests raw residuals and standardized residuals
# - the SAME monitoring concept can be used for performance or
#   biomechanics outcomes once a clinically/practically meaningful
#   direction of concern is defined

dir_18 <- file.path(out_dir, "18_cusum_backtest")
dir.create(dir_18, recursive = TRUE, showWarnings = FALSE)

run_lower_cusum <- function(x, k, h) {
  s <- numeric(length(x))
  signal <- logical(length(x))
  for (i in seq_along(x)) {
    prev <- if (i == 1) 0 else s[i - 1]
    s[i] <- min(0, prev + x[i] + k)
    signal[i] <- s[i] <= -h
  }
  tibble(cusum_lower = s, signal_lower = signal)
}

run_upper_cusum <- function(x, k, h) {
  s <- numeric(length(x))
  signal <- logical(length(x))

```

```

    for (i in seq_along(x)) {
      prev <- if (i == 1) 0 else s[i - 1]
      s[i] <- max(0, prev + x[i] - k)
      signal[i] <- s[i] >= h
    }
    tibble(cusum_upper = s, signal_upper = signal)
  }

summarise_sequence <- function(df_seq) {
  first_signal_idx <- which(df_seq$signal_any)[1]
  first_signal_order <- if (length(first_signal_idx) == 0) NA_real_
  else df_seq$RowIndex[first_signal_idx]
  tibble(
    N_TotalRows = nrow(df_seq),
    N_MonitorRows = sum(df_seq$IsMonitoringRow, na.rm = TRUE),
    BaselineMean = unique(df_seq$BaselineMean)[1],
    BaselineSD = unique(df_seq$BaselineSD)[1],
    N_Signals_Any = sum(df_seq$alarm_start_any, na.rm = TRUE),
    N_Signals_Lower = sum(df_seq$alarm_start_lower, na.rm = TRUE),
    N_Signals_Upper = sum(df_seq$alarm_start_upper, na.rm = TRUE),
    FirstSignalOrder = first_signal_order
  )
}

run_cusum_sequence <- function(df_seq,
                               value_col,
                               baseline_n_appearances = 3,
                               k_sd = 0,
                               h_sd = 4,
                               stat_type = c("raw", "standardized"),
                               direction = c("lower", "two_sided"))
{
  stat_type <- match.arg(stat_type)
  direction <- match.arg(direction)
  df_seq <- df_seq %>% arrange(OrderIndex)

  baseline_rows <- df_seq$AppearanceOrder <= baseline_n_appearances
  if (sum(baseline_rows, na.rm = TRUE) < 2) return(NULL)

  x_all <- df_seq[[value_col]]
  mu0 <- mean(x_all[baseline_rows], na.rm = TRUE)
  sd0 <- sd(x_all[baseline_rows], na.rm = TRUE)

  if (!is.finite(mu0) || !is.finite(sd0) || sd0 <= 0) return(NULL)

  if (stat_type == "raw") {
    x <- x_all - mu0
    k <- k_sd * sd0
    h <- h_sd * sd0
  } else {
    x <- (x_all - mu0) / sd0
    k <- k_sd
  }
}

```

```

    h <- h_sd
  }

  monitor_rows <- !baseline_rows
  x_monitor <- x[monitor_rows]
  if (length(x_monitor) == 0) return(NULL)

  lower_res <- run_lower_cusum(x_monitor, k = k, h = h)
  if (direction == "two_sided") {
    upper_res <- run_upper_cusum(x_monitor, k = k, h = h)
  } else {
    upper_res <- tibble(cusum_upper = rep(NA_real_,
length(x_monitor)),
                        signal_upper = rep(FALSE,
length(x_monitor)))
  }

  out <- df_seq[monitor_rows, , drop = FALSE] %>%
    mutate(
      ValueCentered = x_monitor,
      cusum_lower = lower_res$cusum_lower,
      signal_lower = lower_res$signal_lower,
      cusum_upper = upper_res$cusum_upper,
      signal_upper = upper_res$signal_upper,
      signal_any = signal_lower | signal_upper,
      BaselineMean = mu0,
      BaselineSD = sd0,
      H_Used = h,
      K_Used = k,
      IsMonitoringRow = TRUE
    )

  signal_lag <- dplyr::lag(out$signal_any, default = FALSE)
  out$alarm_start_any <- out$signal_any & !signal_lag
  out$alarm_start_lower <- out$signal_lower &
!dplyr::lag(out$signal_lower, default = FALSE)
  out$alarm_start_upper <- out$signal_upper &
!dplyr::lag(out$signal_upper, default = FALSE)
  out
}

if (exists("velocity_app") && nrow(velocity_app) > 0) {
  app_backtest_rows <- list()

  keys_app <- velocity_app %>% distinct(Season, Pitcher)

  for (i in seq_len(nrow(keys_app))) {
    s <- keys_app$Season[i]
    p <- keys_app$Pitcher[i]

    df_seq <- velocity_app %>%
      filter(Season == s, Pitcher == p) %>%
      arrange(AppearanceOrder) %>%

```

```

mutate(OrderIndex = row_number())

res_raw <- run_cusum_sequence(
  df_seq = df_seq,
  value_col = "MeanResidual",
  baseline_n_appearances = 3,
  k_sd = 0,
  h_sd = 4,
  stat_type = "raw",
  direction = "lower"
)

if (!is.null(res_raw)) {
  res_raw$Season <- s
  res_raw$Pitcher <- p
  res_raw$BacktestLevel <- "appearance"
  res_raw$StatisticType <- "raw"
  res_raw$DirectionType <- "lower"
  app_backtest_rows[[length(app_backtest_rows) + 1]] <- res_raw
}

res_std <- run_cusum_sequence(
  df_seq = df_seq,
  value_col = "MeanResidual",
  baseline_n_appearances = 3,
  k_sd = 0,
  h_sd = 4,
  stat_type = "standardized",
  direction = "lower"
)

if (!is.null(res_std)) {
  res_std$Season <- s
  res_std$Pitcher <- p
  res_std$BacktestLevel <- "appearance"
  res_std$StatisticType <- "standardized"
  res_std$DirectionType <- "lower"
  app_backtest_rows[[length(app_backtest_rows) + 1]] <- res_std
}

app_backtest_tbl <- bind_rows(app_backtest_rows)
write_csv(app_backtest_tbl, file.path(dir_18,
"18_cusum_backtest_appearance_level.csv"))

if (nrow(app_backtest_tbl) > 0) {
  signal_summary_tbl <- app_backtest_tbl %>%
    group_by(BacktestLevel, StatisticType, DirectionType, Season,
Pitcher) %>%
    group_modify(~ summarise_sequence(.x)) %>%
    ungroup()
  write_csv(signal_summary_tbl, file.path(dir_18,
"18_cusum_signal_summary.csv"))
}

```

```

    }
  }

# 19. LIMITED CUSUM GRID-SEARCH VALIDATION

# What is being done:
#   - tunes practical CUSUM settings retrospectively
#   - tests baseline window, min pitches, k, h, and residual scale
#   - uses ranking logic rather than claiming one universally
#     optimal surveillance parameter set
#
# Dissertation design note:
#   The stated plan was limited grid-search validation, not an
#   unrestricted search over every possible control parameter.

dir_19 <- file.path(out_dir, "19_cusum_gridsearch")
dir.create(dir_19, recursive = TRUE, showWarnings = FALSE)

grid_rows <- list()

if (exists("velocity_app") && nrow(velocity_app) > 0) {
  for (baseline_n in c(3, 4, 5)) {
    for (h_sd in c(4, 6, 8)) {
      for (stat_type in c("raw", "standardized")) {

        tmp_rows <- list()
        keys_app <- velocity_app %>% distinct(Season, Pitcher)

        for (i in seq_len(nrow(keys_app))) {
          s <- keys_app$Season[i]
          p <- keys_app$Pitcher[i]

          df_seq <- velocity_app %>%
            filter(Season == s, Pitcher == p) %>%
            arrange(AppearanceOrder) %>%
            mutate(OrderIndex = row_number())

          res <- run_cusum_sequence(
            df_seq = df_seq,
            value_col = "MeanResidual",
            baseline_n_appearances = baseline_n,
            k_sd = 0,
            h_sd = h_sd,
            stat_type = stat_type,
            direction = "lower"
          )

          if (!is.null(res)) {
            tmp_rows[[length(tmp_rows) + 1]] <- res %>%
              summarise(
                NMonitor = n(),
                AnySignal = any(signal_any, na.rm = TRUE),

```

```

        FirstSignalOrder =
suppressWarnings(min(OrderIndex[signal_any], na.rm = TRUE))
        ) %>%
        mutate(Season = s, Pitcher = p)
    }
}

if (length(tmp_rows) > 0) {
  tmp_tbl <- bind_rows(tmp_rows)
  grid_rows[[length(grid_rows) + 1]] <- tmp_tbl %>%
  summarise(
    BaselineN = baseline_n,
    H_SD = h_sd,
    StatisticType = stat_type,
    Coverage = mean(NMonitor > 0, na.rm = TRUE),
    SignalRate = mean(AnySignal, na.rm = TRUE),
    MedianFirstSignal = median(FirstSignalOrder, na.rm =
TRUE)
  )
}
}
}
}

grid_tbl <- bind_rows(grid_rows) %>%
  mutate(
    RankScore = rank(-Coverage) + rank(SignalRate) +
rank(MedianFirstSignal, na.last = "keep")
  ) %>%
  arrange(RankScore)

write_csv(grid_tbl, file.path(dir_19,
"19_gridsearch_ranked_primary.csv"))

# 20. FINAL CUSUM OUTPUTS

# What is being done:
# - selects the final preferred CUSUM specification
# - checks whether the current backtest matches the chosen spec
# - exports final tables and figures suitable for manuscript or
# presentation use

dir_20 <- file.path(out_dir, "20_cusum_final")
dir.create(dir_20, recursive = TRUE, showWarnings = FALSE)

if (exists("grid_tbl") && nrow(grid_tbl) > 0 &&
exists("app_backtest_tbl") && nrow(app_backtest_tbl) > 0) {
  final_spec <- grid_tbl %>% slice(1)
  write_csv(final_spec, file.path(dir_20,
"20_final_selected_specification.csv"))
}

```

```

final_h <- final_spec$H_SD[1]
final_stat <- final_spec$StatisticType[1]
final_baseline <- final_spec$BaselineN[1]

final_app <- list()
keys_app <- velocity_app %>% distinct(Season, Pitcher)

for (i in seq_len(nrow(keys_app))) {
  s <- keys_app$Season[i]
  p <- keys_app$Pitcher[i]

  df_seq <- velocity_app %>%
    filter(Season == s, Pitcher == p) %>%
    arrange(AppearanceOrder) %>%
    mutate(OrderIndex = row_number())

  res <- run_cusum_sequence(
    df_seq = df_seq,
    value_col = "MeanResidual",
    baseline_n_appearances = final_baseline,
    k_sd = 0,
    h_sd = final_h,
    stat_type = final_stat,
    direction = "lower"
  )

  if (!is.null(res)) {
    res$Season <- s
    res$Pitcher <- p
    final_app[[length(final_app) + 1]] <- res
  }
}

final_app_tbl <- bind_rows(final_app)
write_csv(final_app_tbl, file.path(dir_20,
"20_final_backtest_rows.csv"))

final_primary_summary <- final_app_tbl %>%
  group_by(Season, Pitcher) %>%
  group_modify(~ summarise_sequence(.x)) %>%
  ungroup()
write_csv(final_primary_summary, file.path(dir_20,
"20_final_primary_summary.csv"))

final_signal_counts <- final_primary_summary %>%
  count(Season, Pitcher, wt = N_Signals_Any, name = "SignalCount")
write_csv(final_signal_counts, file.path(dir_20,
"20_final_signal_counts_by_pitcher.csv"))

top_sequences <- final_app_tbl %>%
  filter(signal_any) %>%
  group_by(Season, Pitcher) %>%

```



```

    summarise(
      FirstSignalOrder = min(OrderIndex, na.rm = TRUE),
      MinCUSUM = min(cusum_lower, na.rm = TRUE),
      .groups = "drop"
    ) %>%
    arrange(FirstSignalOrder, MinCUSUM)
  write_csv(top_sequences, file.path(dir_20,
    "20_final_top_sequences.csv"))

  example_pitcher <- final_app_tbl %>%
    group_by(Pitcher) %>%
    summarise(NSignals = sum(signal_any, na.rm = TRUE), .groups =
"drop") %>%
    arrange(desc(NSignals)) %>%
    slice(1) %>%
    pull(Pitcher)

  if (length(example_pitcher) == 1 && !is.na(example_pitcher)) {
    ex_df <- final_app_tbl %>% filter(Pitcher == example_pitcher)
    p_ex <- ggplot(ex_df, aes(OrderIndex, cusum_lower)) +
      geom_line() +
      geom_point(aes(shape = signal_any), size = 2) +
      geom_hline(aes(yintercept = H_Used), linetype = 2) +
      theme_minimal(base_size = 12) +
      labs(
        title = paste("Final appearance-level lower CUSUM:",
example_pitcher),
        subtitle = "Velocity residual monitoring based on N-of-1
expected performance",
        x = "Appearance order after baseline",
        y = "Lower CUSUM"
      )
    ggsave(file.path(dir_20, "20_example_pitcher_final_cusum.png"),
p_ex, width = 9, height = 6, dpi = 300)
  }
}

```

```

# FINAL NOTES FOR DEMO USE

```

```

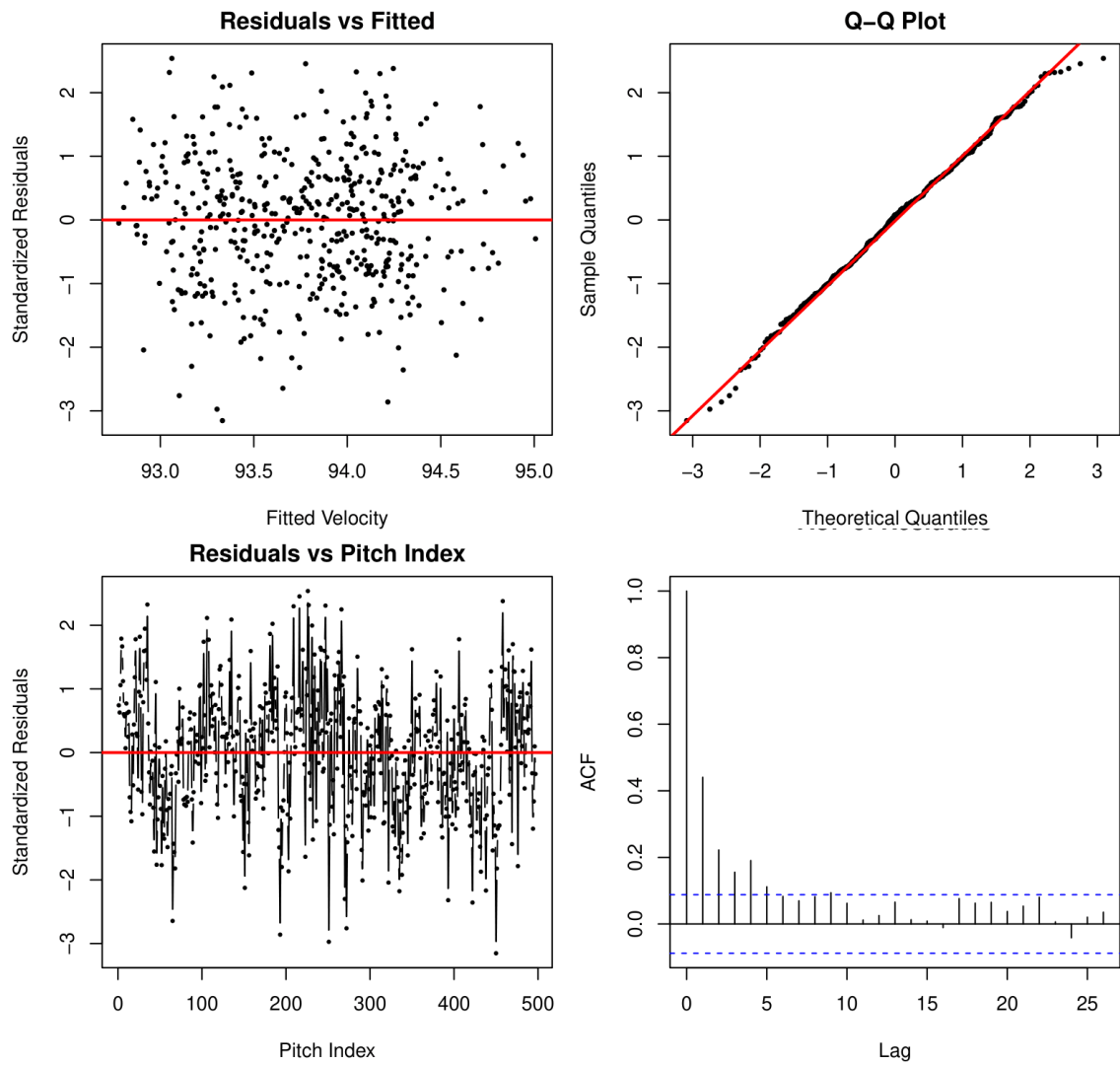
# 1. This file is intentionally a long-form, pedagogic version of
#   the original pipeline rather than a line-for-line copy.
# 2. It combines performance, biomechanics, and downstream CUSUM.
# 3. The individualized model target is always the pitcher-season.
# 4. The CUSUM workflow is downstream and velocity-specific.
# 5. De-identification is applied at the start and reused
everywhere.
# 6. The comments are written to explain the logic for diffent
stages

```

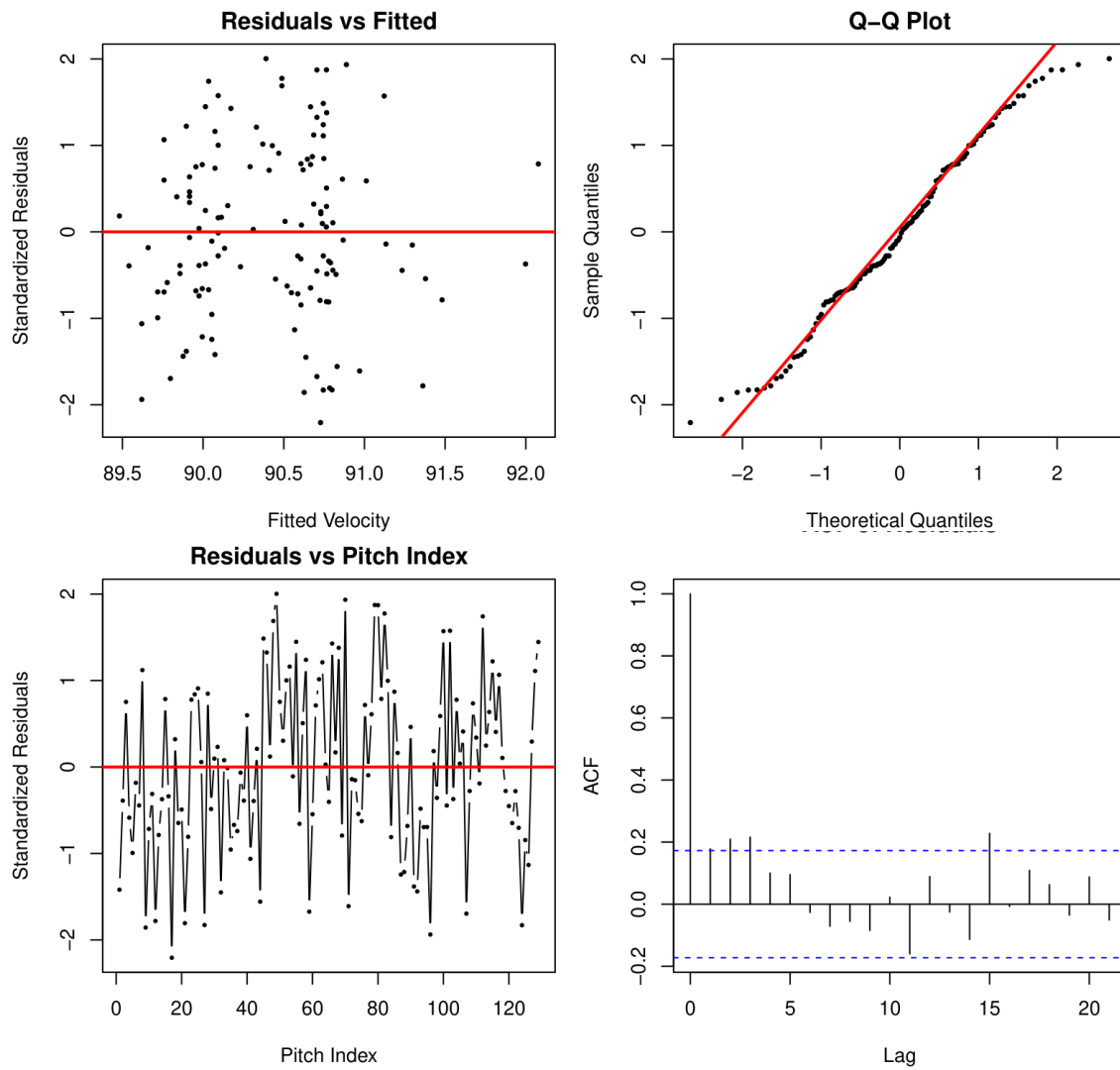
Appendix E – Diagnostic Plot Sample

Because more than 1,000 diagnostic plots were generated for this project, not all are included.

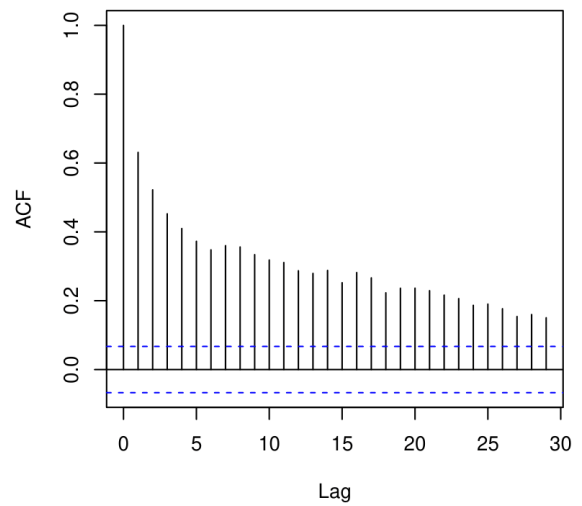
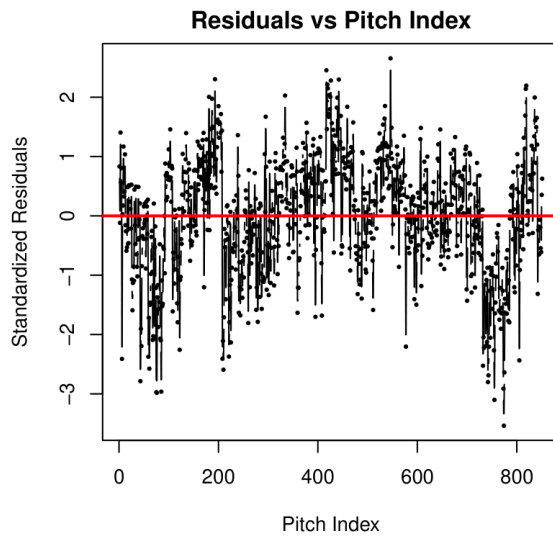
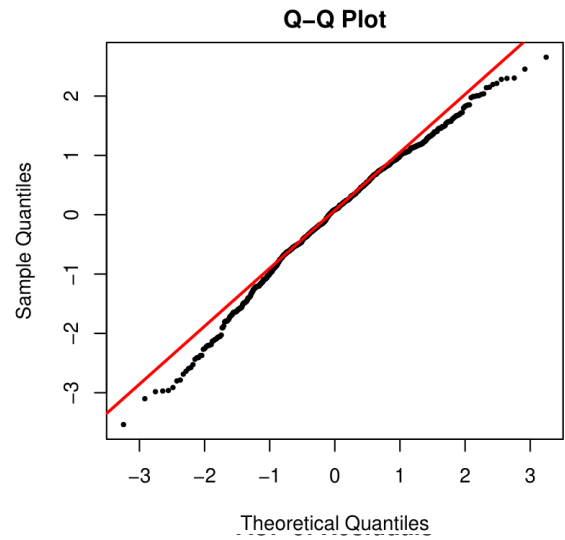
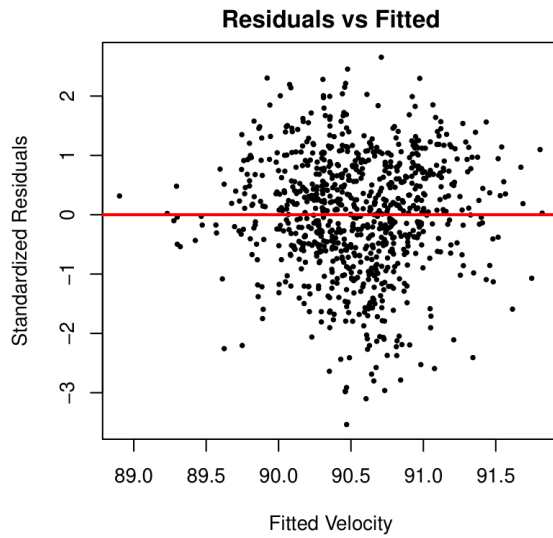
Player 1 | Velocity | Season 2023



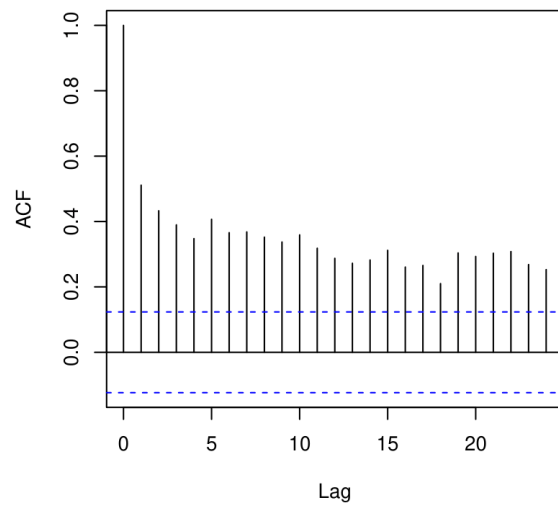
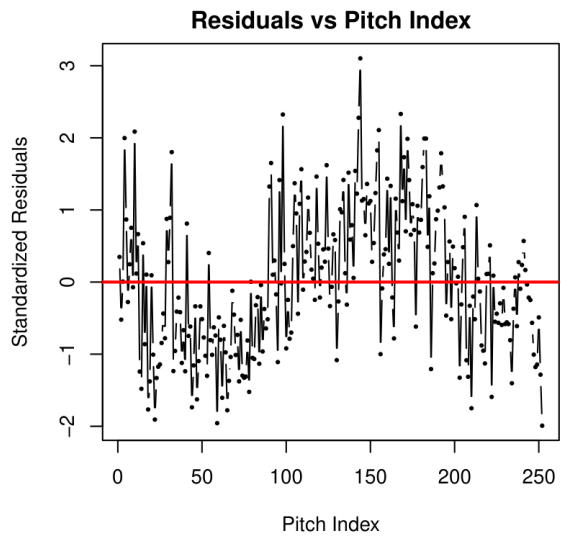
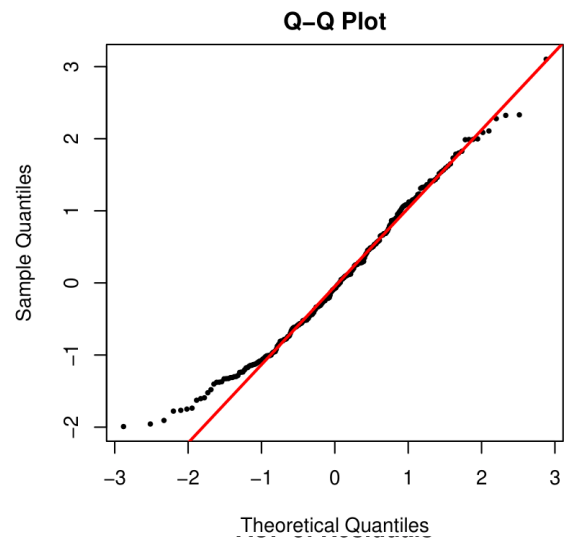
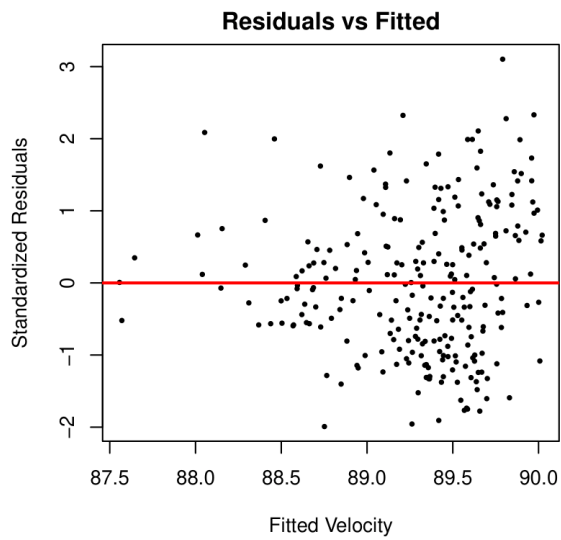
Player 2 | Velocity | Season 2023



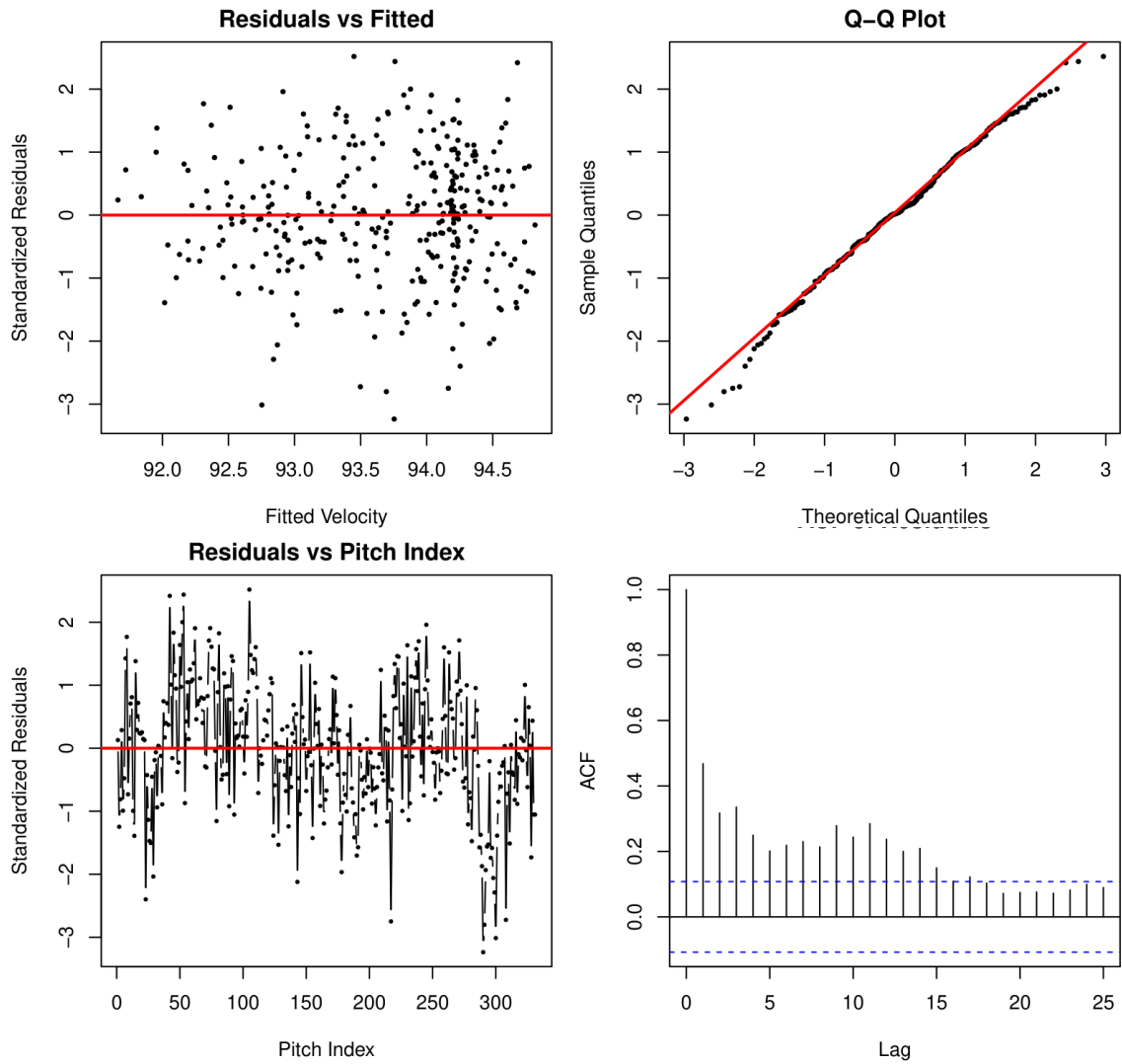
Player 3 | Velocity | Season 2023



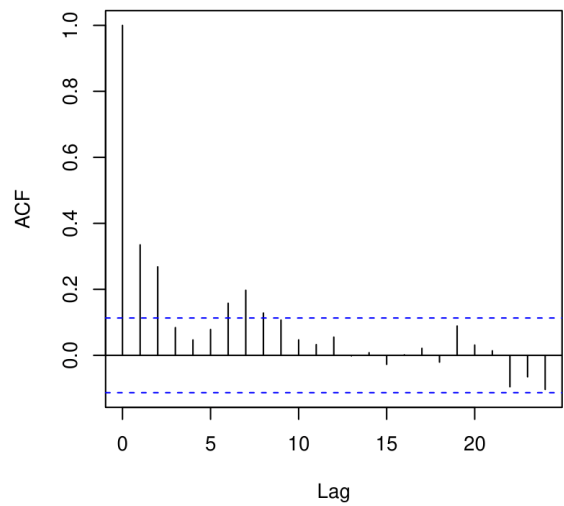
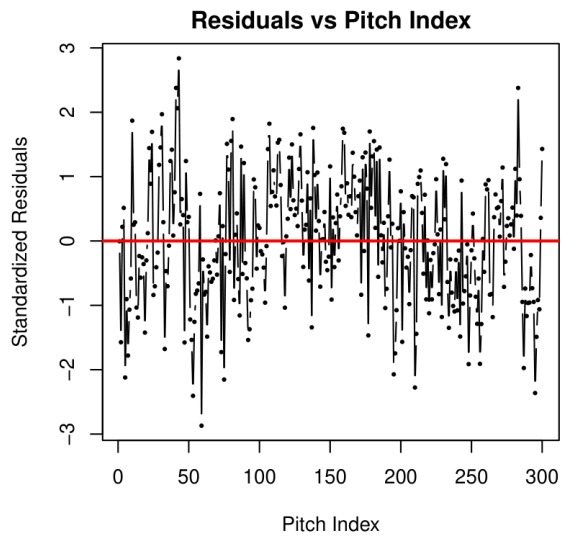
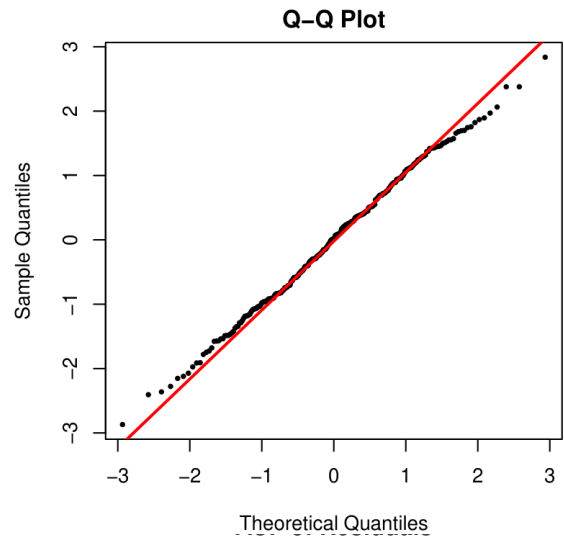
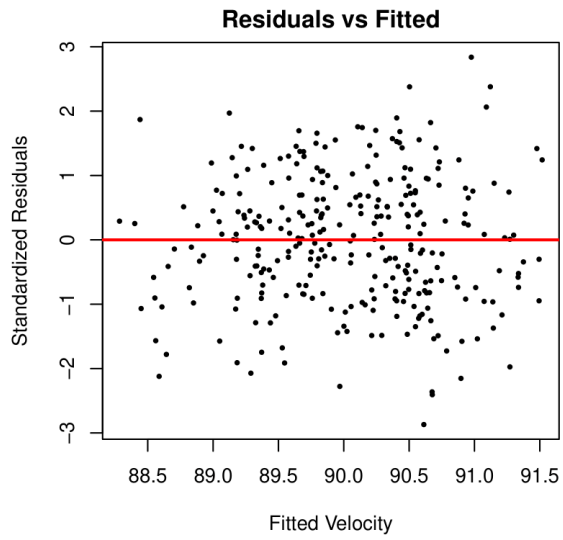
Player 4 | Velocity | Season 2023



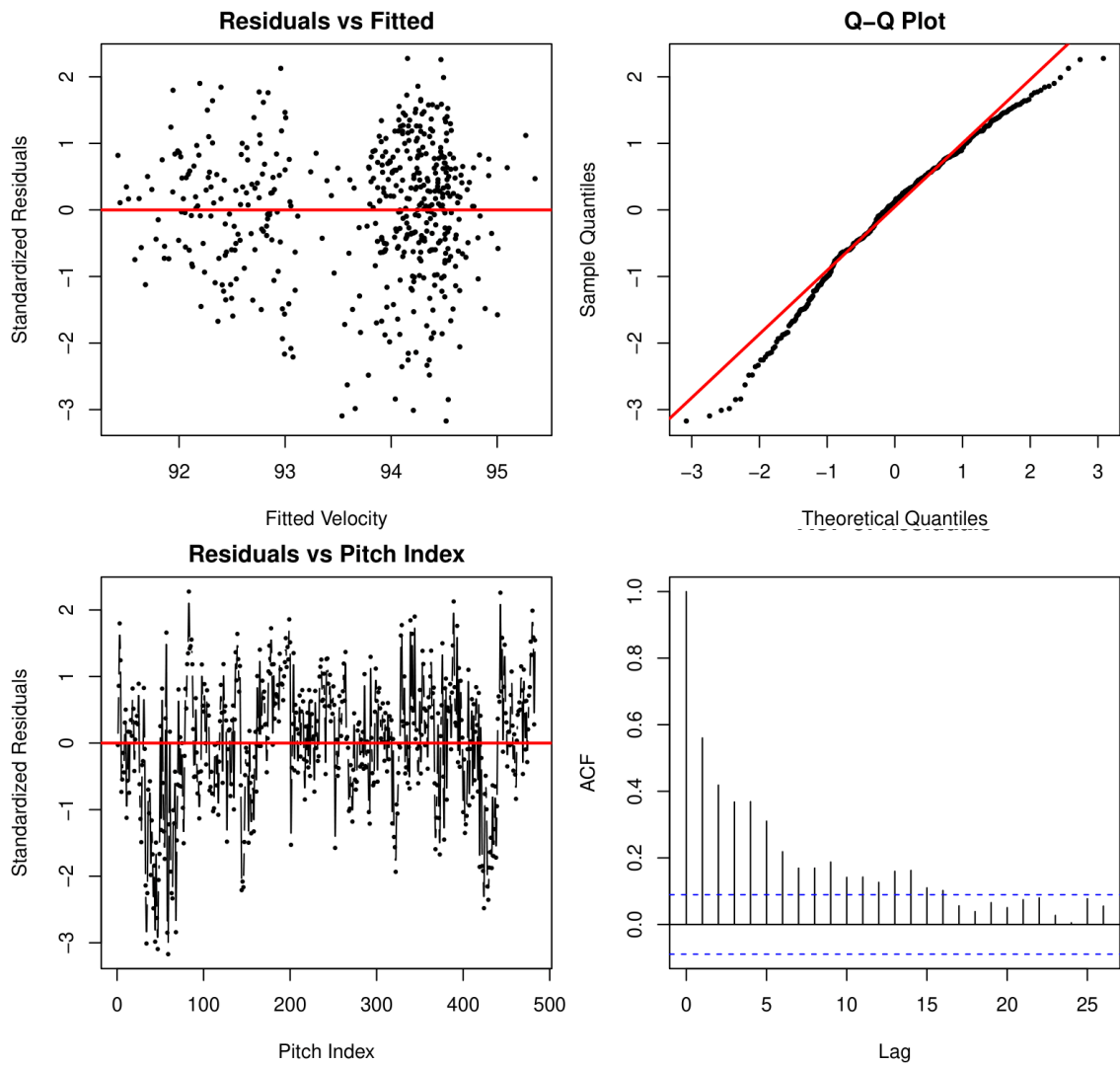
Player 5 | Velocity | Season 2023



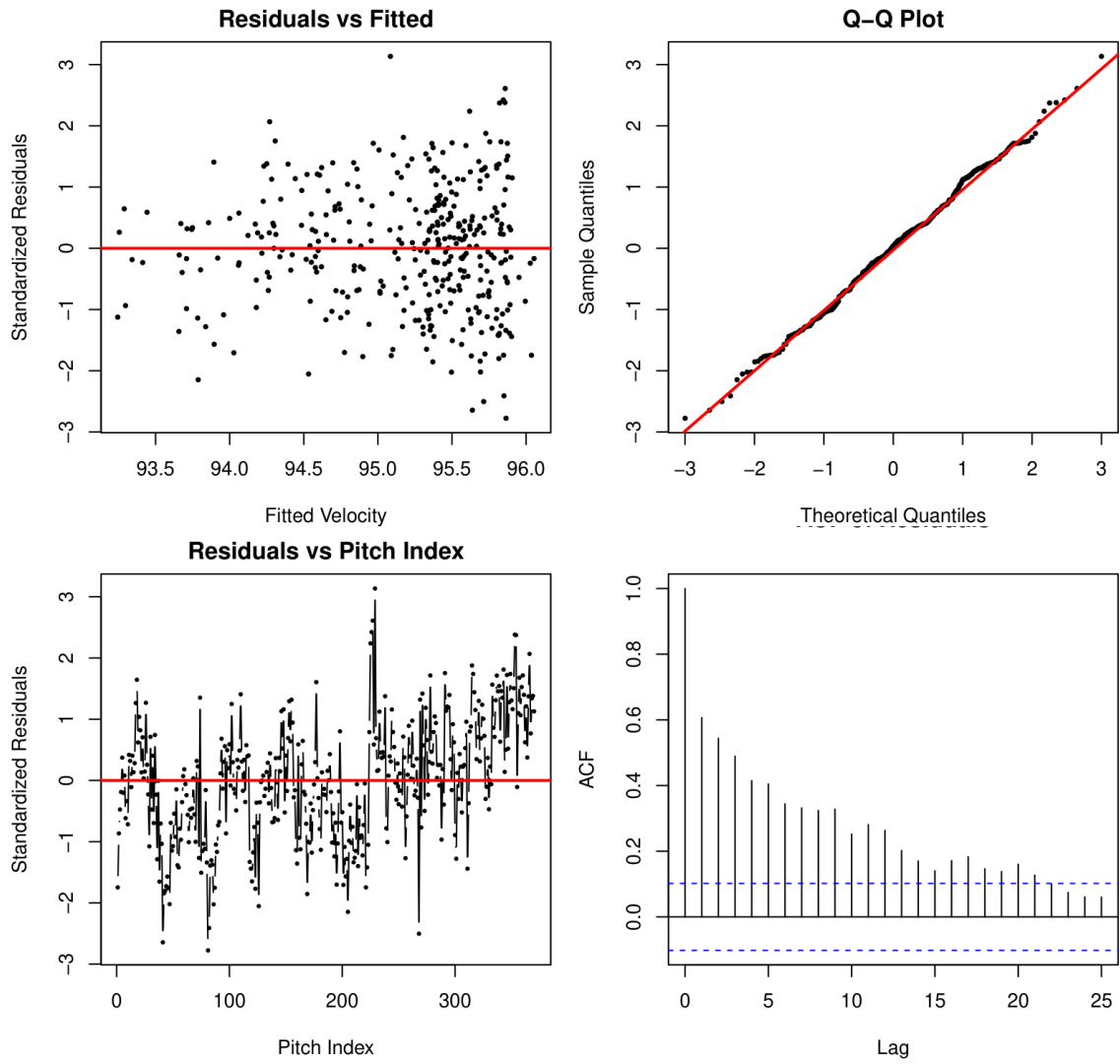
Player 6 | Velocity | Season 2023



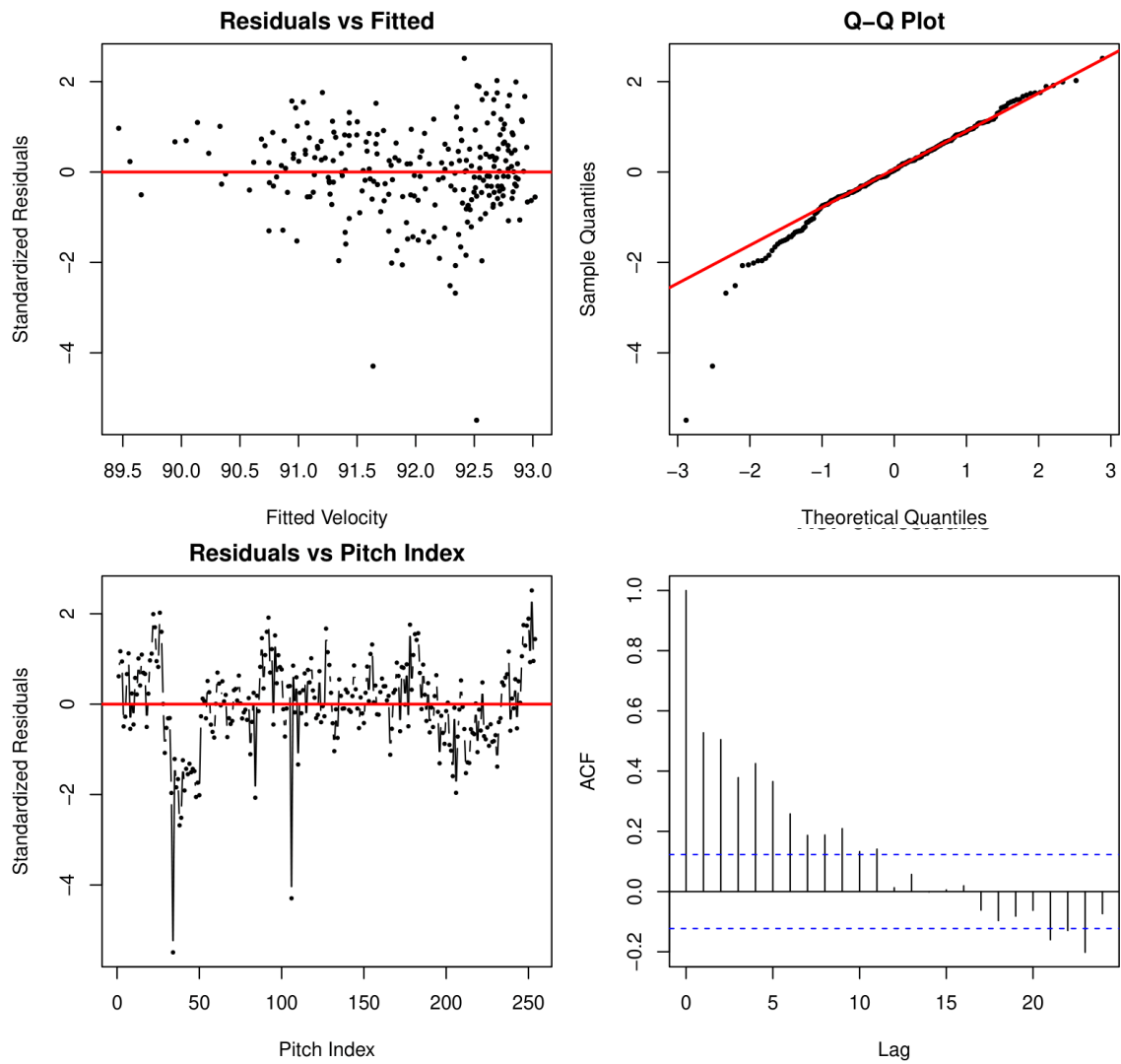
Player 7 | Velocity | Season 2023



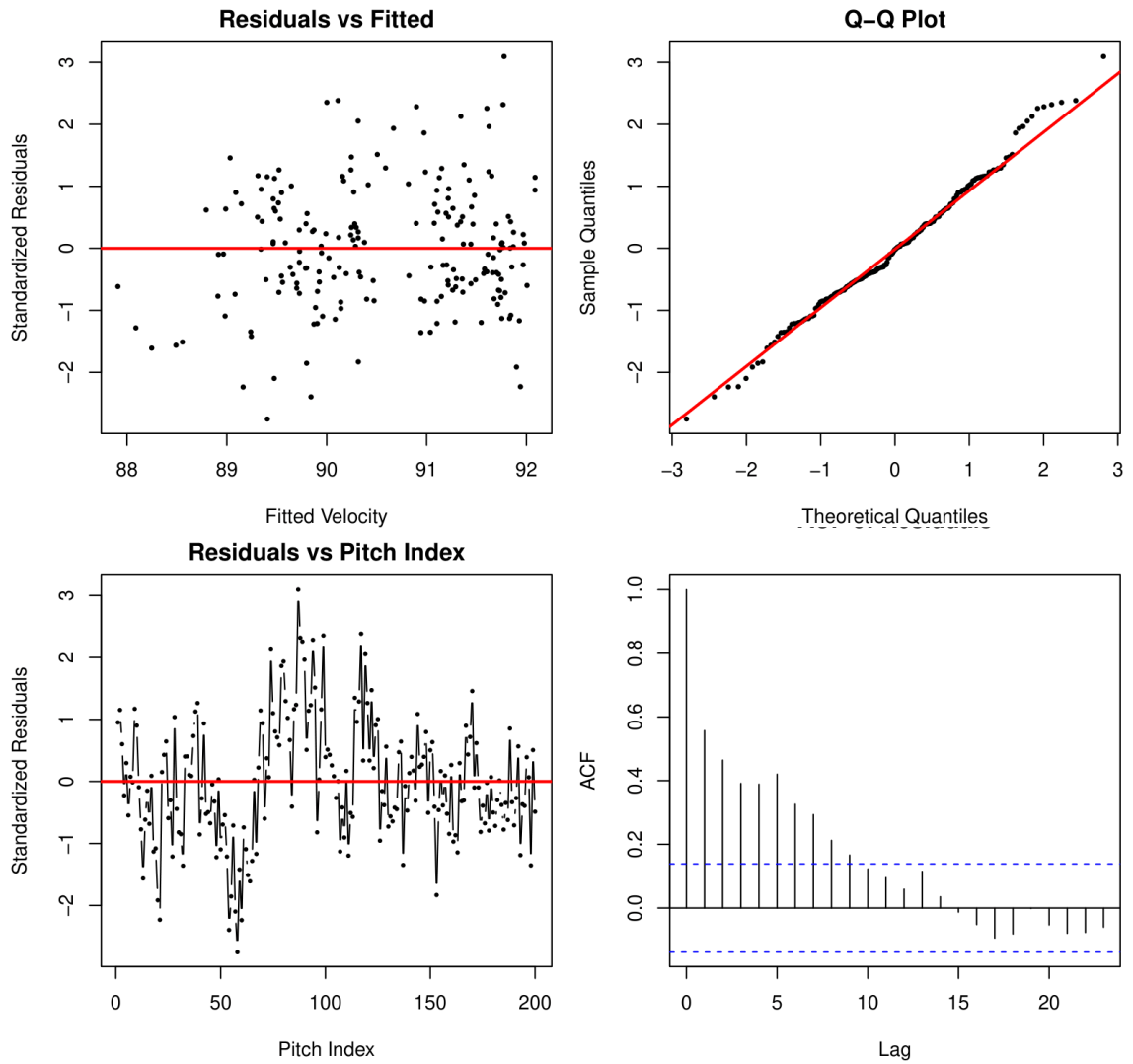
Player 8 | Velocity | Season 2023



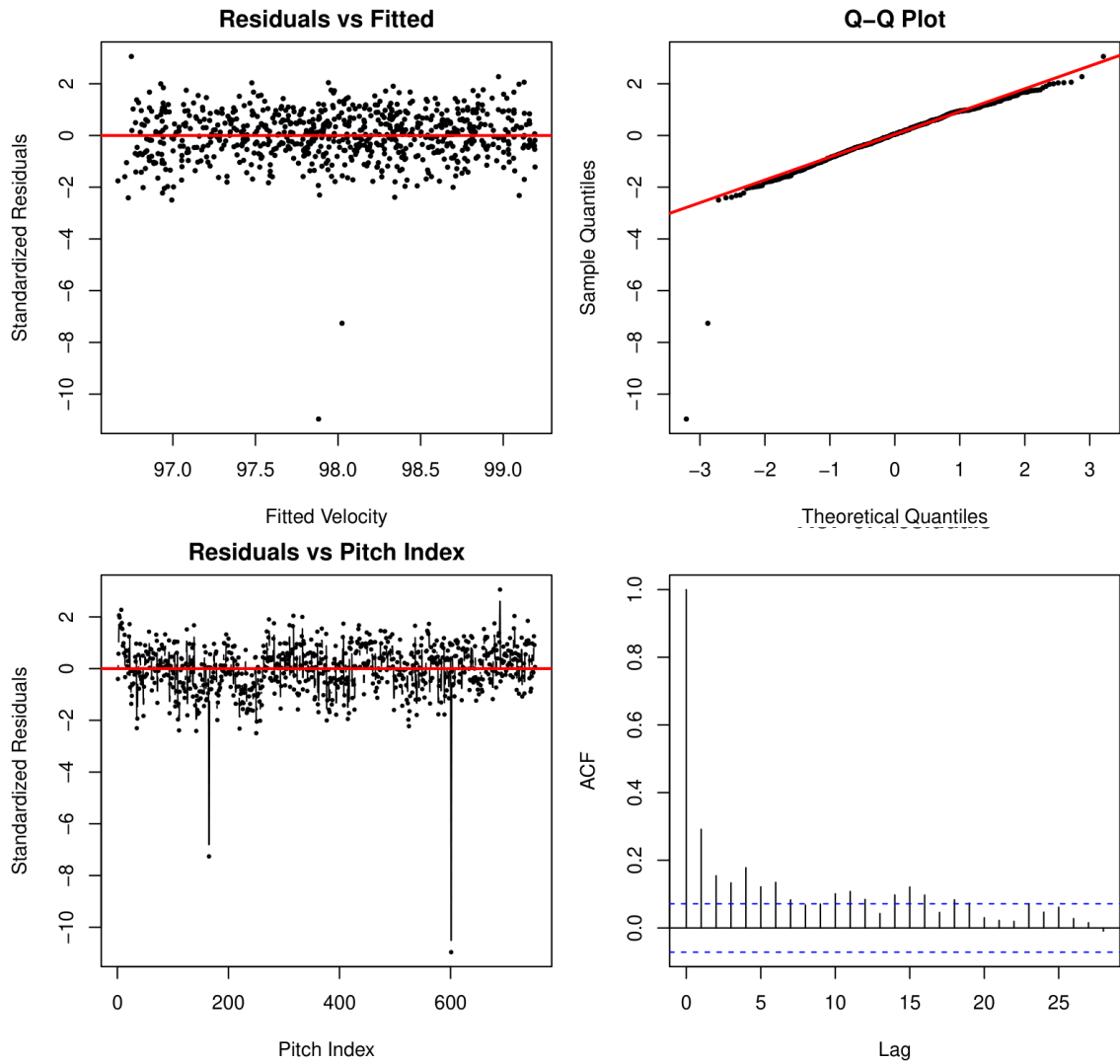
Player 9 | Velocity | Season 2023



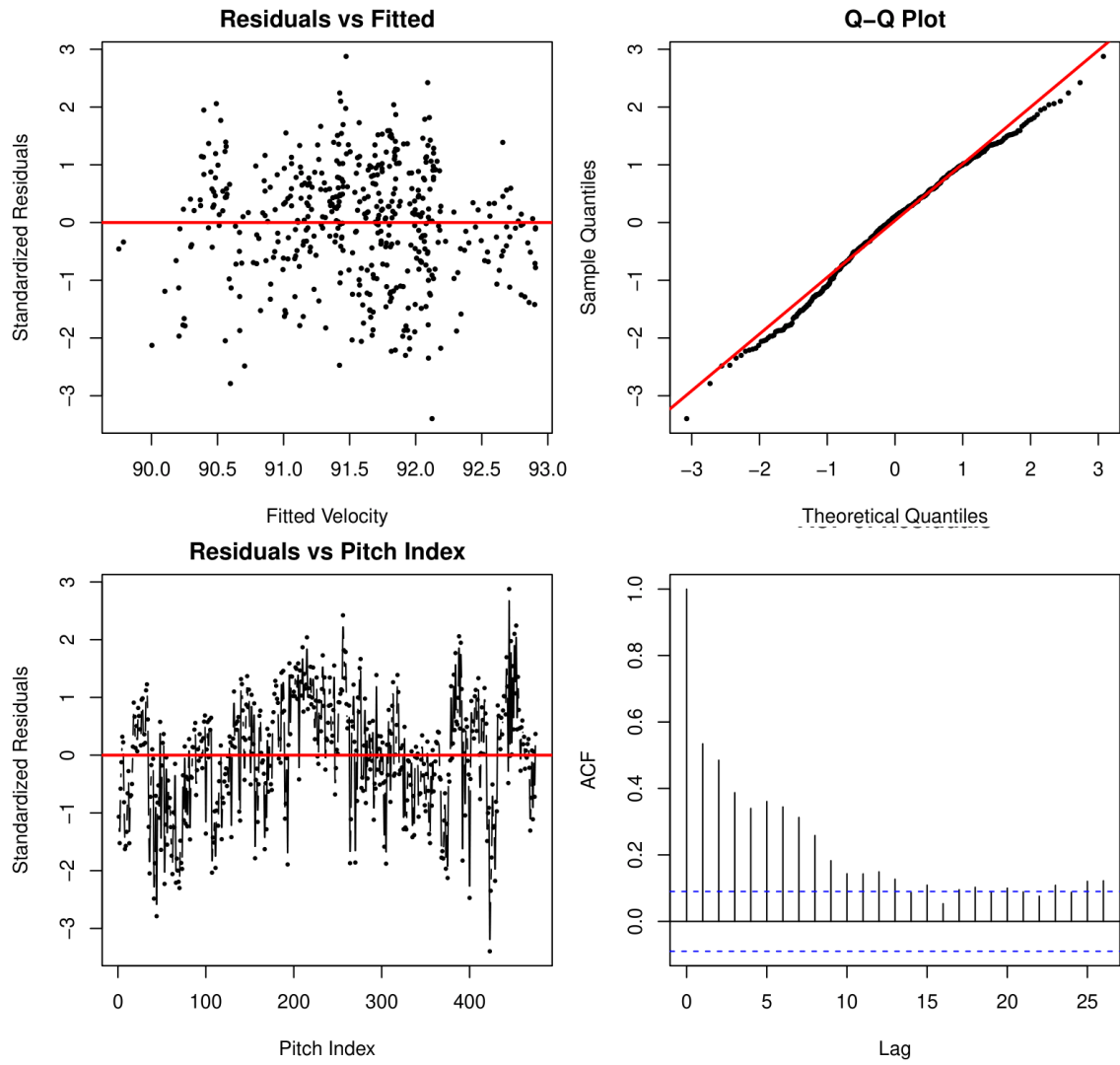
Player 10 | Velocity | Season 2023

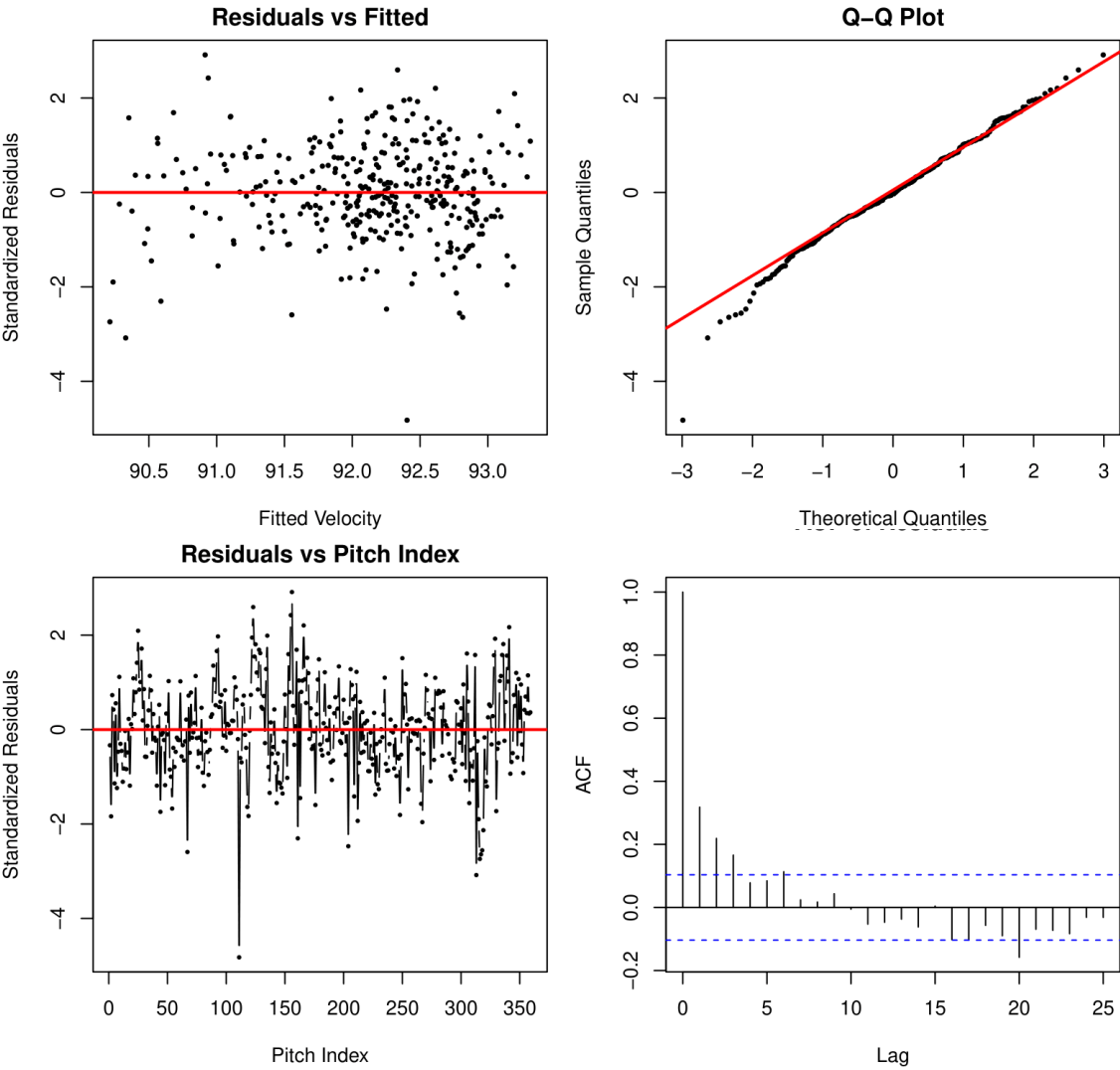


Player 11 | Velocity | Season 2024

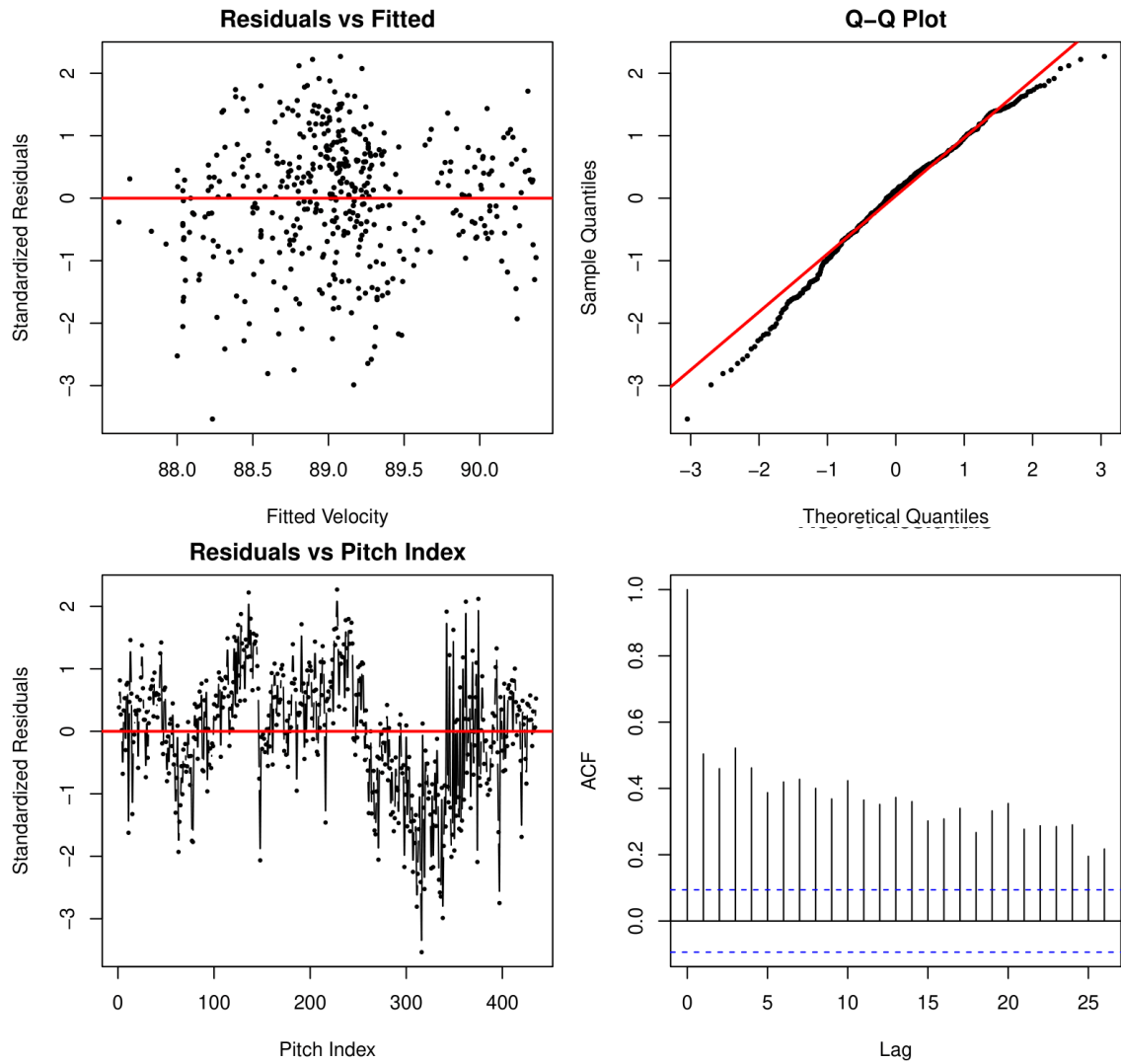


Player 12 | Velocity | Season 2024

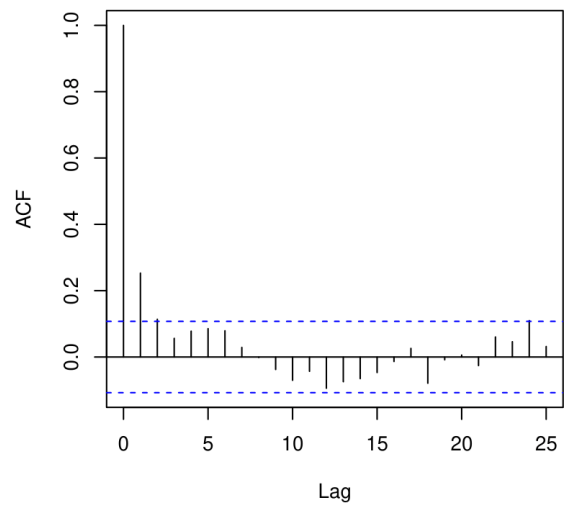
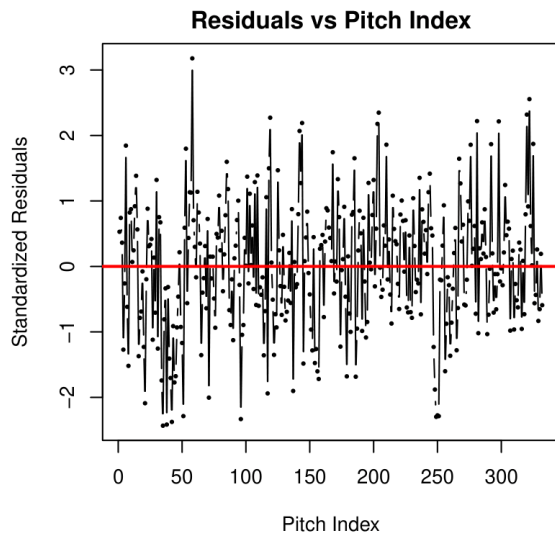
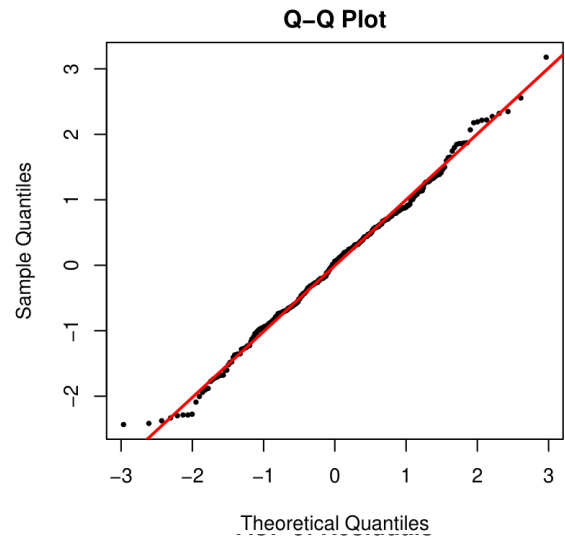
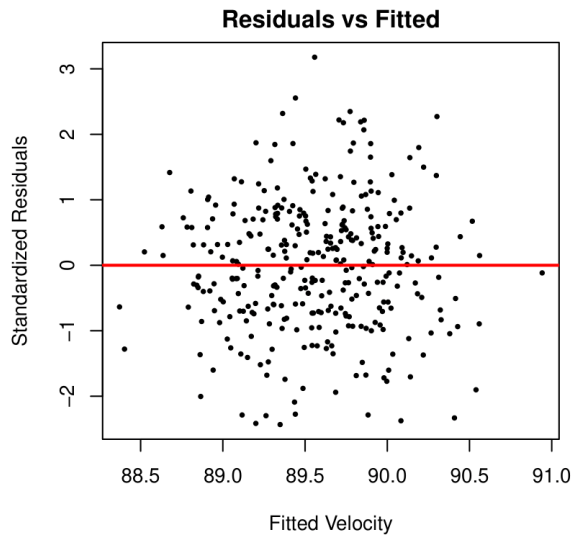




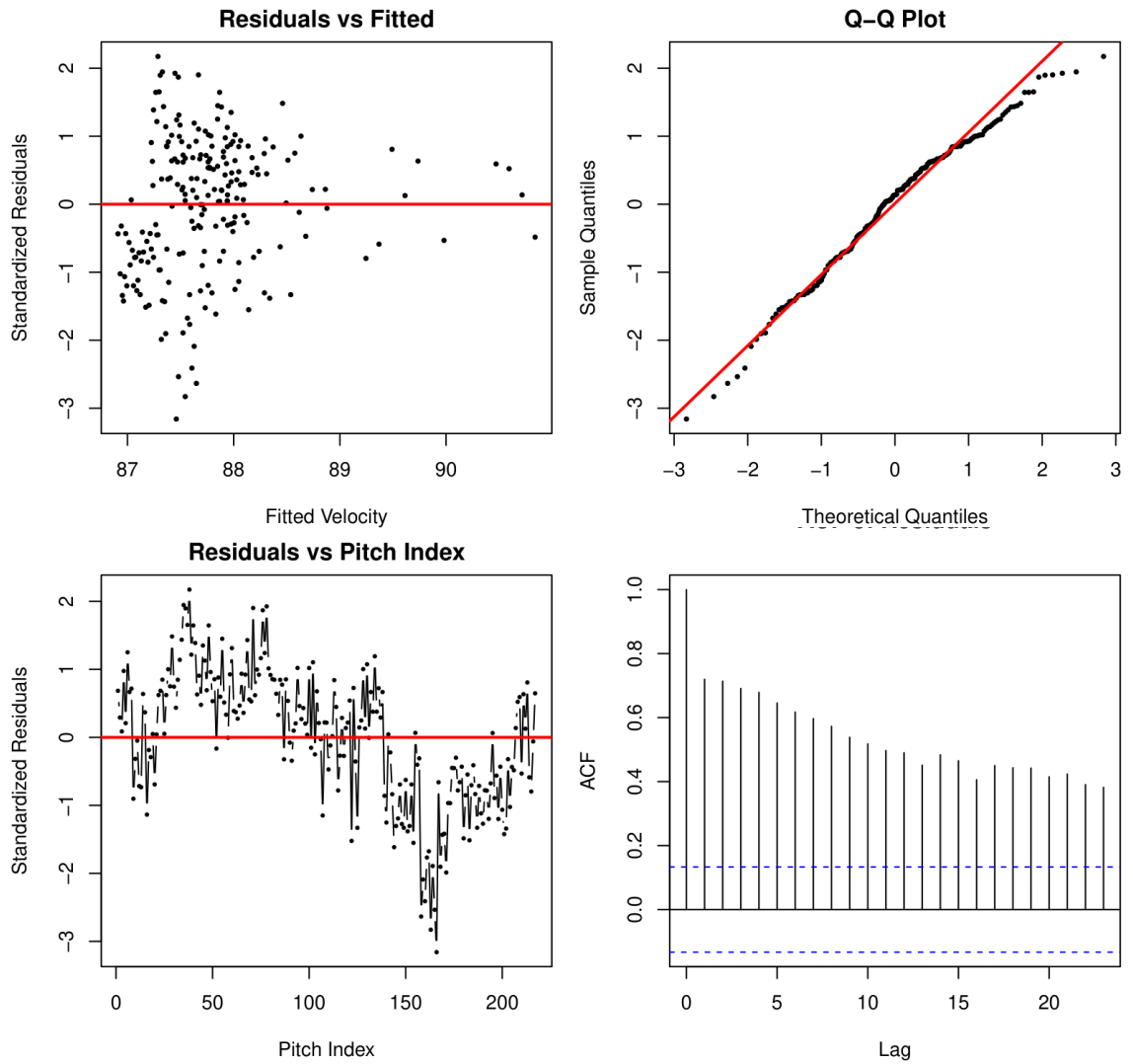
Player 14 | Velocity | Season 2024



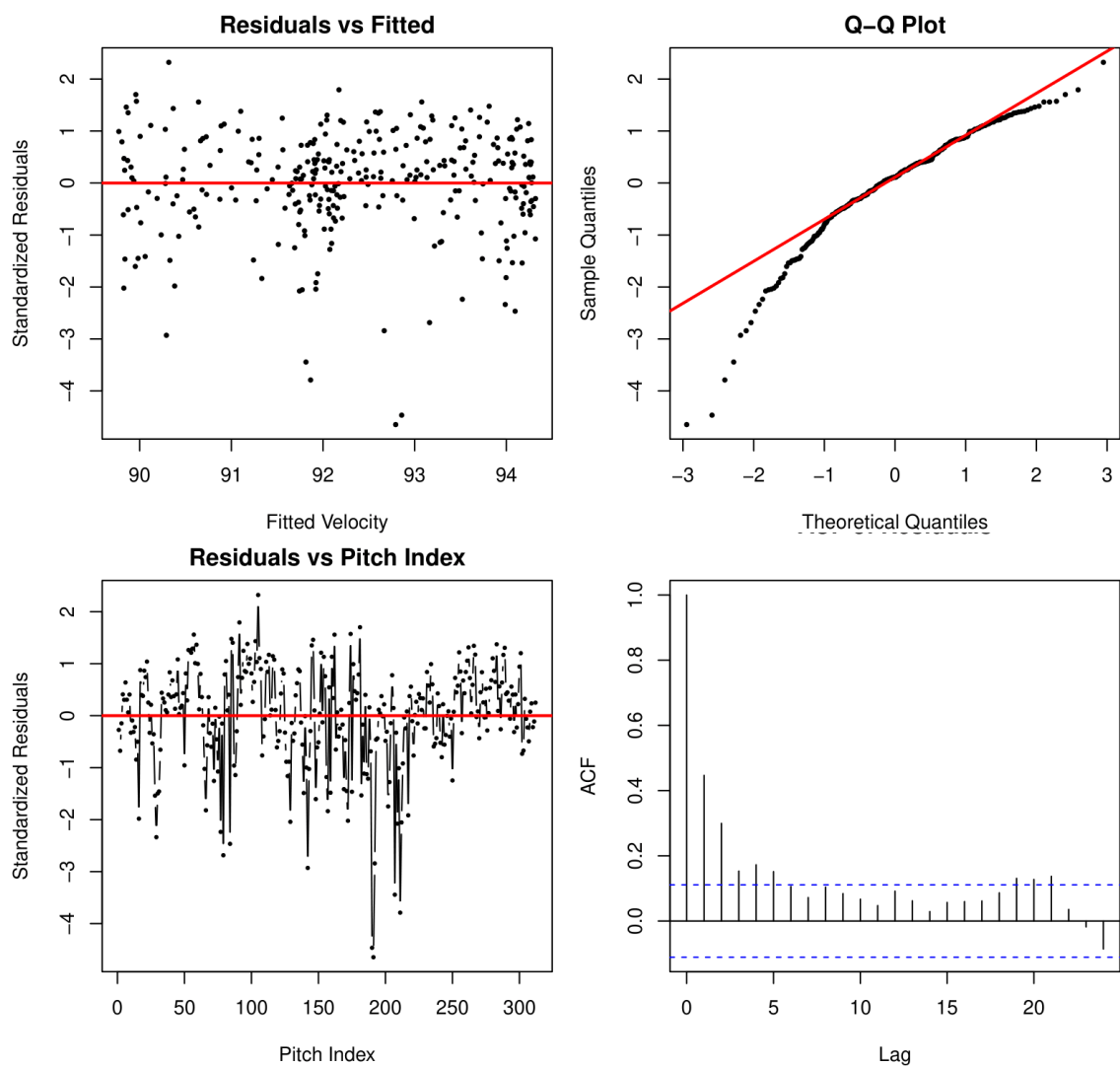
Player 15 | Velocity | Season 2024



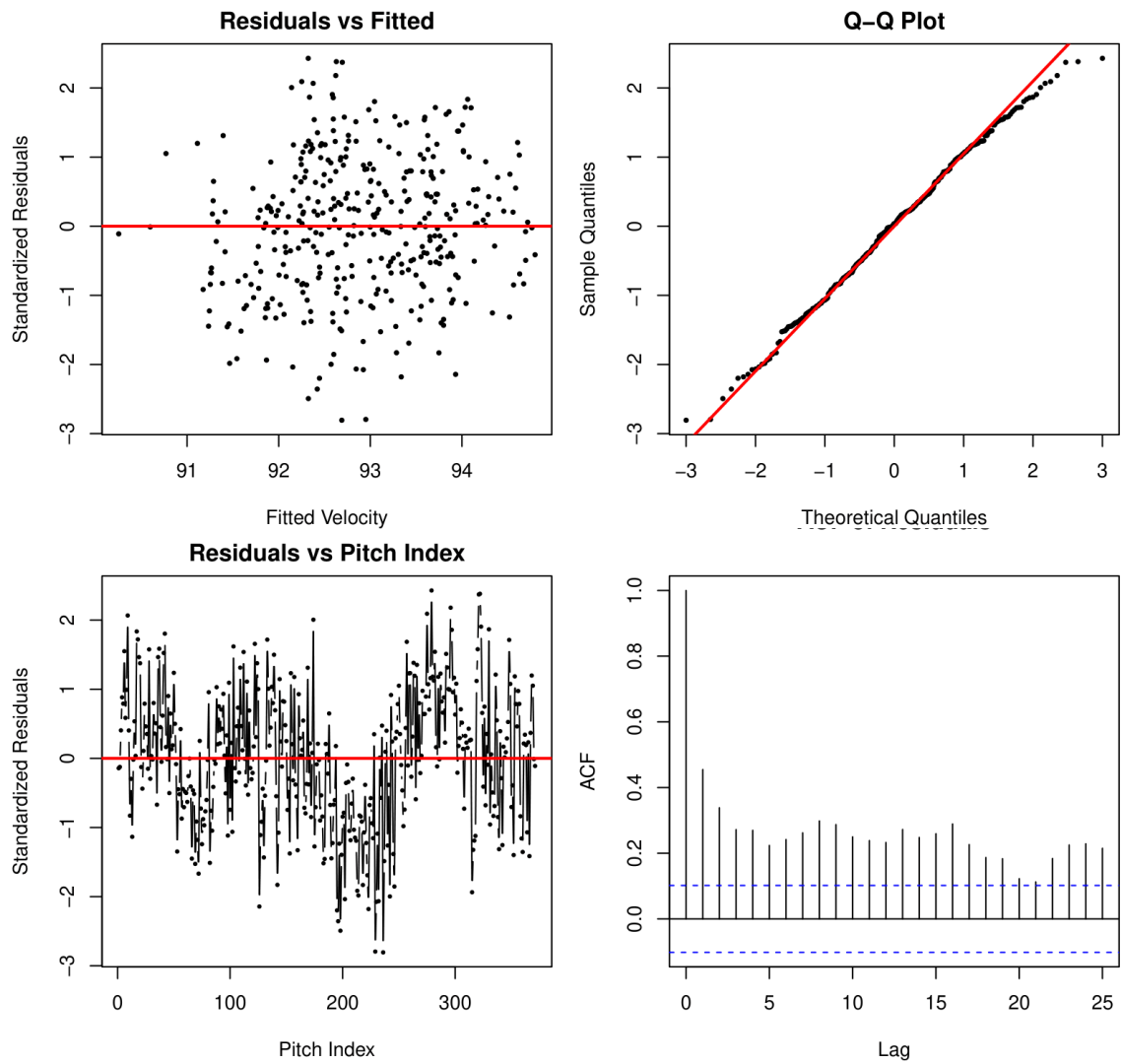
Player 16 | Velocity | Season 2024



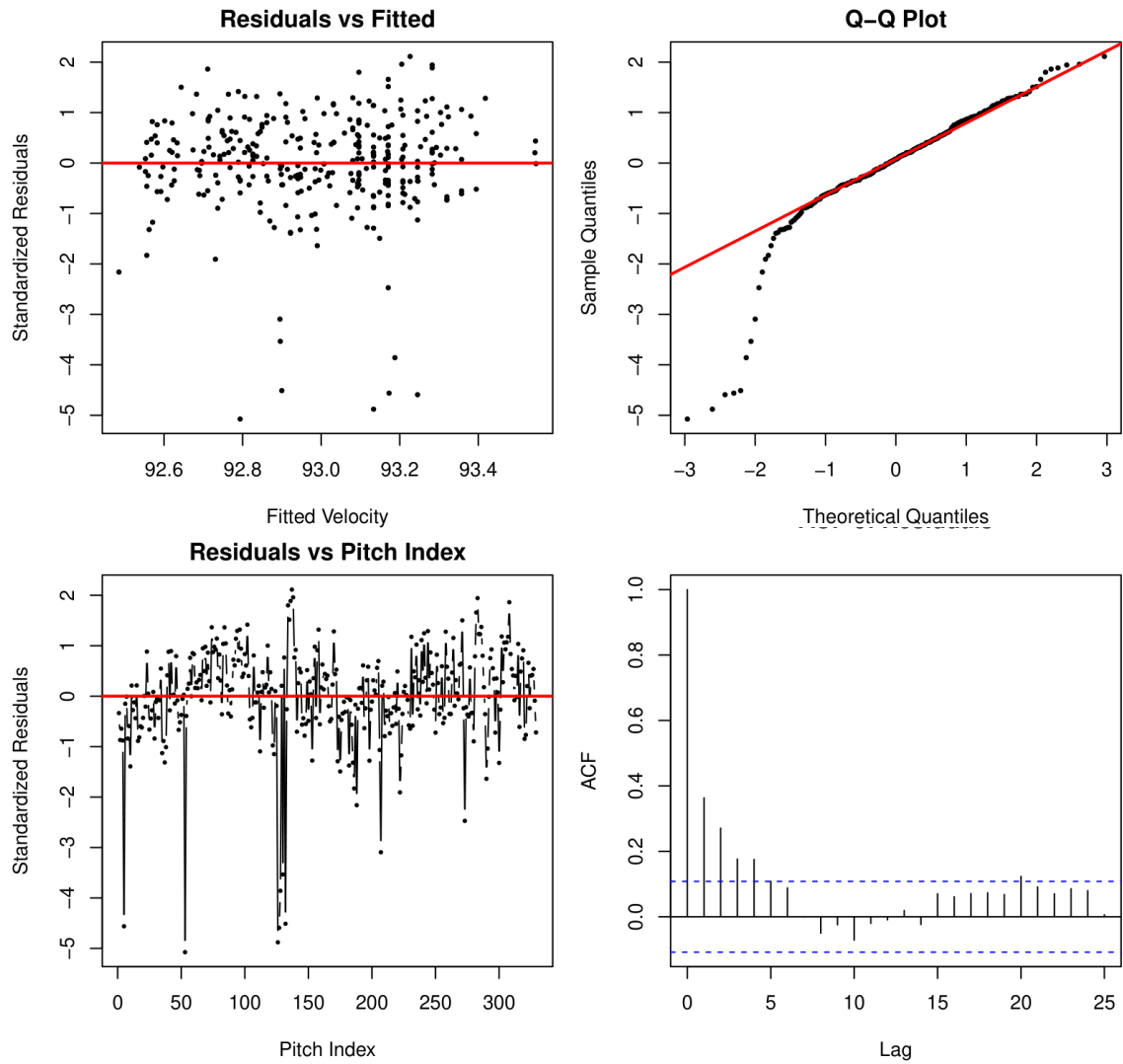
Player 17 | Velocity | Season 2024

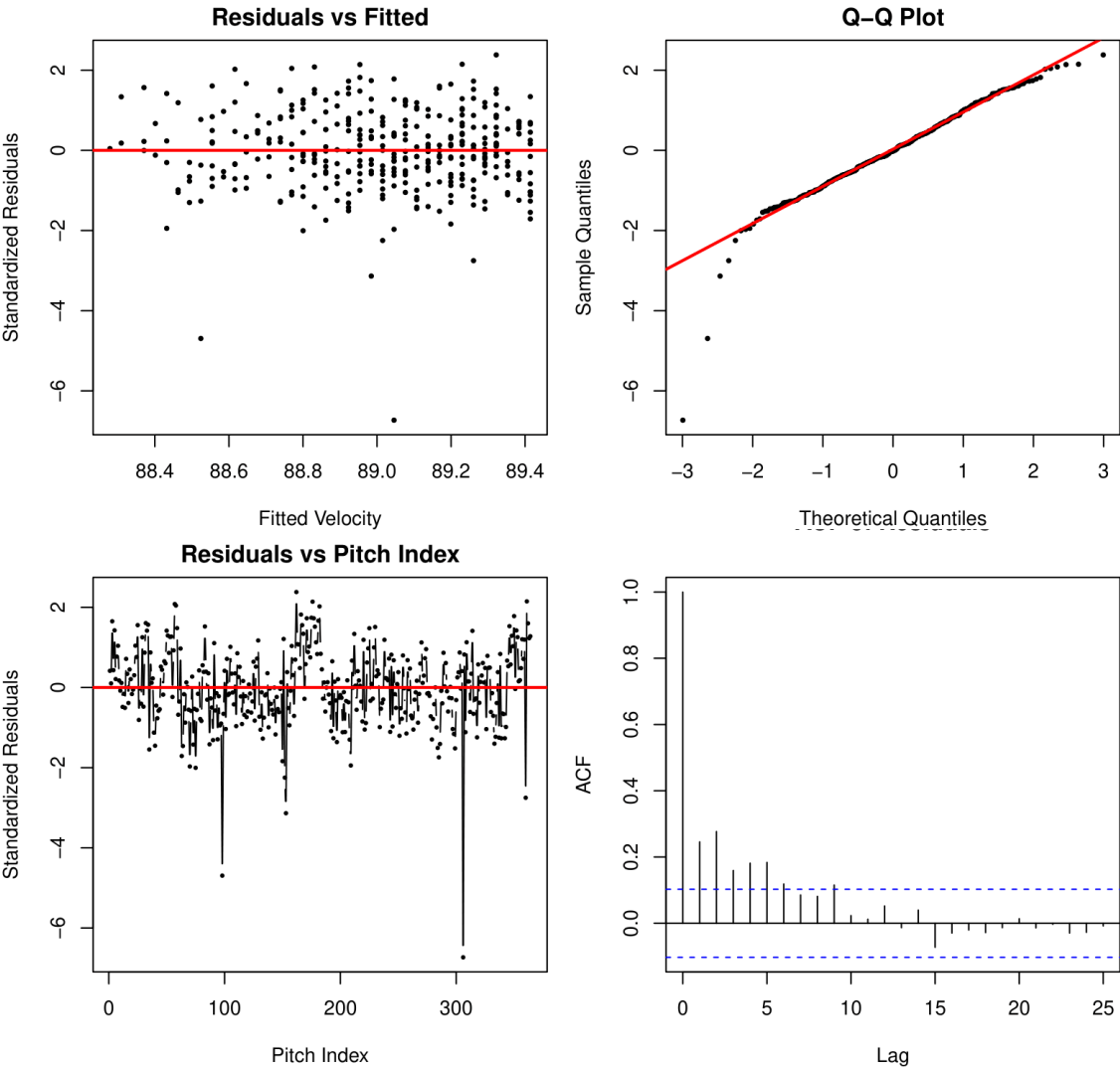


Player 18 | Velocity | Season 2024

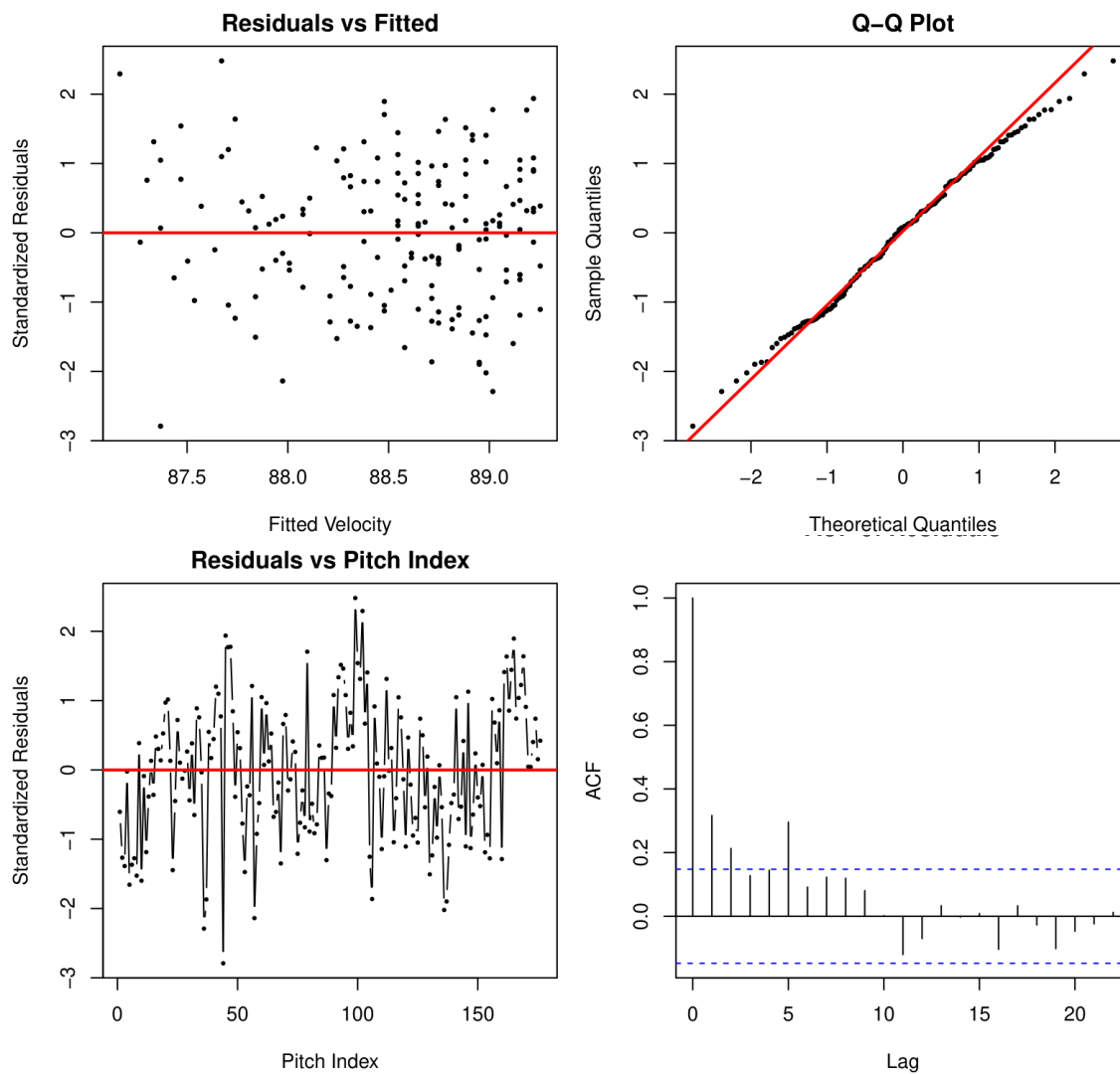


Player 19 | Velocity | Season 2024

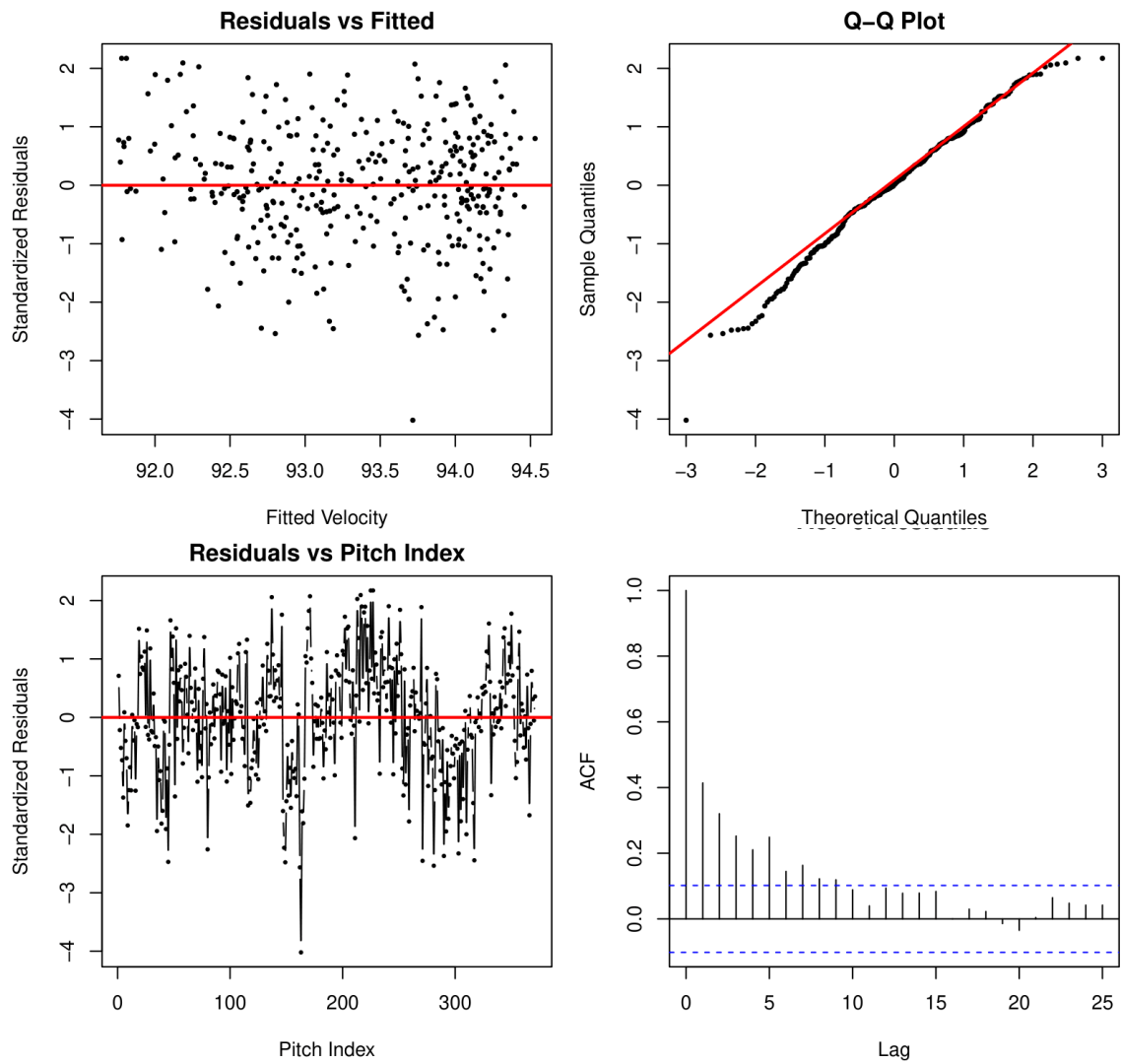




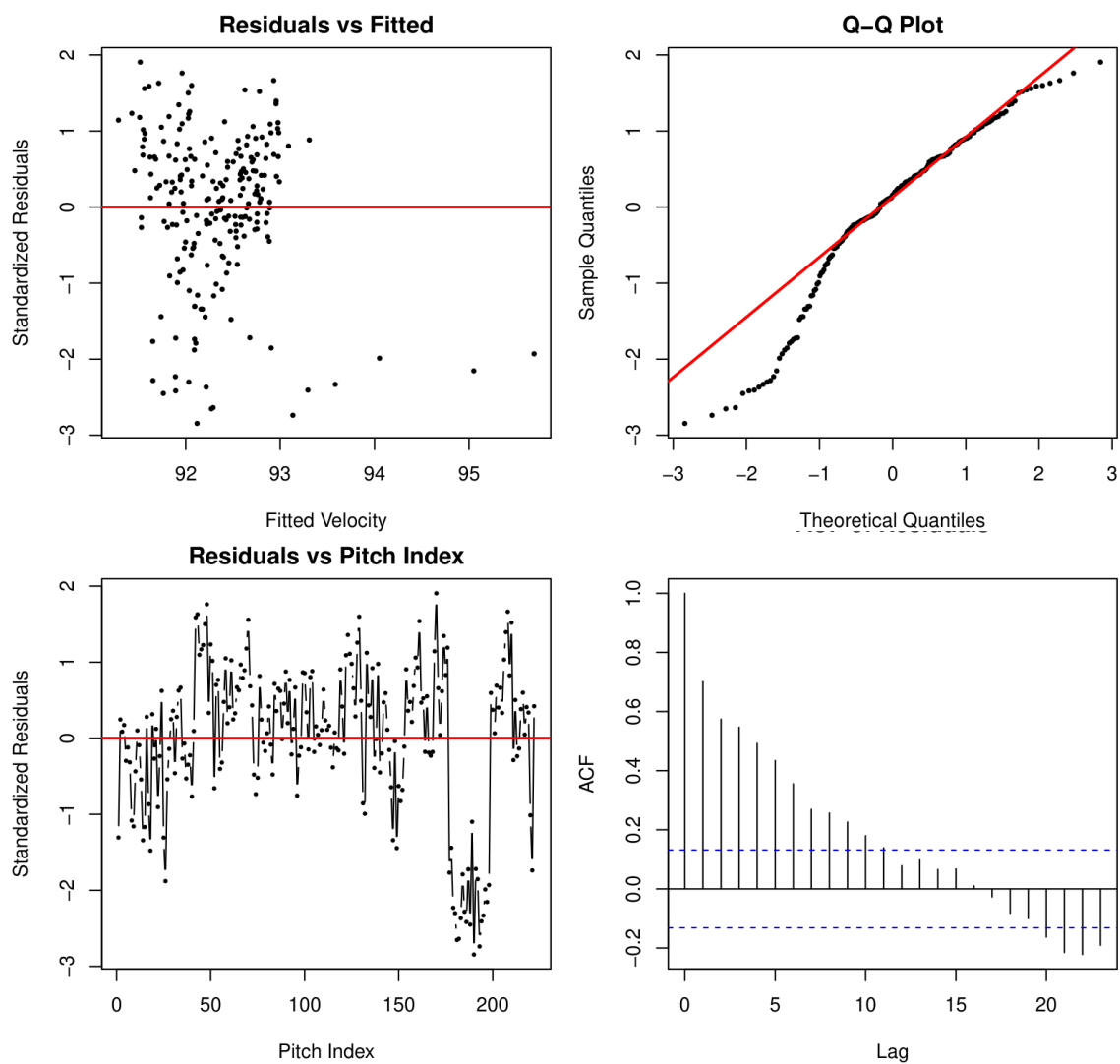
Player 21 | Velocity | Season 2024



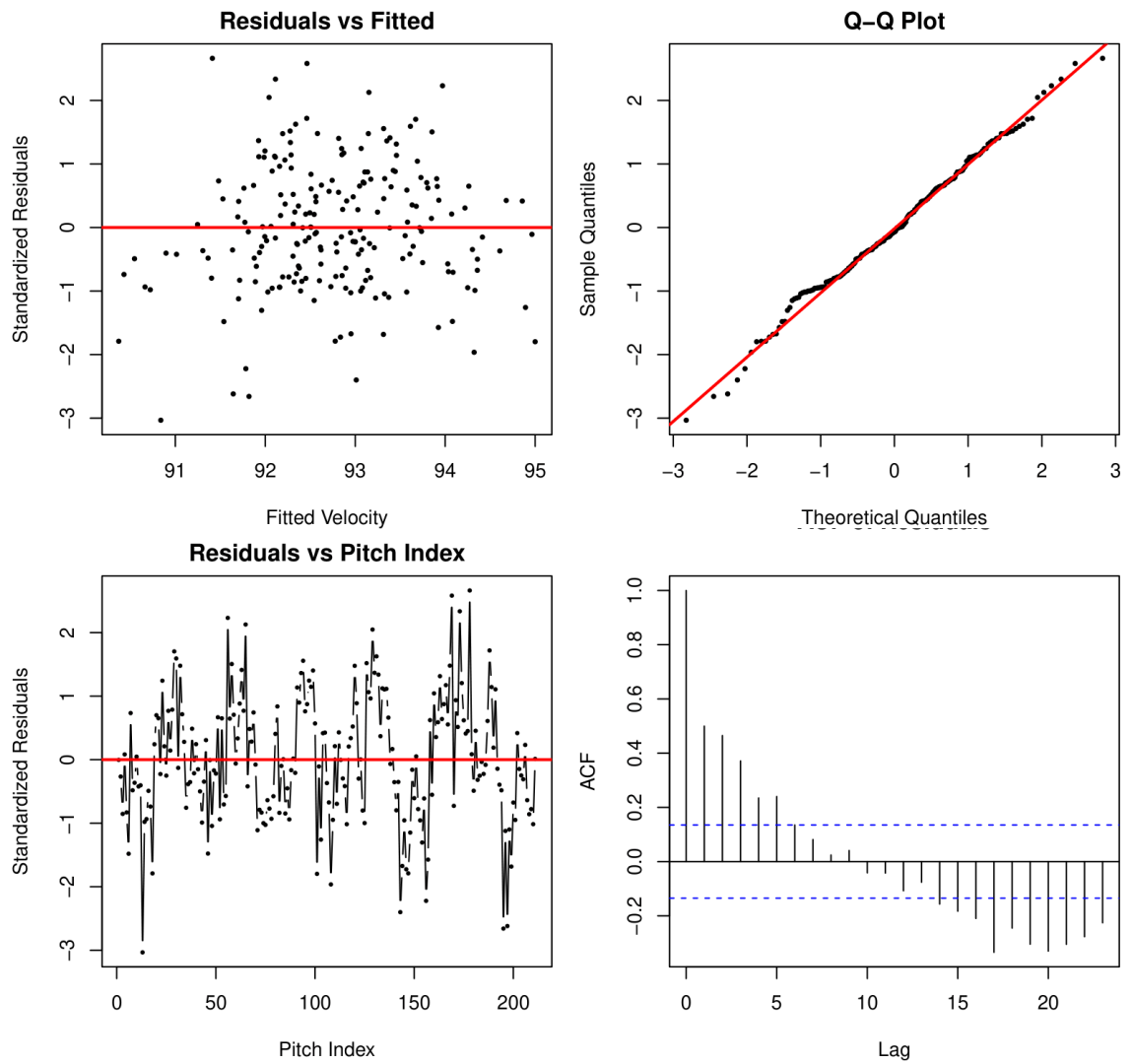
Player 22 | Velocity | Season 2025



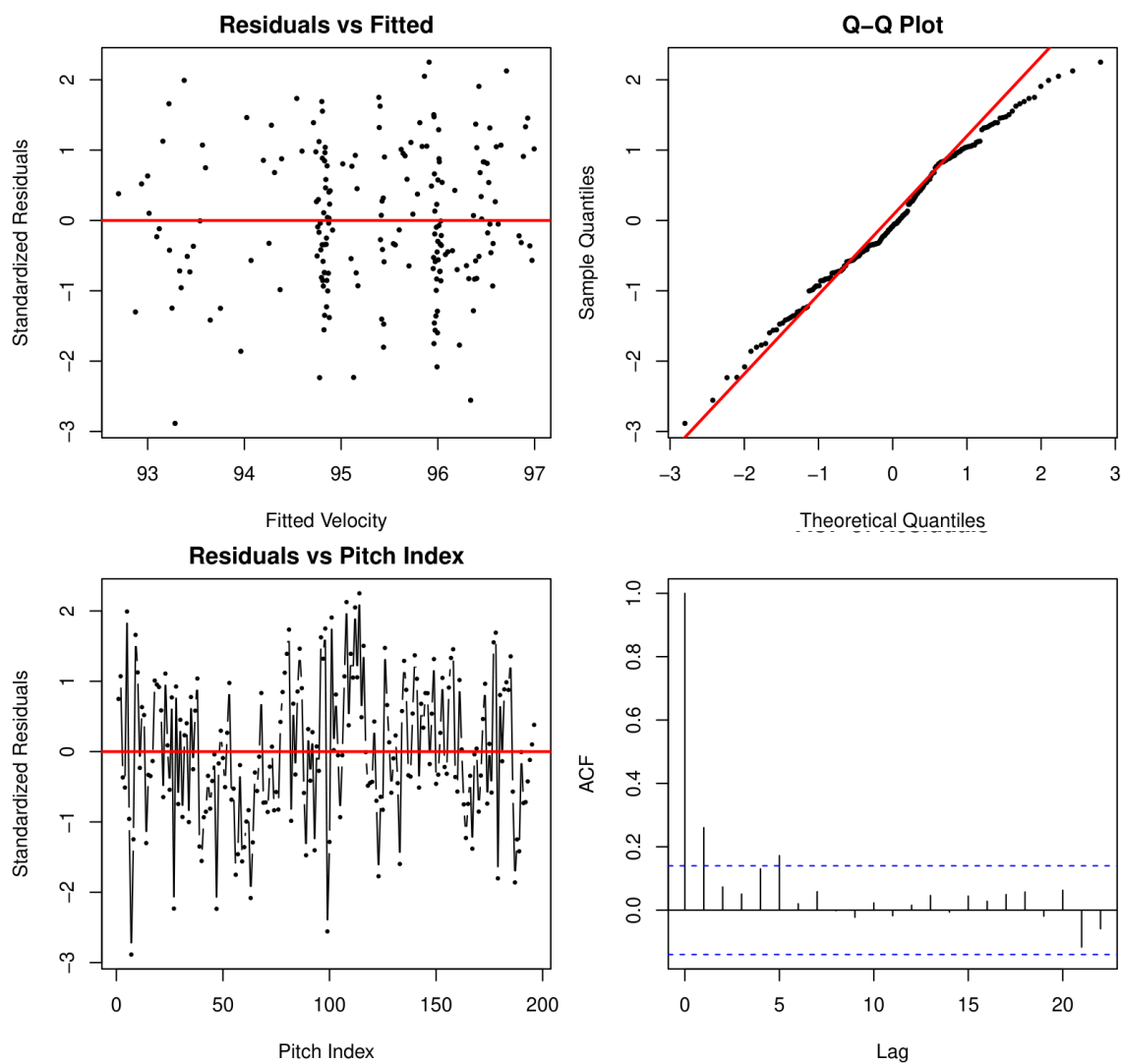
Player 23 | Velocity | Season 2025



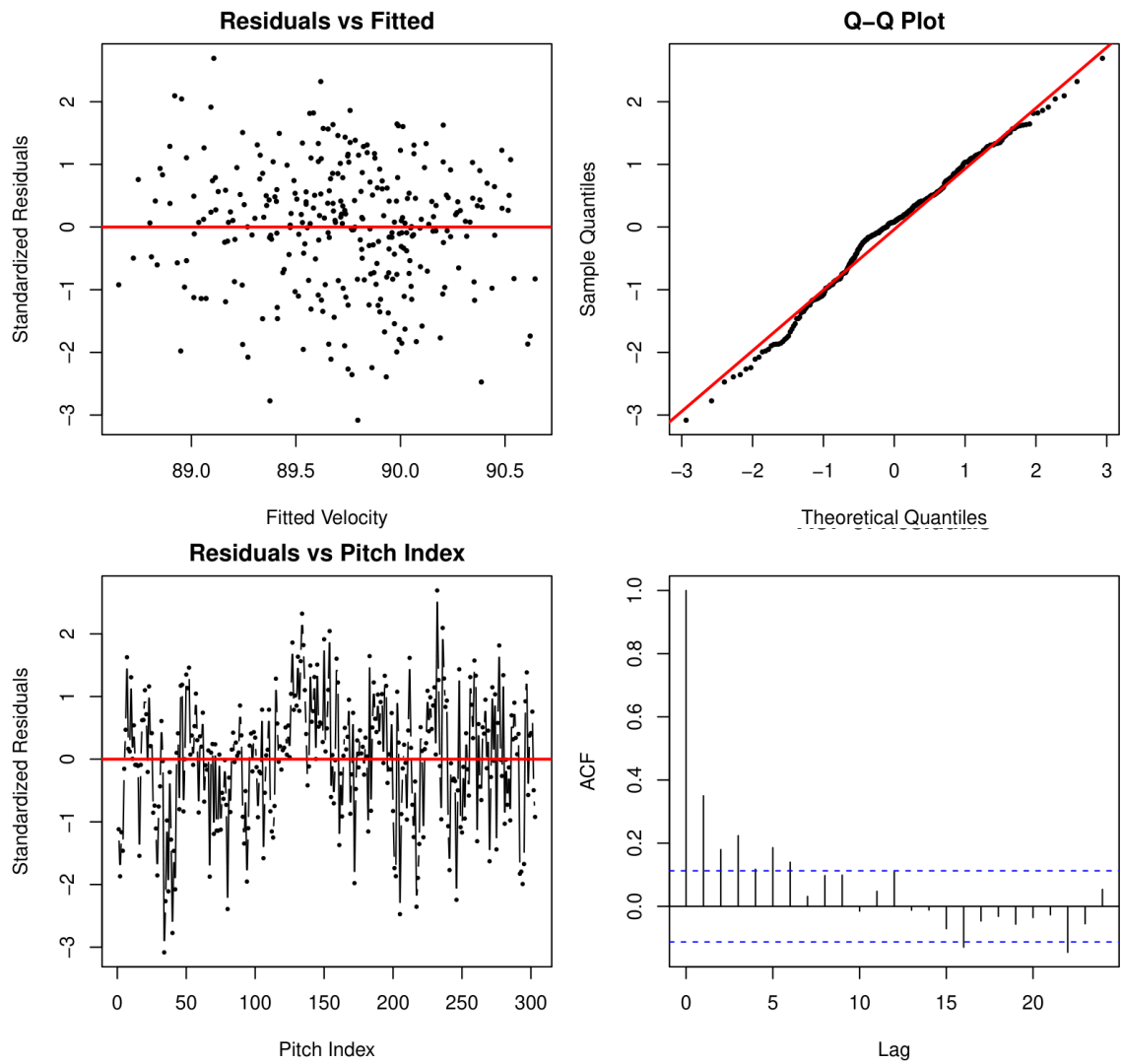
Player 24 | Velocity | Season 2025



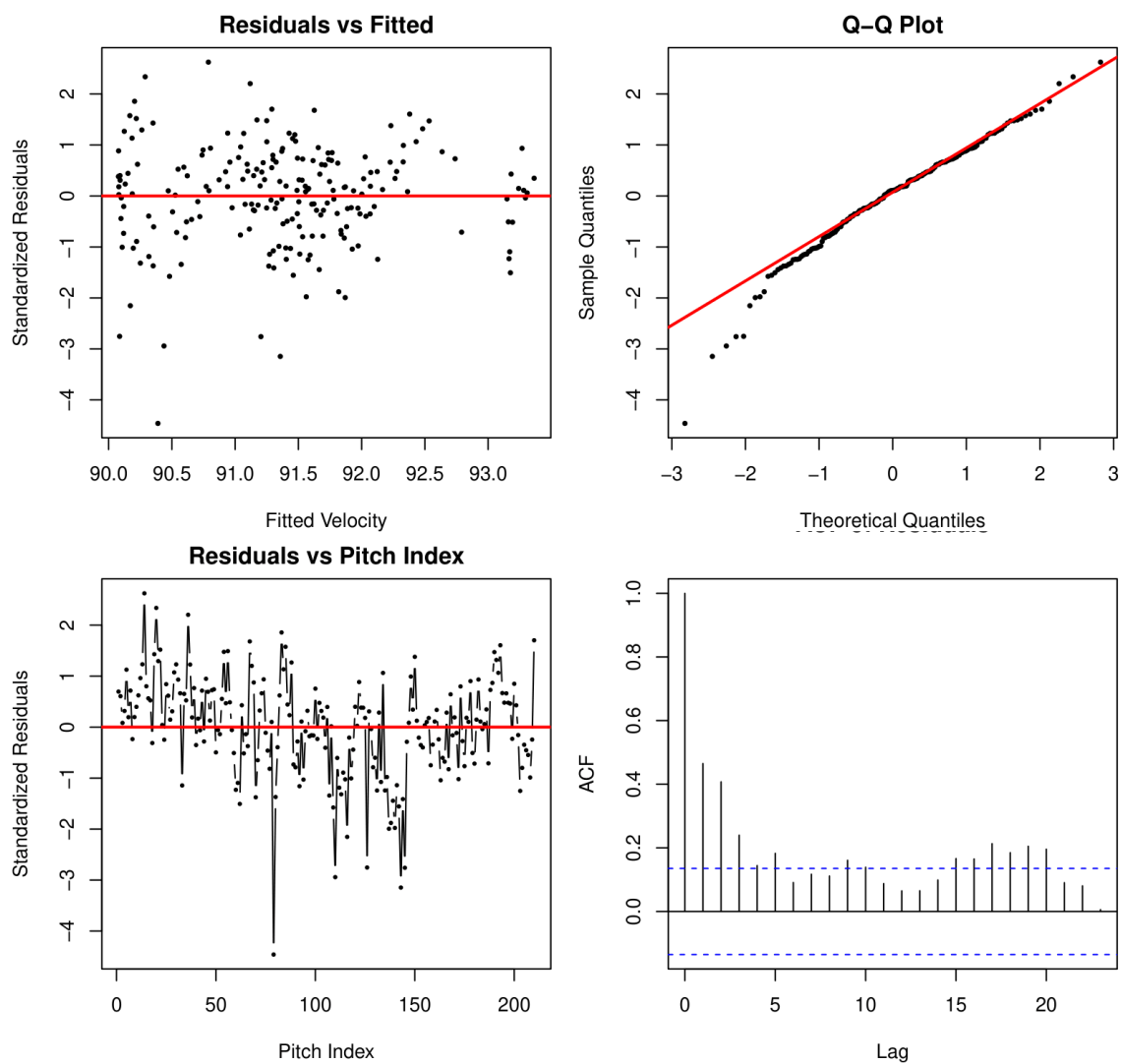
Player 25 | Velocity | Season 2025



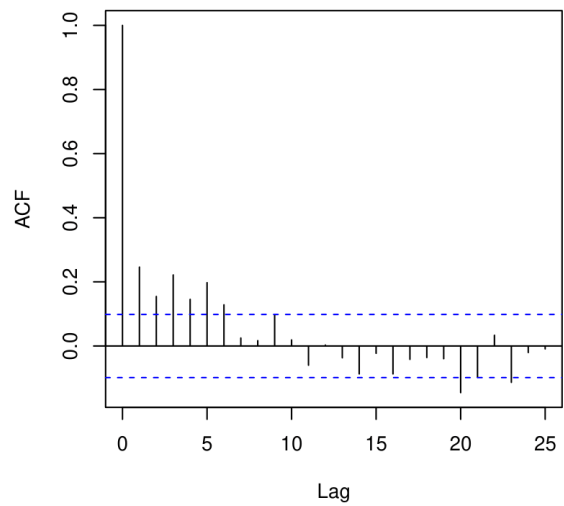
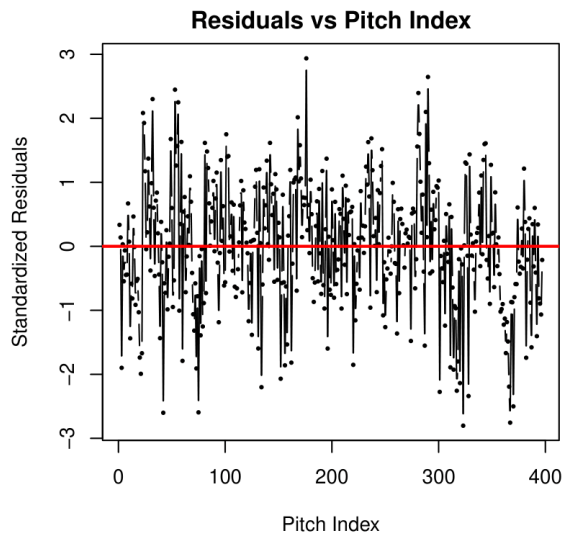
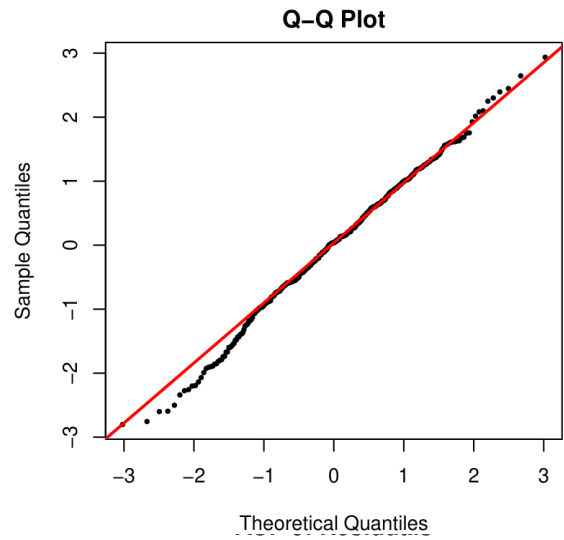
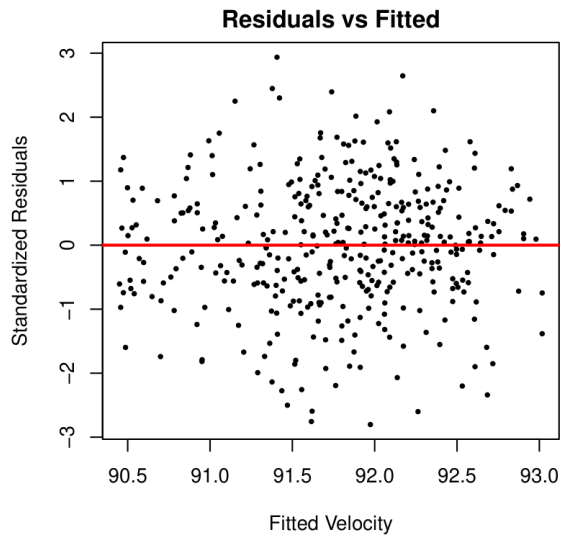
Player 26 | Velocity | Season 2025



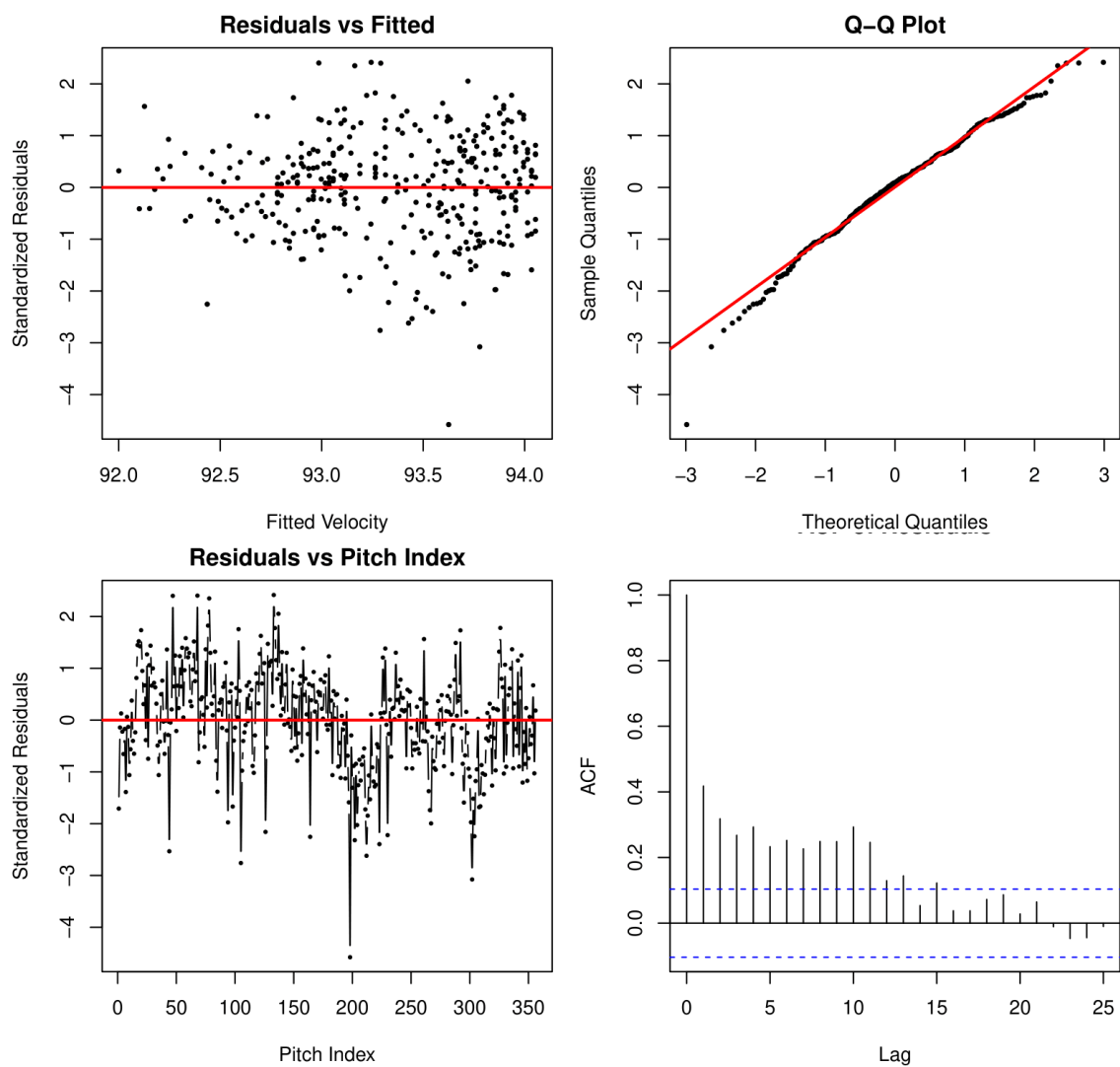
Player 27 | Velocity | Season 2025



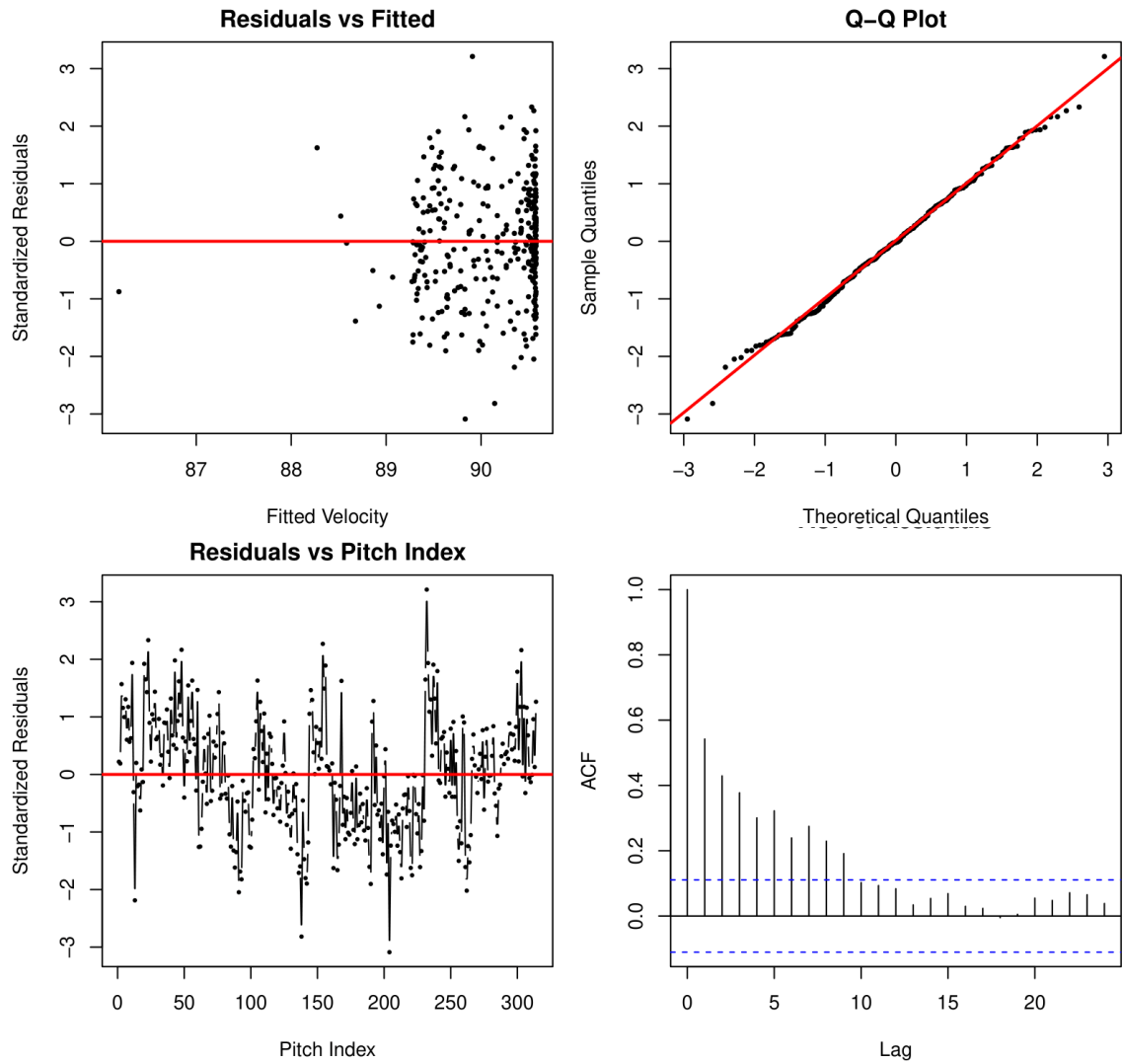
Player 28 | Velocity | Season 2025



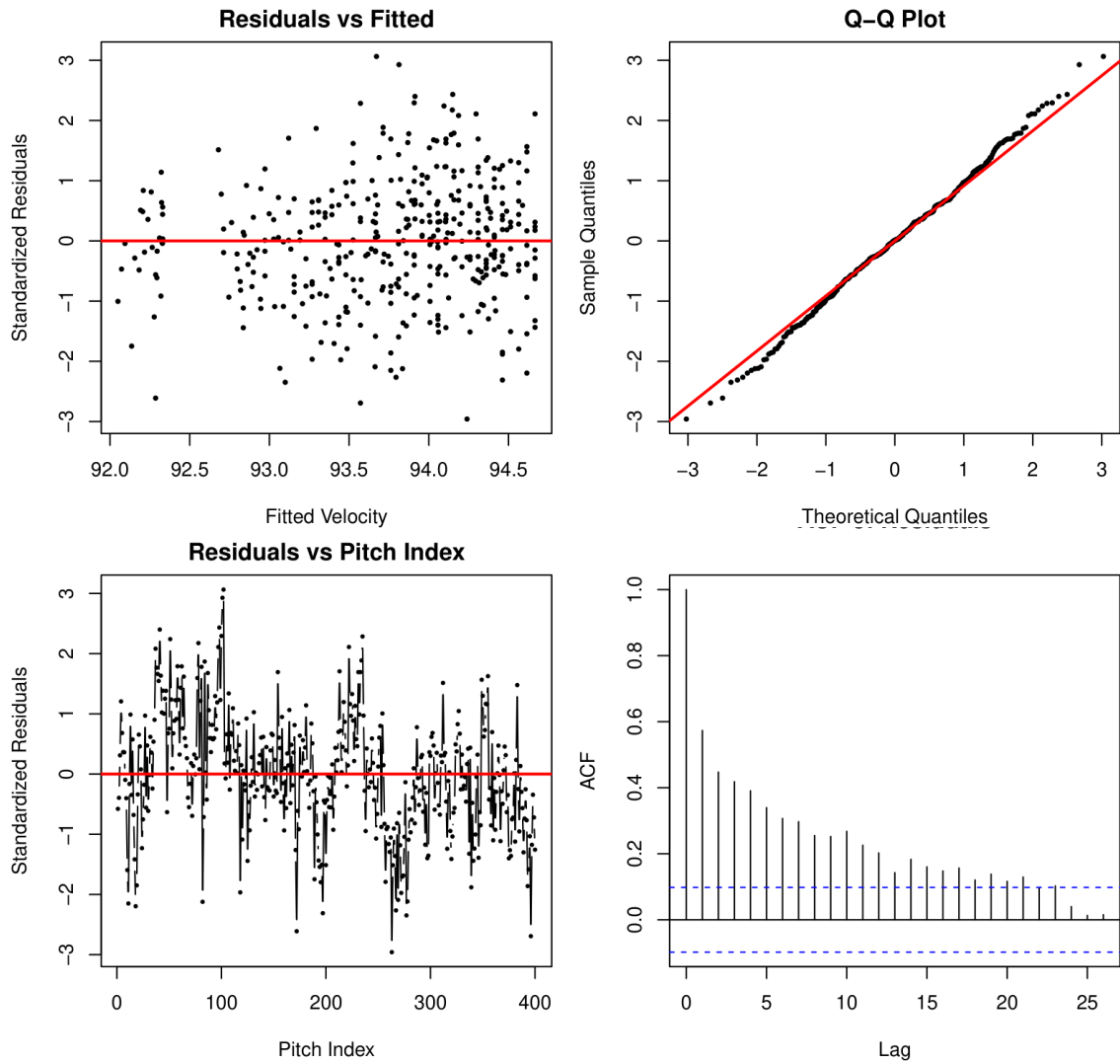
Player 29 | Velocity | Season 2025



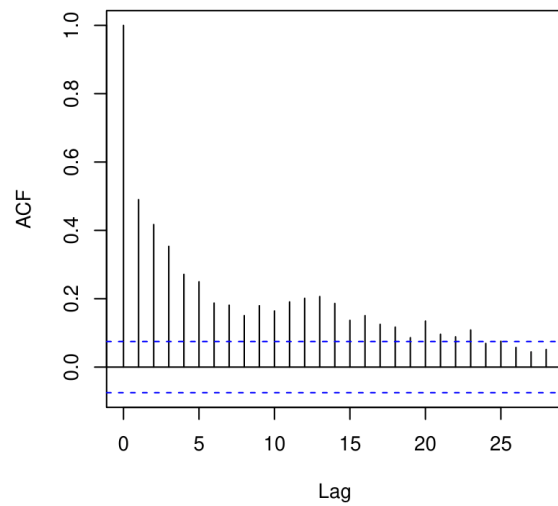
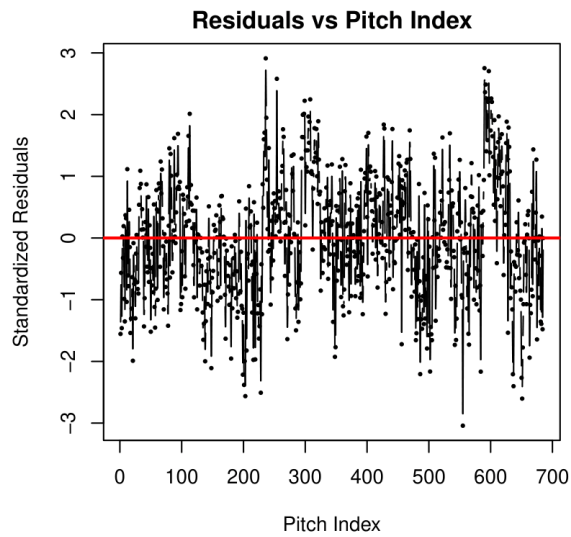
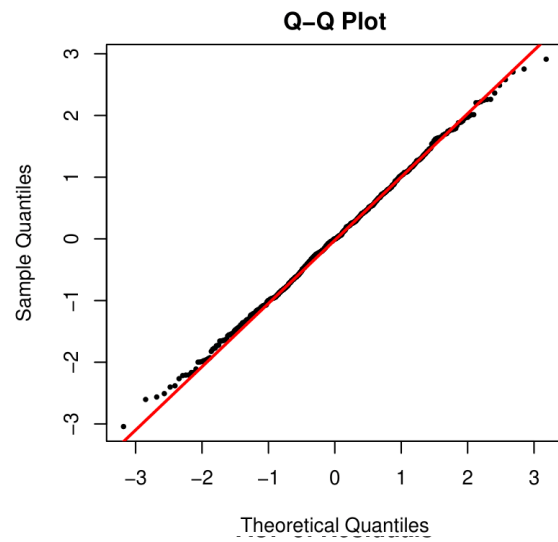
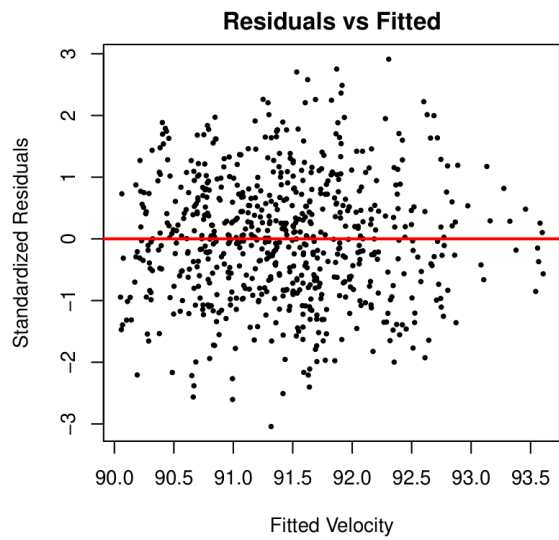
Player 30 | Velocity | Season 2025



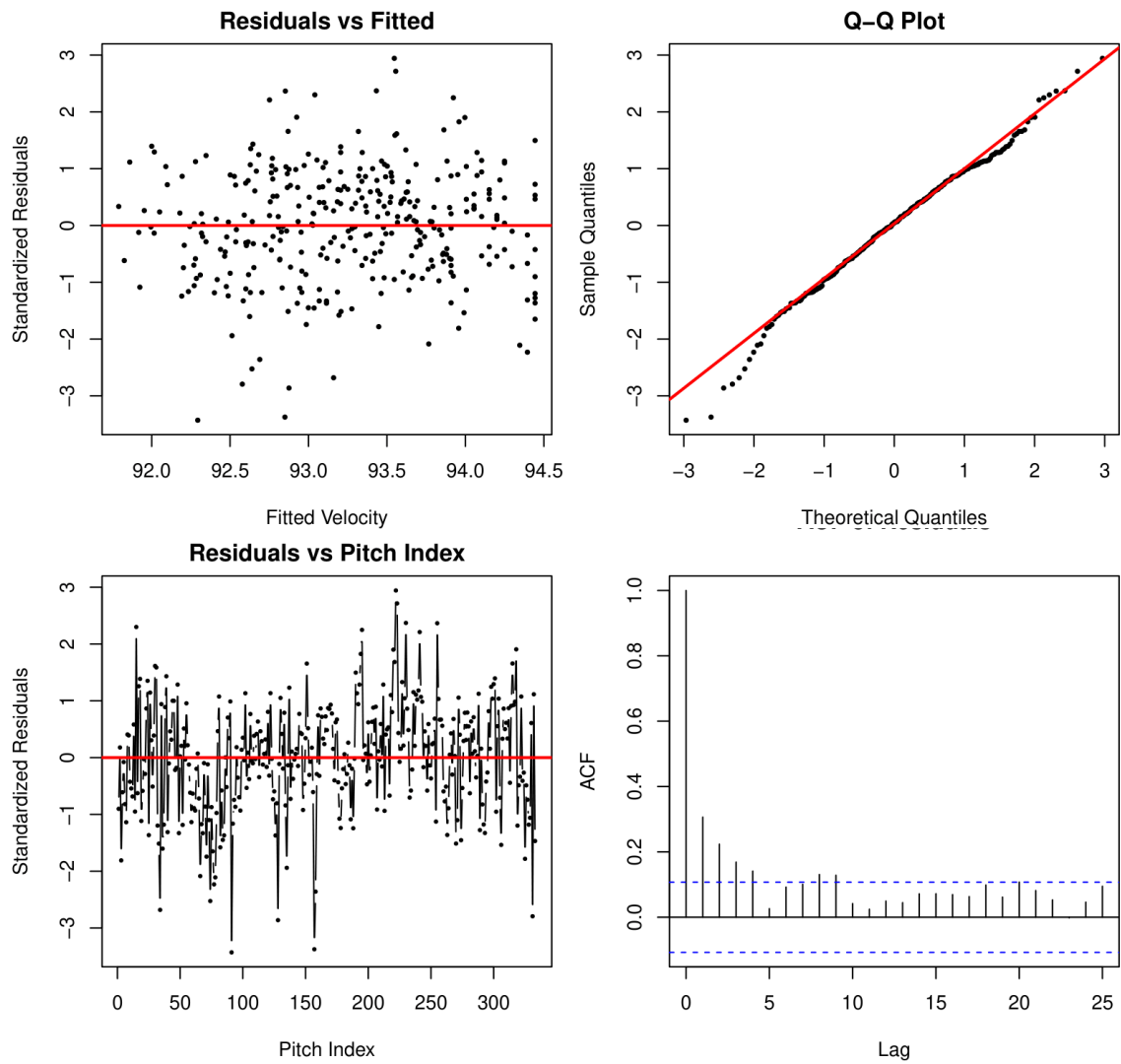
Player 31 | Velocity | Season 2025



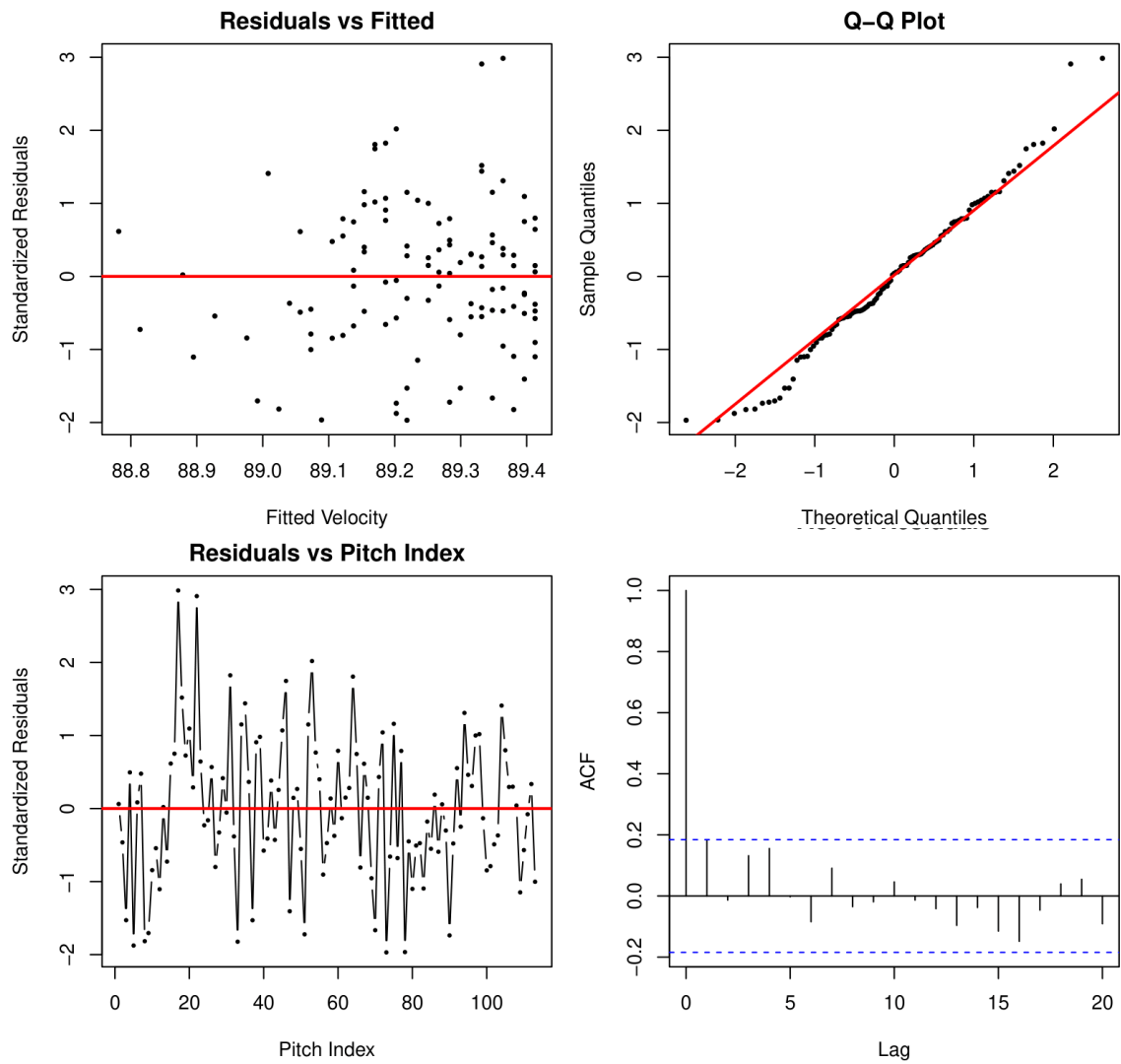
Player 32 | Velocity | Season 2025



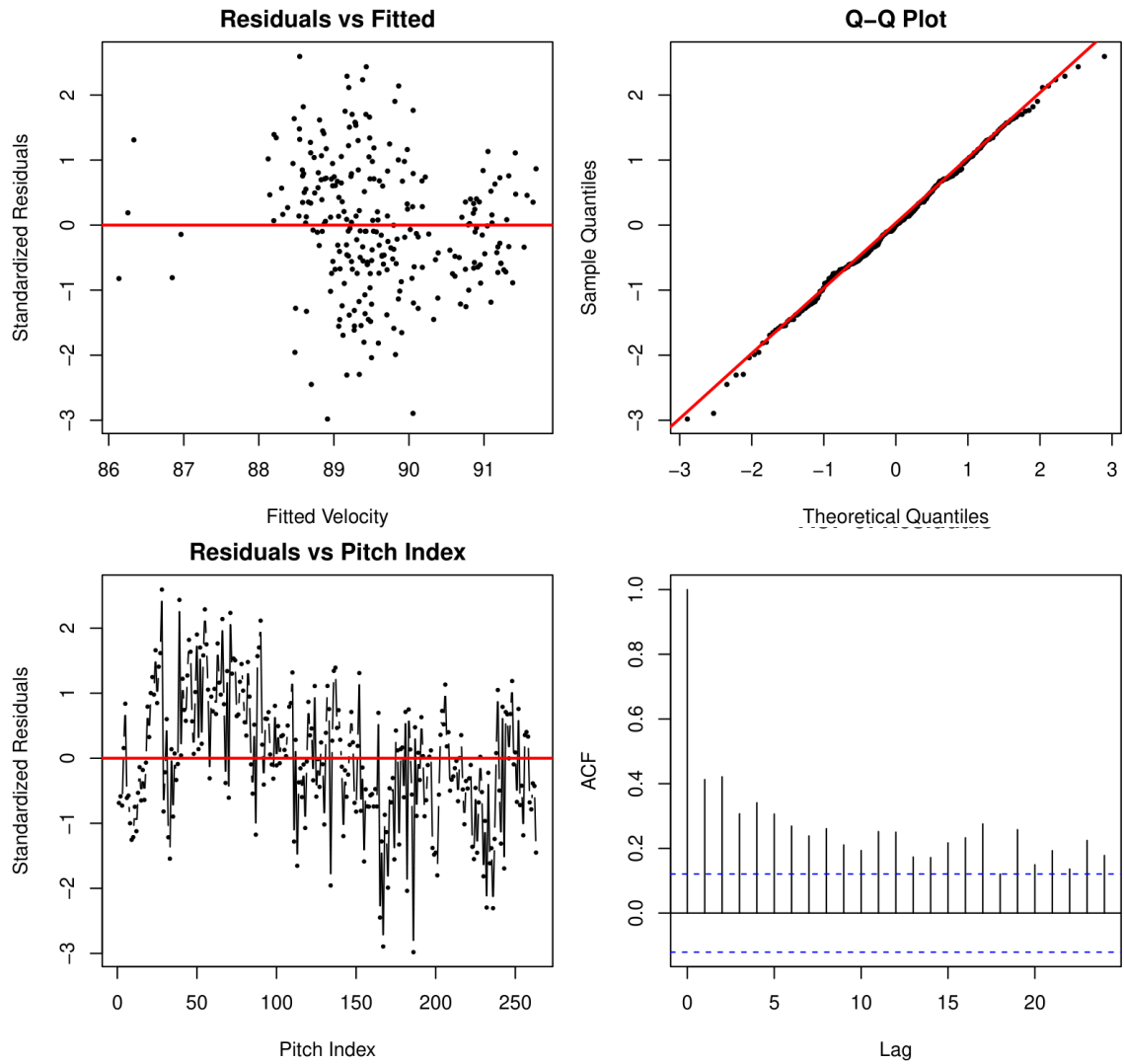
Player 33 | Velocity | Season 2025



Player 34 | Velocity | Season 2025



Player 35 | Velocity | Season 2025



CURRICULUM VITAE

Garrett Fernandez

Winston-Salem, North Carolina
(972) 302-7871 | garrett.fernandez@wfusm.edu | www.linkedin.com/in/garrett-fernandez-BME

Education

Wake Forest University School of Medicine, Winston-Salem, NC

Ph.D. in Biomedical Engineering (Biomechanics), Expected May 2026

Dissertation: *Applying N-of-1 GAMMs to Quantify Pitch Count Associations with Fastball Metrics and Biomechanics: A Comparative Modeling Study*

The University of Texas at San Antonio, San Antonio, TX

B.S. in Biomedical Engineering, Biomechanics Concentration; Minor in Wellness, May 2022

Honors: College of Engineering Dean's List (Fall 2019); Presidential Scholarship Recipient (3 years); UTSA Fall 2021 and Spring 2022 Senior Design Tech Symposium Winner

Research Experience

Applied Sport Science Researcher / Graduate Research Assistant

Wake Forest Pitching Lab, Wake Forest School of Medicine, Winston-Salem, NC
August 2022–April 2026

- Conduct applied biomechanics and workload-monitoring research in collegiate baseball pitchers.
- Collect, process, and analyze motion-capture, force, workload, and performance data using systems including Qualisys and KinaTrax.
- Develop individualized N-of-1 statistical modeling pipelines to evaluate pitcher-specific workload, performance, and biomechanical response patterns.
- Integrate multi-season athlete data into actionable reporting for coaches, medical staff, and performance personnel.
- Conduct motion-capture, strength, functional movement, and range-of-motion assessments.
- Collaborate with coaches, athletic trainers, physicians, and performance staff to support training and return-to-play decision-making.
- Develop data visualizations, presentations, and interactive tools, including Shiny applications, for applied sport science use.

Sport Science Intern

UFC Performance Institute, Las Vegas, NV
January 2024–April 2024

- Supported daily athlete evaluations and individualized performance monitoring.

- Worked across strength and conditioning, sports medicine, nutrition, and performance staff to integrate sport-science data into athlete-specific training discussions.
- Conducted physiological and energy-system testing, including VO2 max and maximal power output assessments.
- Developed and refined data visualizations and applied testing approaches in response to staff feedback and evolving performance questions.

Research Assistant

Musculoskeletal & Orthopedic Biomechanics Laboratory, University of Texas at San Antonio, San Antonio, TX
2020

- Assisted with cadaver-based biomechanics research under laboratory supervision.
- Organized cadaver specimens and supporting records for laboratory studies.
- Performed tissue preparation and shoulder dissections for musculoskeletal biomechanics projects.
- Assisted with ultrasound-based testing of rotator cuff stress and strain.
- Maintained laboratory equipment and supported general laboratory operations.

Publications

Wolf J, Nicholson KF, Fernandez G, Bullock GS, et al. **Examination of the Construct Validity of Nextiles Sleeve as a Measure of Peak Elbow Varus Torque.** *Journal of Athletic Training*. 2025.

Martin CL, Evenson KR, Moore JB, Bullock GS, Fernandez G, et al. **Using Biomechanical Devices in Elite Baseball Pitchers: A Preliminary Feasibility Study.** *International Journal of Sports Physical Therapy*. 2025.

Bullock GS, Ward P, Impellizzeri FM, Collins GS, Fernandez G, et al. **Up Front and Open? Shrouded in Secrecy? Or Somewhere in Between? A Meta-Research Systematic Review of Open Science Practices in Sport Medicine Research.** *Journal of Orthopaedic and Sports Physical Therapy*. 2023.

Manuscripts Submitted / Under Journal Consideration

Fernandez G, et al. **Biomechanical Drivers of Fastball Velocity Identified from In-Game Motion Capture Using Super Learner Ensemble Modeling.** Submitted to *Sports Biomechanics*.

Fernandez G, et al. **Evaluating Performance Metrics Before and After Return from Injuries in Baseball Pitchers: An Interrupted Time Series Longitudinal Study.** Submitted to *Journal of Sports Sciences*.

Manuscripts in Preparation

Fernandez G, et al. **N-of-1 GAMM Associations Between Rolling Pitch Count and Fastball Velocity, Spin Rate, and Vertical Break in Collegiate Baseball Pitchers.** Manuscript in preparation.

Fernandez G, et al. **Associations Between Rolling Pitch Count and Key Pitching Biomechanics in Collegiate Baseball Pitchers Using Individualized N-of-1 Modeling.** Manuscript in preparation.

Fernandez G, et al. **Comparison of Individualized N-of-1 GAMMs and Population-Level Mixed-Effects Models for Quantifying Rolling Pitch Count Associations in Collegiate Baseball Pitchers.** Manuscript in preparation.

Fernandez G, et al. **CUSUM-Based Monitoring of Individualized Expected Fastball Velocity for Detecting Systematic Performance Decline in Collegiate Baseball Pitchers.** Manuscript in preparation.

Conference and Research Presentations

Poster Presentation, SBES 2025. Presented work on individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers.

Poster Presentation, 2025 American Society of Biomechanics (ASB) Annual Meeting, Pittsburgh, Pennsylvania. August 13–16, 2025. Presented work on individualized associations between rolling pitch count and fastball performance metrics in collegiate baseball pitchers.

Poster Presentation, 2025 Biomedical Engineering Society (BMES) Annual Meeting, San Diego, California. October 8–12, 2025. Presented work on individualized associations between rolling pitch count and pitching biomechanics in collegiate baseball pitchers.

Grand Rounds Presentation, Orthopaedic Department, Atrium Health Wake Forest Baptist Medical Center. February 25, 2026.

Research Presentation, SBES, April 2026. Presented work on CUSUM-based monitoring for real-time detection of systematic performance decline in collegiate baseball pitchers.

Applying N-of-1 GAMMs to Quantify Pitch Count Associations with Fastball Metrics and Biomechanics: A Comparative Modeling Study. Dissertation Defense 2026

Technical Skills

- Statistical programming: R, MATLAB, SAS, Python
- Statistical methods: generalized additive mixed models, mixed-effects modeling, N-of-1 methods, LASSO regression, longitudinal modeling

- Data analysis and visualization
- Biomechanical analysis (marker-based and markerless motion capture)
- Motion tracking systems: Qualisys, KinaTrax
- Performance testing and athlete evaluations
- Workload monitoring and athlete management
- CAD modeling: SolidWorks
- Cell culture and maintenance
- Fluorescence microscopy
- Tissue preparation techniques
- Spirometry and ECG experience

Professional Memberships

- Biomedical Engineering Society (2017-2025)
- Society of Hispanic Professional Engineers (2019-2022)

Honors and Awards

- Presidential Scholarship Recipient, University of Texas at San Antonio
- College of Engineering Dean's List, Fall 2019
- UTSA Fall 2021 Senior Design Tech Symposium Winner
- UTSA Spring 2022 Senior Design Tech Symposium Winner
- 2nd Place Oral Presentation Award – Data Analytics for Risk Prevention SBES 2026