A Statistical Look at Roger Clemens' Pitching Career

By

Eric T. Bradlow
Shane T. Jensen
Justin Wolfers
Abraham (Adi) J. Wyner*

* Eric T. Bradlow is the K.P. Chao Professor, Professor of Marketing Statistics and Education and Academic Director of the Wharton Small Business Development Center, Shane T. Jensen (stjensen@wharton.upenn.edu) is an Assistant Professor of Statistics, Justin Wolfers (jwolfers@wharton.upenn.edu) is an Assistant Professor of Business and Public Policy, and Adi J. Wyner (ajw@wharton.upenn.edu) is an Associate Professor of Statistics, all at the Wharton School of the University of Pennsylvania. All correspondence on this manuscript should be sent to Eric T. Bradlow, 761 Jon M. Huntsman Hall, The Wharton School, Philadelphia PA 19104 (ebradlow@wharton.upenn.edu, 215-898-8255). The order of authorship is alphabetical and all authors contributed equally on this manuscript.

Abstract

A recent report (Hendricks Sports Management, LP, et al, 2008) issued by Hendricks Sports Management, LP, claims to provide evidence for the lack of use of performance-enhancing substances (PESs) by Hall-of-Fame caliber pitcher Roger Clemens, a claim based on an analysis of his career statistics (using ERA = earned run average, K rate = strikeout rate, innings pitched), both in isolation and in comparison to other power pitchers of his era (Randy Johnson, Nolan Ryan, and Curt Schilling).

In this research, we re-examine Roger Clemens' career using a more complete and stable set of pitching measures (WHIP = walks + hits per inning pitched, BAA = batting average against, ERA, BB rate = rate of walks per batter faced, K rate), and by using a broader (census) comparison set of pitchers with similar longevity in order to reduce the selection bias inherent in the Hendricks report. In contrast to Hendricks' report, our analysis examines not only late career performance but also early- and mid-career trends. Our findings can be summarized as follows:

Using simple quadratic functions, and an occasional spline to relate the above pitching measures to age, we demonstrate a number of empirical regularities:

- Roger Clemens' career is atypical with respect to his peer group. While most pitchers with comparable longevity improve for the first half of their career, peaking just past the age of 30 and then declining (an inverted-U shape), Roger Clemens' career statistics shows a decrease into his early thirties followed by a marked improvement late in his career (more of a U-shape).

- This pattern is consistent across most measures for Roger Clemens, yet for certain measures is not unique to him. That is, other pitchers have atypical patterns as well for some, but not all other tested measures.

Our analyses suggest what we, as statisticians, have postulated all along: empirical association is not causation, and neither the Hendricks report nor ours can prove or disprove the use of PESs by any given player. This is because players are indeed unique, and due to the short-time series and sparseness of comparable players there is low power to assess specific hypotheses. However, our analyses clearly suggest that Roger Clemens' career pitching trajectory is atypical.

1. Introduction

Baseball is 'America's Pastime' and with attendance and interest at an all-time high, it is clear that baseball is a big business. Furthermore, many of the sport's hallowed records (the yearly home run record, the total home run record, the 500 home run club, etc.) are being assailed and passed at a pace never before seen. Yet, due to the admitted use and accusations documented in the "Mitchell report" (2007), of performance enhancing substances (PES's), the "shadow" (Fainaru-Wada and Williams, 2007) over these accomplishments is receiving as much press, if not more, than the breaking of the records themselves.

A particularly salient example comes from a recently released report by Hendricks Sports Management, LP (Hendricks Sports Management, LP, et al) which led to widespread national coverage. This report purported to demonstrate the innocence from claims of PES use directed at Roger Clemens, one of baseball's all-time highest-performing pitchers. Using well-established baseball statistics including ERA (number of earned runs allowed per nine innings pitched) and K-rate (strikeout rate per nine-innings pitched), the report compares Roger Clemens' career to those of other great power pitchers of his era (Randy Johnson, Nolan Ryan, and Curt Schilling) and proclaims that Roger Clemens' career *trajectory* on these measures is not atypical. Based on this finding, the report suggests that the pitching data themselves are not an indictment (nor does it provide proof) of Clemens guilt; in fact, just the opposite.

While we concur with the Hendricks report that a statistical analysis of Clemens' career can provide prima facie 'evidence'[1], our approach provides a new look at his career pitching trajectory using a broader set of measures as well as a broader comparison set of pitchers. This is important as there has been a lot of recent research as to what are the most reliable and stable measures of pitching performance (Albert, 2006) and our attempt is to be inclusive in this regard. In Section 2, we provide a closer examination of Clemens' career with respect to these additional pitching measures.

Even more importantly, one of the pitfalls that all analyses of extraordinary events (the immense success of Clemens as a pitcher) have is 'right-tail self-selection'. If one compares extraordinary players only to other extraordinary players, and selects that set of comparison players based on their behavior on that extraordinary dimension, then one does not obtain a representative (appropriate) comparison set. By focusing only on pitchers who pitched effectively into their mid-forties, the Hendricks report minimized the possibility that Clemens would look atypical. Here we use more reasonable criteria for pitchers that are based on their longevity and the number of innings pitched in their career to form the comparison set, rather than performance at any specific point in their career.

The focus of this paper is an analysis of Clemens' career using a more sophisticated and comprehensive database which we describe in Section 3. Section 4 contains our analysis of this larger comparison set, which suggests that Roger Clemens'

[1] It is important to point out that neither the Hendricks report nor this one can prove or disprove the guilt or innocence of any player based on data alone. Rather, statistics can provide a lens through which we can compare a focal player to other comparable sets of players.

performance is very atypical along these dimensions. We conclude in Section 5 with a set of prescriptive advice for those who wish to perform similar analyses.

2. A Closer Look at the Career of Roger Clemens

Before we begin our full analysis and discussion, we first take a closer look at Clemens' entire career. To be sure, this unavoidable act of data "snooping" was part of our research method, and it is instructive to unfold our insights in the order in which they actually occurred. For the average fan, the most salient measures of success are winning percentage and ERA, which are a good place to start.

Figure 1a: Clemens' Winning Percentage over Time

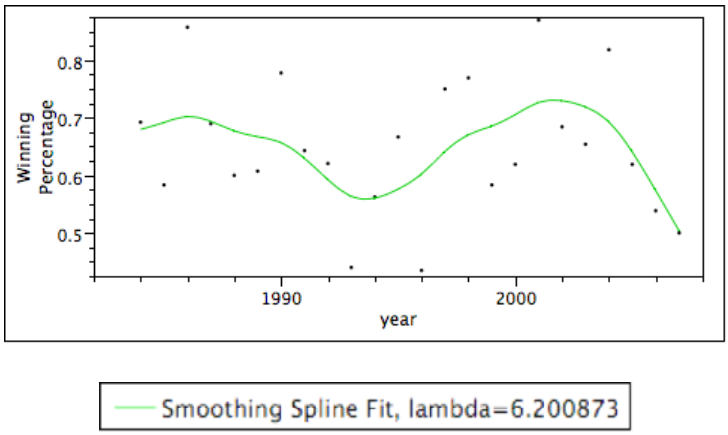
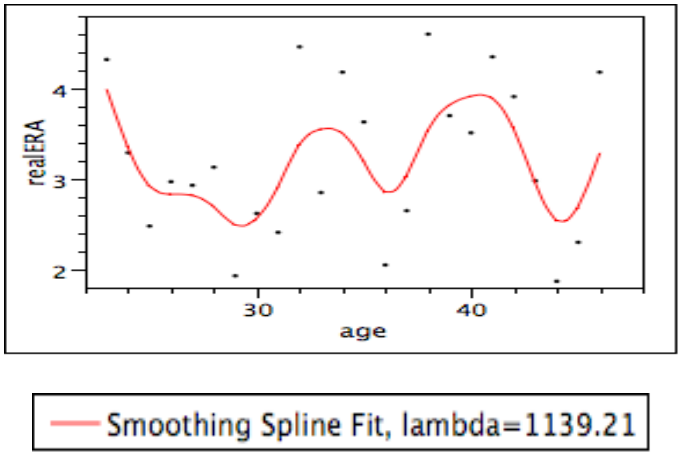


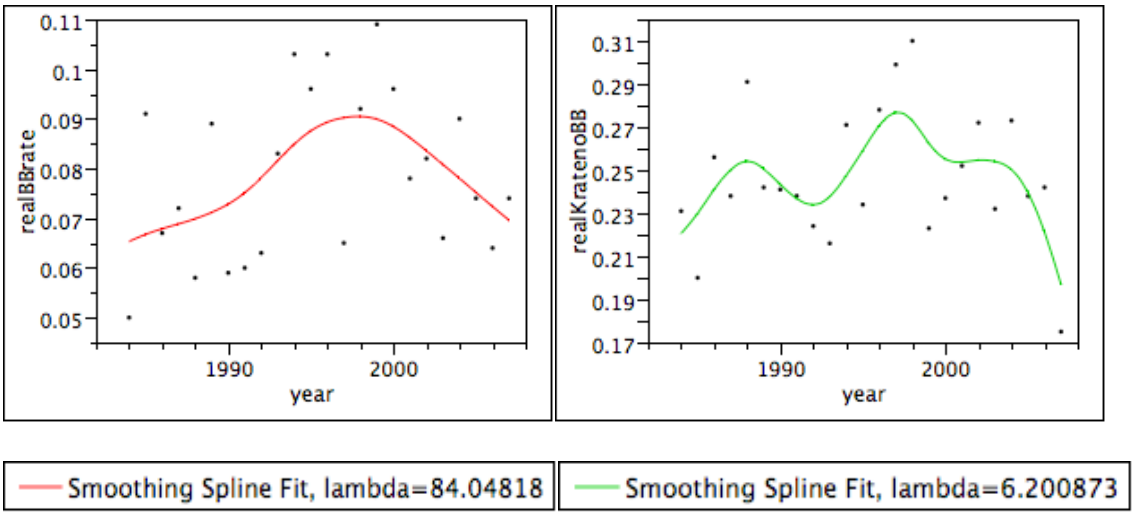
Figure 1b: Clemens' ERA over Time



What these graphs show is that Clemens quickly established himself as a star and in the early 1990s he lost his "relative" luster. His final 4 years with the Red Sox were certifiably mediocre (compared to his history), so much so that the future Hall of Famer was considered to be in the "twilight of his career" (Dan Duquette quoted in Silverman, 1996). However, as our graph clearly demonstrates, Clemens recovered and climbed to new heights at the comparatively old age of 35. His last few years showed a second period of decline.

Now any 'well-read' student of baseball understands that winning percentage and ERA (Earned Run Average) are fairly noisy measures of quality. Both measures are readily affected by factors outside a pitcher's ability, such as fielding and the order in which batting events occur. Additionally, winning percentage depends critically on run support. Analysts who specialize in pitching evaluation, instead use measures of component events[2], such as rates of Strike Outs (K) and Walks (BB). We graph the career trajectory of K rate and BB rate for Roger Clemens below:

Figure 2: Clemens' K rate and BB rate over Time



[2] For example, see http://en.wikipedia.org/wiki/Baseball_statistics#Pitching_statistics.

Again we see Clemens' strong start, a gradual decline in BB rate as he entered the "first-twilight" of his career, followed by a marked improvement. His strikeout rate improved in his early career and then suddenly declined and then rose again peaking at age of 35 in 1998 in his second year with Toronto.

To put these career trajectories in an appropriate context, we require a comparison group. Our first effort (but not our last) was a handful of star-level contemporaries, including Greg Maddux, Randy Johnson and Curt Schilling, whose career trajectories on these statistics are graphed below:

Figure 3: Greg Maddux K rate and BB rate over Time

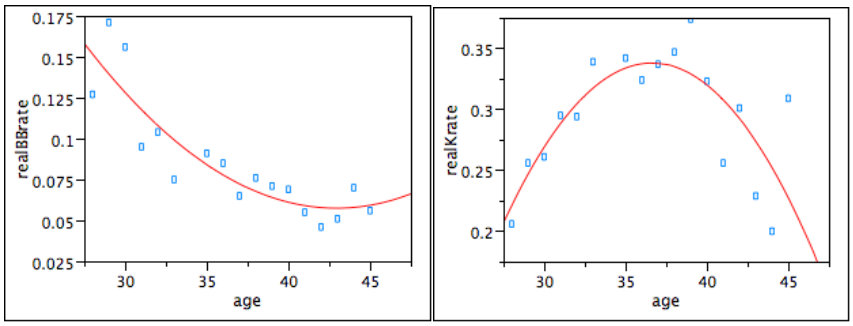


Figure 4: Randy Johnson K rate and BB rate over Time

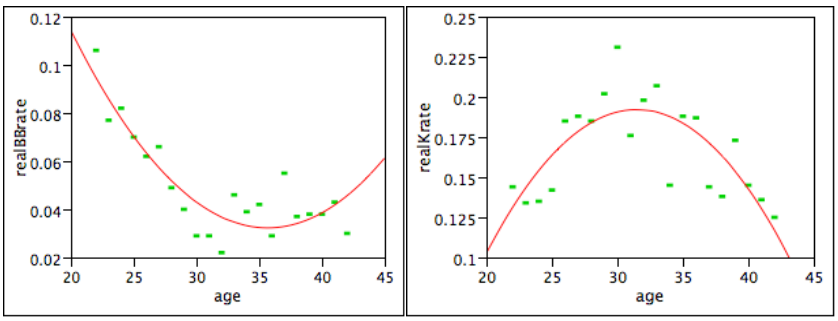
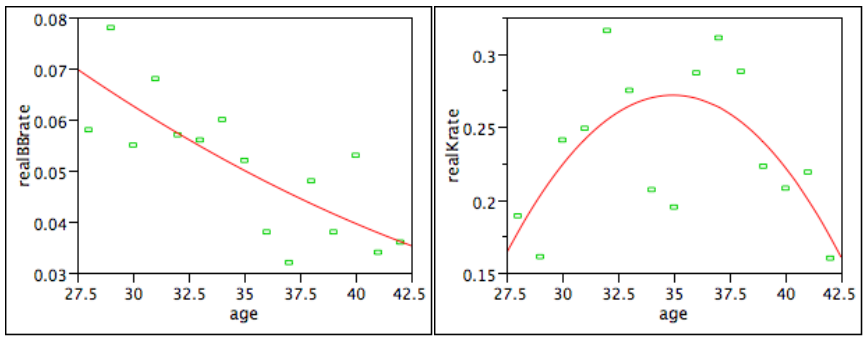


Figure 5: Curt Schilling K rate and BB rate over Time



The career trajectories for all three of these star contemporaries of Clemens are nicely fit with quadratic curves. In terms of performance, the curves clearly show steady improvement as they entered their primes followed by a marked decline in their strikeout rate and a leveling off in their walk rates.

The contrast with Clemens' career trajectory is quite stark. The second act for Clemens is unusual when compared to these other greats because his later success follows such an unprecedented period of decline. This leaves us with the following question: how unusual is it for a durable pitcher to have suffered a mid-career decline and then recovery?

3. Database Construction

In order to perform our statistical analyses, we first obtained data from the Lahman Database, Version 5.5 (www.baseball1.com) on all major league baseball pitchers, and their associated statistics, whose careers were contained in the years 1969-2007. The starting year of 1969 was selected because of the change in the height of the pitchers' mound (Vescey, 1968), which launched the 'modern era' in baseball. From that set of pitchers, we constructed a comparison set of all durable starting pitchers by looking at all pitchers who played at least 15 years (with 10 or more games started per year) and had at least 3000 innings pitched.[3] There were thirty-one pitchers besides Roger Clemens that fit these criteria. All of these starting pitchers therefore had comparably long careers (in years) and innings pitched similar to Roger Clemens, and hence were a relevant comparison set; albeit others could certainly be chosen. Appendix 1 gives the names and a set of descriptive statistics for the 31 players and Clemens.

For each of these pitchers, we looked at the following well-established pitching statistics for each of the years in which they pitched:

[1] WHIP = Walks + Hits per inning pitched;

[2] BAA = Batting average for hitters when facing the given pitcher;

[3] ERA = Earned run average per nine innings pitched;

[4] BB Rate = Walk rate;

[3] Sensitivity analyses run that included minor perturbations in these criterion indicated that the results are quite stable.

[5] K rate = Batter strike out rate per plate appearance (not including walks)

Together, these statistics provide a fairly complete picture of the career trajectory for a starting pitcher. We describe next in Section 4 a set of analyses we performed on the data.

4. Trajectory Analyses

In order to understand and summarize the trajectory that each of the five (j = 1, …J=5) aforementioned statistics take, for each of the thirty-two (i =1, .., I = 32) focal pitchers (including Clemens), we fit a quadratic function to each pitcher's data at year t as follows:

$$S_{ijt} = \beta_{0ij} + \beta_{1ij} Age_{it} + \beta_{2ij} Age_{it}^2 + \varepsilon_{ijt} \qquad [1]$$

where S_{ijt} = value of statistic j for pitcher i in their t-th season, Age_{it} = age of pitcher i in their t-th major league season, β_{0ij}, β_{1ij}, and β_{2ij} are an intercept and coefficients describing how Age and Age^2 influence the prediction of the statistics, and ε_{ijt} is a randomly distributed normal error term.[4] We acknowledge that a quadratic curve may not be the best model for every pitcher's career, including Roger Clemens. However, the

[4] As none of these statistics were near boundaries, taking transformations to make the normal residual error more plausible had little impact.

quadratic curve is a simple model with interpretable coefficients that provide a common basis of comparison for all pitchers in our study.

Our primary interest centers around the coefficient β_{2ij} which describes whether the pitchers trajectory for that statistic is purely linear as they age ($\beta_{2ij} = 0$), "hump-shaped" ($\beta_{2ij} < 0$) or "U-shaped" ($\beta_{2ij} > 0$). To provide some context around this, one might predict the following patterns, corresponding a priori to a pitcher hitting a mid-career "prime" and then falling off near the end of his career.

[1] WHIP ($\beta_{2ij} > 0$ and career peak = $-\beta_{1ij}/\beta_{2ij} \approx 31$ [5]).

[2] BAA ($\beta_{2ij} > 0$ and career peak = $-\beta_{1ij}/\beta_{2ij} \approx 29$)

[3] ERA ($\beta_{2ij} > 0$ and career peak = $-\beta_{1ij}/\beta_{2ij} \approx 29$)

[4] BB Rate ($\beta_{2ij} > 0$ and career peak = $-\beta_{1ij}/\beta_{2ij} \approx 33$)

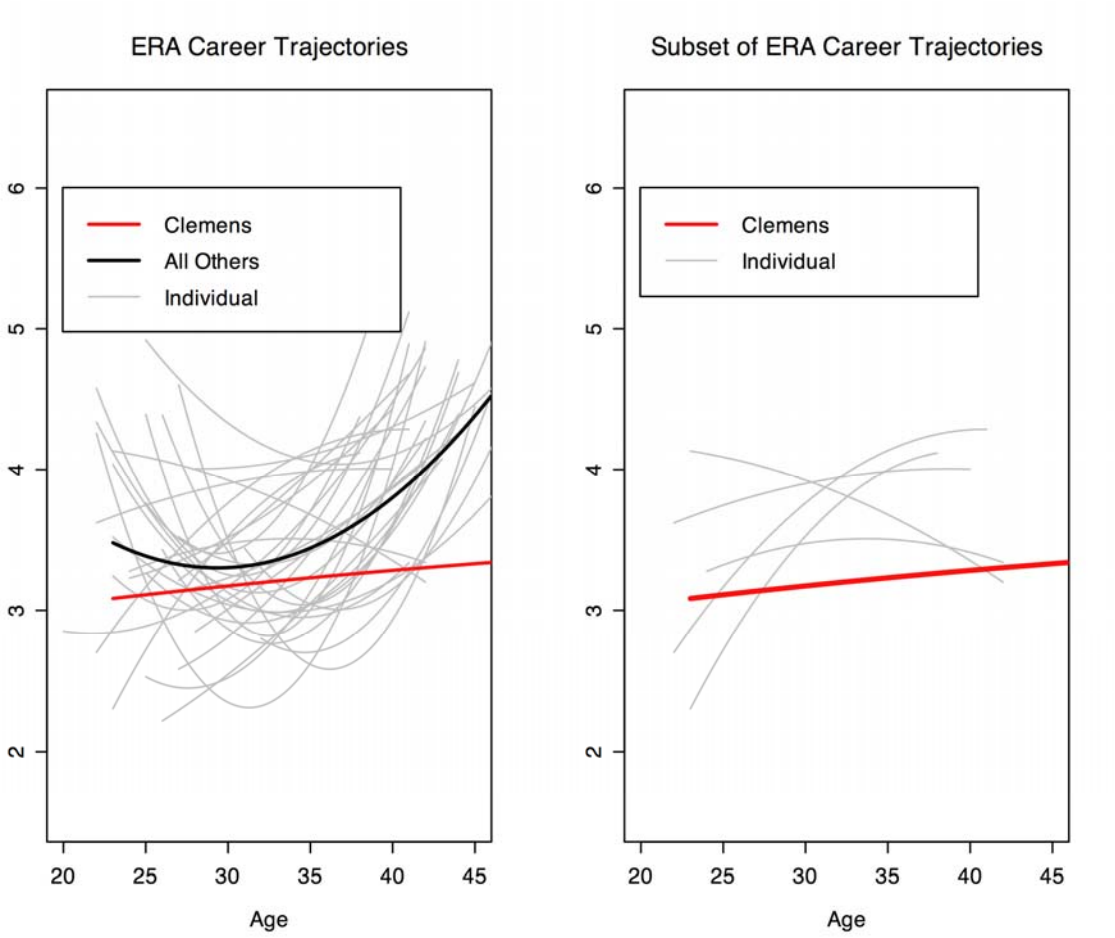
[5] K rate ($\beta_{2ij} < 0$ and career low = $-\beta_{1ij}/\beta_{2ij} \approx 29$).

Note the sign change for K rate for β_{2ij} as more strikeouts is better, while a lower value for the other statistics is better. Figures 6A and 6B below contain a more detailed analysis of the data from the Hendricks report, using ERA. We first present in Figure 6a the ERA curves for the 32 relevant players (31 pitchers + Clemens). Each individual trajectory is depicted with a gray curve, except for Clemens who has a red curve. Also given is a black curve which is the quadratic trajectory fit to the data for all 31 players except

[5] We report the empirically estimated values here out of interest.

Clemens. Figure 6b contains the players with curves that have quadratic terms which are "atypical" ($\beta_{2ij} \leq 0$) compared to the prior hypothesis of a mid-career prime. Six players, including Clemens, have these atypical curves, and in fact Clemens' curve looks quite atypical even within this subset of six players

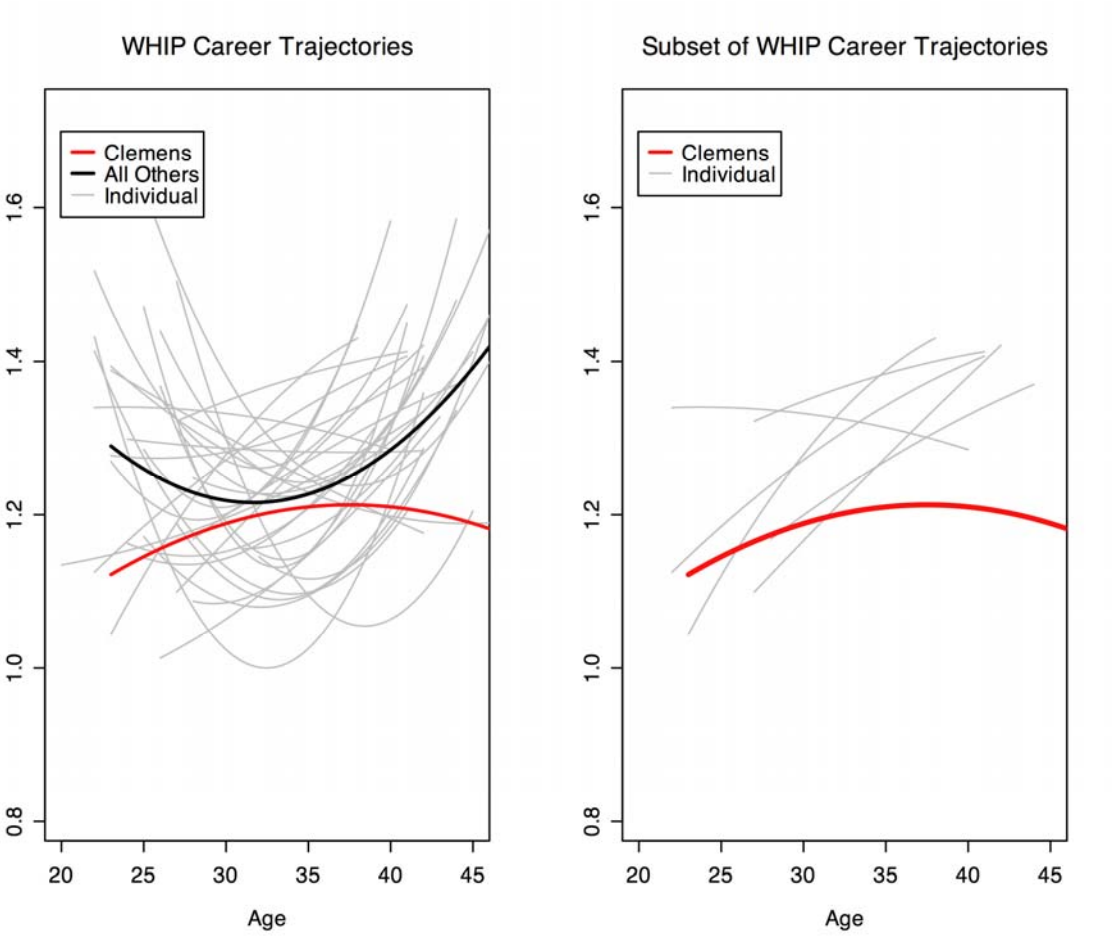
Figure 6A Figure 6B



Figures 7A and 7B below contain career trajectories of WHIP for the same 32 players. Roger Clemens is again within a small subset of seven pitchers that show atypical career

paths. Further inspection of his WHIP curve suggests that he was the only pitcher to get worse as his career went on and then improve at the end of his career.

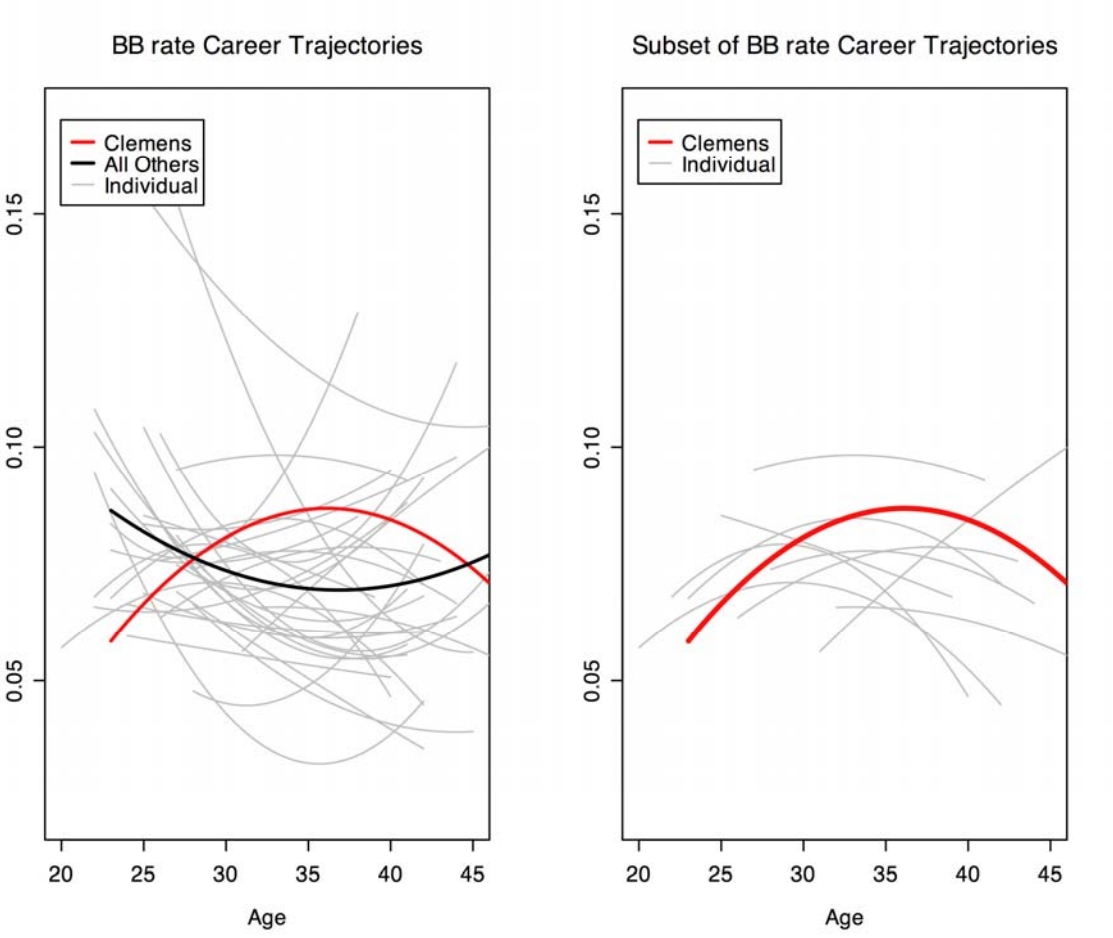
Figure 7A Figure 7B



Two additional analyses we performed using ERA and WHIP were to compute the same figures as Figures 6 and 7, but instead using ERA margin and WHIP margin, defined as the difference between the individual ERA and the league average. In Appendix 2, we show the ERA margin and WHIP margin curves for Roger Clemens and for the average over the 31 other pitchers, and we see little difference between the raw curves (Figures 6 and 7) and the margin curves.

Figures 8A and 8B below contain career trajectories of BB rate (Walks per Batter Faced) for the same 32 players. For BB rate, we note that there are 10 pitchers who have 'inverted-U' fits to their data with Clemens being one of them. Furthermore, the "steepness" of his improvement is particularly noticeable in the later years, even amongst this set of 10.

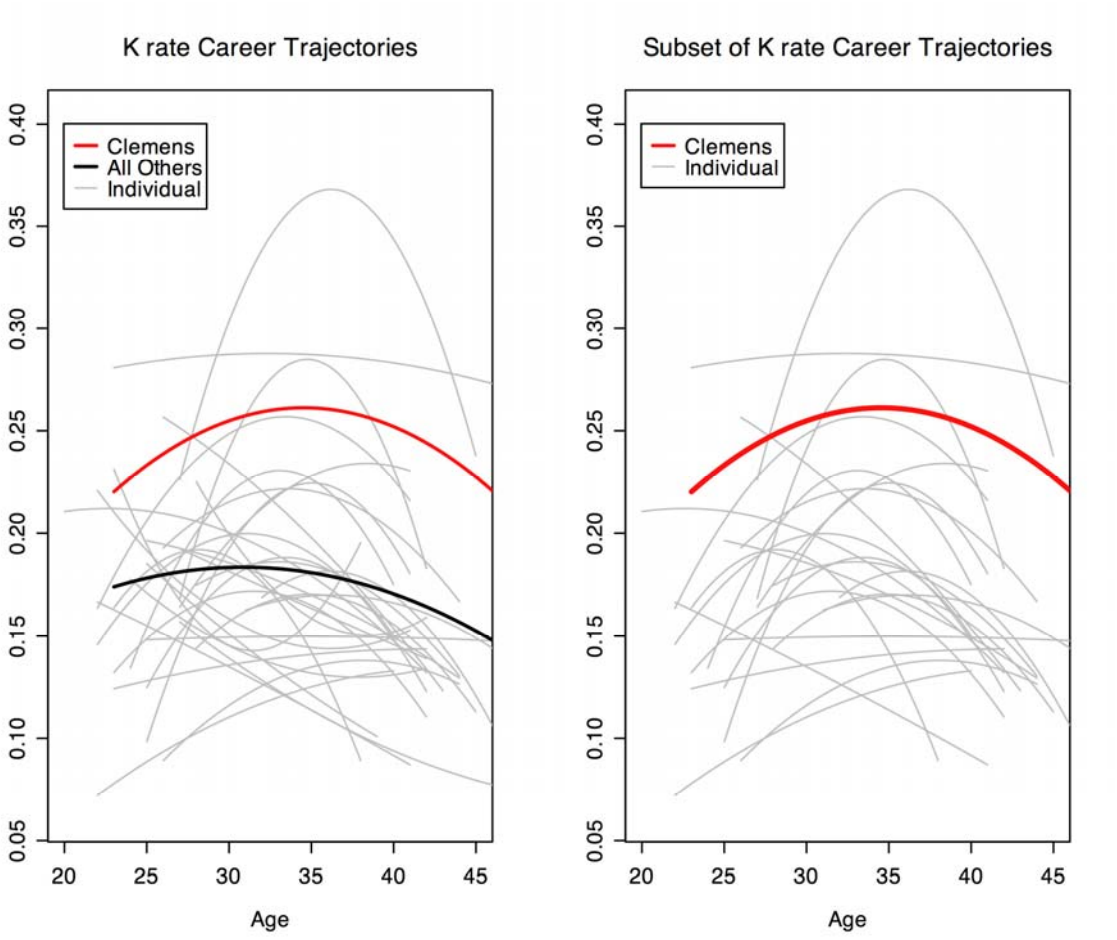
Figure 8A **Figure 8B**



There are several pitching measures for which Clemens' career trajectory does not look atypical, which is the central assertion of the Hendricks report. In Figures 9A and 9B below, we give the strikeout rate (K per non-BB batters faced) for each of the 32

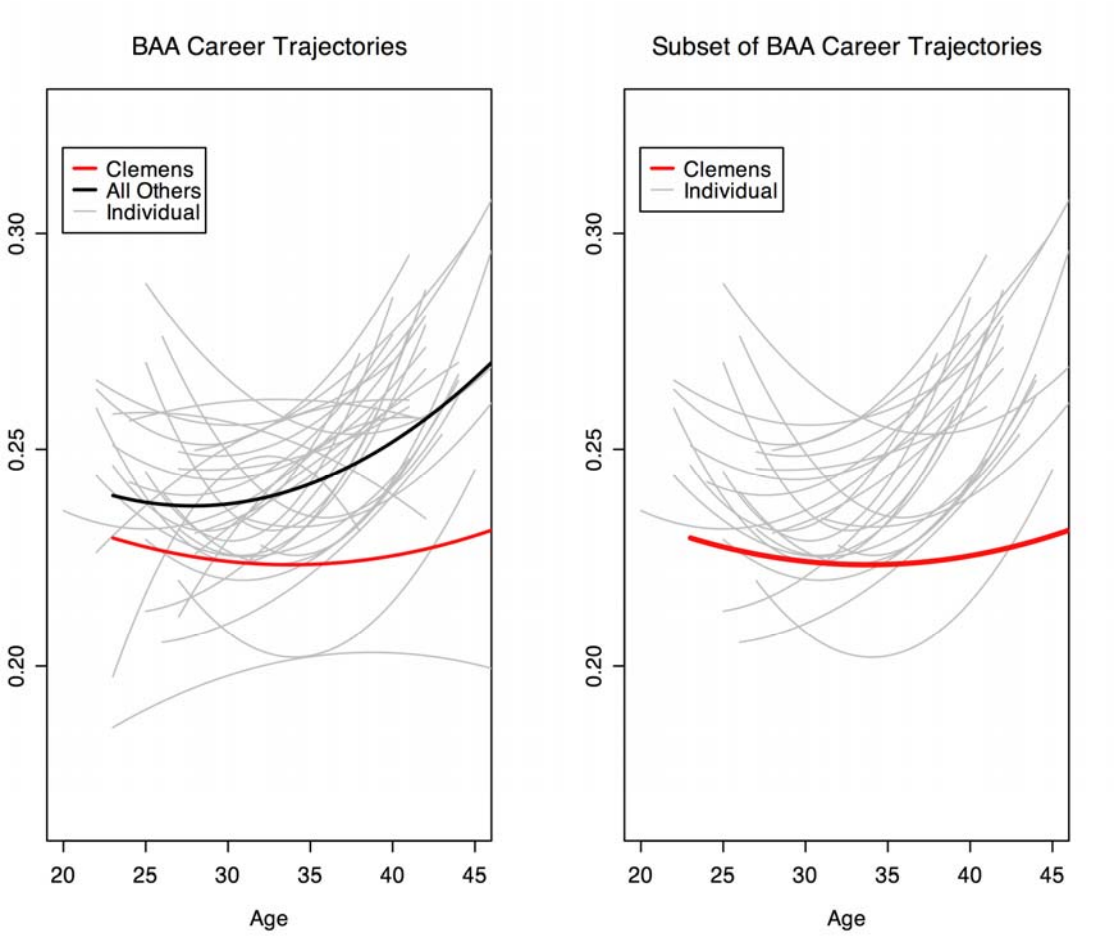
durable starting pitchers. Clemens does have an overall higher K rate than most pitchers in this set, but his career trajectory follows a similar shape (β_{2ij} <0) to 24 of the other 31 players at least with respect to the quadratic fit.

Figure 9A Figure 9B



In Figures 10A and 10B below, we examine BAA (Batting Average Against) for each of the 32 pitchers. Similar to K rate, we again observe that Clemens has a typical shape to his career trajectory to most (24 out of 31) of these other starting pitchers, albeit his curve is somewhat flatter.

Figure 10A Figure 10B



5. Summary Conclusions

Through the use of simple exploratory curve fitting applied to a number of pitching statistics, and for a well-defined set of long-career pitchers, we assessed whether Roger Clemens pitching trajectories were atypical. Our evidence is suggestive that while

most long-term pitchers have peaked mid-career and decline thereafter, Roger Clemens (for some key statistics) worsened mid-career and improved thereafter.

So, what can we conclude? We can conclude that his pitching career was statistically atypical for long-term pitchers in terms of WHIP, BB rate and ERA, and in particular, Clemens shows an end-of-career improvement that is rarely seen. The data does not exonerate (nor indict) Roger Clemens, since an exploratory statistical analysis of this type never proves innocence or guilt. After analyzing this data set there are at least as many questions remaining as before.

References

Albert, James, (2006), Pitching Statistics, Talent and Luck, and the Best Strikeout Seasons of All-Time, Journal of Quantitative Analysis in Sports, Vol 2, Issue 1, Article 2.

Fainaru-Wada, Mark and Lance Williams (2007), *Game of Shadows*: *Barry Bonds, BALCO, and the Steroids Scandal That Rocked Professional Sports.* Gotham Books: March 2007.

Randal A. Hendricks, Stephen Mann and Bret R. Larson-Hendricks (2008) "An Analysis of the Career of Roger Clemens." January 18, 2008 ,<http://www.rogerclemensreport.com>

Mitchell, George J. (2007),"Report to the Commissioner of Baseball of an Independent Investigation into the Illegal Use of Steroids and Other Performance Enhancing Substances by Players in Major League Baseball", DLA Piper US LLP: December 13, 2007. (Accessed at: http://mlb.mlb.com/mlb/news/mitchell/index.jsp.)

Silverman, M. (1996), "Baseball END OF AN ERA: No return fire from Sox Brass tried to keep ace", Boston Herald, December 14th, 1996, page 40.

Vecsey, George. (1968). "Baseball rules committee makes 3 decisions to produce more hits and runs." New York Times. December 4, 1968. p. 57.

Appendix 1: Durable Starting Pitchers (1968-2007)

Name	Number of Years	First Year	Last Year	Innings Pitched
Doyle Alexander	18	1971	1989	3261
Vida Blue	15	1971	1986	3263
Bert Blyleven	21	1970	1992	4950
Kevin Brown	17	1989	2005	3228
Steve Carlton	20	1968	1987	4938
Roger Clemens	24	1984	2007	4917
Chuck Finley	15	1988	2002	3060
Tom Glavine	20	1988	2007	4300
Orel Hershiser	15	1984	1999	3072
Fergie Jenkins	16	1968	1983	4015
Randy Johnson	18	1989	2007	3768
Tommy John	21	1968	1989	4011
Jerry Koosman	18	1968	1985	3817
Greg Maddux	21	1987	2007	4783
Denis Martinez	20	1977	1996	3832
Jack Morris	16	1979	1994	3672
Jamie Moyer	20	1986	2007	3519
Mike Mussina	17	1991	2007	3362
Joe Niekro	16	1968	1987	3201
Phil Neikro	20	1968	1987	5057
Jim Palmer	15	1969	1983	3581
Gaylord Perry	16	1968	1983	4281
Rick Reuschel	17	1972	1990	3517
Jerry Ruess	20	1970	1989	3655
Nolan Ryan	26	1968	1993	5383
Curt Schilling	16	1992	2007	3116
Tom Seaver	19	1968	1986	4532
John Smoltz	15	1988	2007	3082
Don Sutton	21	1968	1988	4824
Frank Tanana	20	1974	1993	4162
Bob Welch	16	1978	1993	3023
David Wells	18	1990	2007	3259

Appendix 2: ERA and WHIP Margin Curves