"Productivity in Baseball: How Babe Ruth Beats the Benchmark"

Peter A. Groothuis*
Walker College of Business
Appalachian State University
Boone, NC 28608
Telephone: 828-262-6077
Email: groothuispete@gmail.com

Kurt W. Rotthoff
Stillman School of Business
Seton Hall University
South Orange, NJ 07079
Telephone: 973-761-9102
Email: Rotthoff@gmail.com

Mark C. Strazicich
Department of Economics
Appalachian State University
Boone, NC 28608
Telephone: 828-262-6124
Email: strazicichmc@appstate.edu

March 9, 2016

*Corresponding author.

Electronic copy available at: http://ssrn.com/abstract=2822082

"Productivity in Baseball: How Babe Ruth Beats the Benchmark"

Abstract

Many count statistics are used to measure the productivity of hitters in Major League Baseball, such as the number of home runs and the number of runs batted in a season. However, comparing the talent of individual players across time is difficult as rules and technologies change. In this paper, we propose applying a practice commonly utilized in the finance literature when comparing the performance of individual stocks and other assets, namely, we "benchmark" the productivity of each player's performance relative to their cohort. Applying our benchmarking strategy to annual Major League Baseball data from 1871-2010, we find that Babe Ruth is the greatest hitter of all time.

JEL Classifications: Z2, J24, C1

Keywords: Benchmarking, Major League Baseball, Structural Change

Electronic copy available at: http://ssrn.com/abstract=2822082

1. Introduction

Productivity for the national economy is typically measured as total output (real GDP) divided by the total hours of labor employed for a given period of time. This number provides a measure of productivity for the average worker and time series on this measure are available for many years. Using this measure, we can compare the productivity of the average worker in 2010, for example, with that of the average worker in 1929. Of course, we expect that the productivity of the average worker in 2010 will be higher than that of the average worker in 1929 due to innovations in technology and greater physical and human capital per worker. Similarly, at the micro level, if we compare the productivity of individual workers across time and include workers from 1929 and 2010, we expect that the most productive would come from 2010 for the same reasons described above.

Given that productivity changes, is there a more accurate way to measure and compare the talent of individuals across time? In this paper, we propose applying a practice commonly utilized in the finance literature when comparing the performance of individual stocks and other assets, namely, we "benchmark" the productivity of each player's performance relative to their cohort in the same time period. We argue that by doing so we can control for changing rules and technologies that may have impacted the productivity of players in general. After applying our benchmarking approach to several measures of hitting performance using data from 1871-2010, we find that Babe Ruth is the greatest hitter of all time.

In the next section, we discuss some of the relevant literature and provide additional background discussion. In Section three, we describe the data and methods that

we utilize to identify the benchmark. In Section four, we evaluate talent both by comparing players to an absolute standard and to a changing benchmark. We conclude in Section five.

2. Background

The sports business is a convenient source of data to examine talent over time. Kahn (2000) suggests that the sports business provides a labor market laboratory given the large amount of productivity data available. Schmidt (2001) uses time series analyze to explain changes in the competitive balance in baseball to see if talent dilutes due to expansion of the number of teams in the league. Groothuis, Hill and Perri (2009) use National Basketball Association data to explain the dilemma of identifying superstars in the draft. Krautmann (2009) uses MLB data to test if home market size and the revenues generated influence managerial decisions in hiring the most talented players. Barros and Leach (2006) analyze the efficiency of management in sports using data from the English Premier League.

As innovations occur, the productive outcomes of players can change. In many sports it is the equipment that leads to changes in the game, such as innovations in tennis rackets or golf club technologies. In other sports change might arise from the development of a new defensive technique, a new way to swing the bat, throw a pitch, shoot a basket, or hold a putter. Often when players develop successful innovations they are mimicked and the game changes.

As the game changes, comparing talent across different time periods becomes increasingly difficult. Many researchers have attempted to address these concerns. For instance, Gould (2003) hypothesized that the population of athletes is always improving,

which lowers the standard deviation of all performance measures. He suggests that the current generation of superstars is equivalent with past generations, while the average player is improving over time.

In the present paper, we argue that performance of the average player may change due to changes in technology and other rule changes. As a result, we suggest that it is better to look at the relative performance in each year to control for changes that have made players in general more productive over time. Berry et al (1999) use overlapping talent between decades and Bayesian updating techniques to control for the change in talent over time. This technique, however, does not account for "structural breaks" in the game. [1] Shell (1999) comes closest to our technique by utilizing peer effects but measures performance at the career level and not the season level. In addition, he uses many different controls other than peer effects, such as controls for which ball park was the hitter's home park, which position did the hitter play, and the career length of the hitter to control for declining talent. He suggests that his technique provides a direct comparison between the era and players.

Controlling for change over time, however, is difficult when there are structural breaks in the game. To overcome the difficulty of controlling for changing talent over time we propose a simple benchmarking or z-statistic technique that Albert (2006) used on pitching data to address the question: How good are players when benchmarked to their peers? Although this technique does not give us the ability to compare Babe Ruth to Barry Bonds in absolute terms, it does provide the answer to the question: When compared to their peers which player has a better performance? This type of benchmark

[1] By structural break, we imply a significant, but infrequent, permanent change in the level and/or trend of a time series.

technique is common in finance, where performance of an asset is not simply measured by the absolute return, but the return relative to some benchmark. In such cases, the benchmark is established as a market portfolio or Security Market Line (Roll 1978) where the portfolio manager's goal is to 'beat the market'. Similar benchmarking is used in many other ways. For example, salaries are benchmarked to relative pay. Given changes in technology over time, research output, teaching performance, and other measures can be similarly benchmarked. Applying a similar relative measure to sporting events will allow us to more accurately compare individual players who may have played in very different eras.

The use of benchmarking allows for more accurate analysis of performance. For example, benchmarking can be expanded to provide more detailed analysis of talent levels; not just of today's performances, but as a viable measure of talent across different eras with different innovations.[2] Given that talent is highly valued, providing an accurate measure of relative talent today, and comparisons across time may provide valuable information.

In some industries, however, talent is difficult to measure. This measurement difficulty increases as you measure the talent within an industry over long periods of time, for instance from 1871-2010. This problem is also complicated by the fact that when the opportunity to reveal talent is limited, true talent does not have the opportunity to reveal itself (Terviö 2009).

Given that talent changes over time, using a z-statistic technique can provide a more accurate measure of a player's performance at a given point in time. Throughout the

[2] For example, a firm comparing the absolute level of productivity of employees pre- and post-computers would provide an inaccurate measure of an individual employee's relative productivity.

years technologies, skills, strength, and training have changed. Having a benchmark increases the accuracy of measurement and the ability to compare talent of individual players over time. Moreover, utilizing a benchmark provides a convenient method to identify superstars and may provide new insights to help identify innovative players who changed the game.

3. Data and Structural Breaks

MLB has a long history beginning in the late 1800s that continues to this day. As in all sports leagues, superstars are commonly identified in the record books using an absolute standard. In this paper, we propose adopting a benchmarking strategy to identify superstars by examining the deviation in performance from the mean of their peers. To perform our calculations, we utilize time series on slugging percentage (SLUG), home runs per hundred at bats (HR), batting average (BAVE), and runs batted in per hundred at bats (RBI) from Sean Lahman's Baseball Database on all players from 1871-2010 with at least 100 at-bats.[3] We calculate the mean and standard deviation of each performance measure for each season. This provides annual time series from 1871-2010 with 140 seasonal observations for each series. With 35,728 single season observations we find that the average player hit 7 homeruns per season (with a maximum of 73), had 42.5 runs batted in (RBI), and a slugging percentage of .379.

In a recent paper, Groothuis, Rotthoff, and Strazicich (forthcoming) find "structural breaks" in several measures of average MLB batting performance. In particular, they find that MLB performance had a structural break around 1920-1921. As a result of this finding, the authors suggest that with the success of Babe Ruth's "free

[3] Sean Lahman's Baseball Database: http://baseball1.com/2011/01/baseball-database-updated-2010/. Slugging percentage is calculated as total bases divided by the number of at-bats.

swinging" style others that could mimicked his innovation. They find another notable break in 1992, which is closely associated with the early years of the modern steriod era. The existence of breaks and (stationary) deterministic trends in average player performance provides additional support for adopting a benchmark technique when evaluating talent over time.

4. Benchmarking

The deliberation on superstars and their relative performance is oft debated and hard to measure, particularly when the comparison happens over different periods of time. When structural breaks occur in the game it makes accurate comparisons nearly impossible over time. A more accurate way to measure talent across time should yield more accurate identification of truly great stars. Given a seemingly endless set of debates and lists of superstars we propose a measurement technique to compare stars relative to their same generational cohort.

In Tables 1-4, we report the means of batting average, slugging percentage, home runs per hundred at bats, and runs batted in per hundred at bats, respectively. In each table we report the top ten talented players as measured in absolute terms by the overall standard deviations above the overall mean of all years and the benchmark measure as the yearly standard deviation above the yearly mean. The first measure treats the entire population as peers and does not account for changes in the game. The second technique compares talent directly to peers during the time of play.

In Table 1, we report the ten players with the best batting average. We find that when using the absolute measure the ten best players all occur in the early years of baseball with eight of the ten in the late 1800s, one in 1901, and the last, Roger Hornsby,

in 1924. However, when using the benchmarking measure we find that the ten best players come from all eras in baseball. Manny Ramirez is the most recent, hitting 3.75 standard deviations above the season mean. Other notables on this list are Ted Williams in 1941, George Brett in 1980, and Tony Gwynn in 1994.

We next report results of the slugging percentage for both measures of talent in Table 2. Using the absolute standard (SD above the absolute mean), Babe Ruth makes the top ten list four times and Barry Bonds three times. The other three making the top ten are Lou Gehrig, Roger Hornsby, and Mark McGwire. Using the z-statistic (SD above the season mean) we find that Babe Ruth makes the list five times including the top two rankings in 1920 and 1921. Interestingly, these years coincide to the time period where Groothuis, Rotthoff, and Strazicich (forthcoming) find a structural break in the mean slugging percentage series of all players. Using the same benchmarking standard, Barry Bonds makes the top ten list in 2001, 2002 and 2004, which follows the second notable structural break (in 1992) identified by Groothuis, Rotthoff, and Strazicich. Other players that make the list in the benchmarking standard are Lou Gehrig in the eighth position and Ted Williams in the ninth.

We next turn our attention to home runs. The means are reported in Table 3. Using the absolute standard, Barry Bonds and Mark McGwire dominate the list of the top thirteen players. Bonds is in the first position hitting home runs 7.37 standard deviations above the mean and making the list three times followed by McGwire making the list six times in the second through seventh position. Note that the majority of these stars come from the latter years of baseball. In comparison, Babe Ruth only makes the list in the tenth position in 1920 hitting 5.45 standard deviations above the absolute mean. Using

the absolute standard Babe Ruth is not the best home run hitter in baseball. In contrast, when applying the benchmarking standard by utilizing home runs per at-bats for each individual player and ranking the standard deviations above the mean for each year, Babe Ruth is the top ranked home run hitter in 1920 (Yankees), 1921 (Yankees), 1919 (Boston), and 1927 (Yankees). In particular, Babe Ruth was 10.58, 8.07, 7.26, and 7.04, respectively, standard deviations above the mean during these years. When compared to his peers Babe Ruth was clearly the best home run hitter in history. The fifth highest ranked player is Ned Williamson (1884 Chicago), followed by Ruth (1926), Ruth (1924), Buck Freeman (1899 Washington Senators), Ruth (1928), and Gavvy Cravath (1915 Phillies). From the modern era the highest ranked players are Barry Bonds (2001 San Francisco), in thirteenth place, at 5.85 standard deviations above the mean, and Mark McGwire (1998 and 1997 St Louis) in nineteenth and twentieth place at 5.4 standard deviations above the mean.

In particular, Babe Ruth, in his 1920 playing season with the New York Yankees was 10.58 standard deviations above the mean. This is simply amazing and displays his level of performance relative to the competition that he faced. To put this in perspective, if Babe Ruth was 10.58 standard deviations above the mean in 2001, when Barry Bonds set the single season home run record, and had the same 476 at-bats that Barry Bonds did, he would have hit 120 home runs. At the time of this writing, Barry Bonds still holds the single season record with 73 home runs.

Next, we measure the RBIs per at-bat of players throughout time to measure how each player performs relative to the mean of the year played, again with at least 100 at-bats. In Table 4, we report the results of the superstars as measured by standard

deviations above the mean. We find that Reb Russell, playing for the Pittsburgh Pirates, has the highest ranking of RBIs both using the absolute and benchmark standards.[4] He was 5.04 above the absolute mean and 4.93 standard deviations above the season mean. Other notable players on the absolute standard list are Babe Ruth in 1921 in the sixth position, Manny Ramirez in 1999 in the seventh position and Mark McGwire in 2000 in the ninth position. Using the benchmark standard Babe Ruth has five of the top ten rankings of RBIs. Babe Ruth ranks third, fourth, fifth, sixth and tenth. No player from the modern era makes the top ten. We do find that Manny Ramirez is ranked thirteenth and twentieth as the highest ranked modern era player.

5. Conclusion

As innovations occur and productivity changes, individual performance becomes increasingly difficult to compare across time. When this occurs, relative measures have more value. In this paper, we utilize a common practice in the finance literature and suggest that adopting a benchmark measurement of performance provides a more accurate method to compare individual player performance across time and identify superstars. Applying our benchmarking technique to annual MLB data from 1871-2010, we find that Babe Ruth was the best power hitter compared to his peers. In particular, Babe Ruth was more than ten standard deviations above the mean in 1920, which is simply amazing. Even among current players the best are more than five standard deviations above the mean. With performance measures this high it is no surprise that the most highly productive players are paid so much more than the average wage in the league.

[4] Reb Russell was a pitcher from 1912-1917 with the Chicago White Sox. He did not become a big hitter until after developing arm troubles and finding his hitting in the minor leagues.

References

Albert, James (2006). "Pitching Statistics, Talent and Luck, and the Best Strikeout Seasons of All-Time," *Journal of Quantitative Analysis in Sports* 2:1, 1-30.

Barros, Carlos Pestana, and Stephanie Leach (2006). "Performance evaluation of the English Premier Football League with data envelopment analysis," *Applied Economics* Volume 38, Issue 12, 1449-1458.

Berry, Scott, C., Rees, Shane, and Patrick D. Larkey (1999). "Bridging Different Eras in Sports," *Journal of the American Statistical Association* 94:447, 661-667.

Gould, S. (2003). *Triumph and Tragedy in Mudville: A Lifelong Passion for Baseball.* New York: W. W. Norton.

Groothuis, Peter A., James Richard Hill, and Timothy Perri (2009). "The dilemma of choosing talent: Michael Jordon's are hard to find," *Applied Economics* 41 (25), 3193-3198.

Groothuis, Peter A., Kurt W. Rotthoff, and Mark C. Strazicich (forthcoming). "Structural Breaks in the Game: The Case of Major League Baseball," *Journal of Sports Economics.*

Kahn, Lawrence (2000). "The Sports Business as a Labor Market Laboratory," *Journal of Economic Perspectives* 14:3, 75-94.

Krautmann, Anthony (2009). "Market size and the demand for talent in major league baseball," *Applied Economics* 41:25, 3267-3273.

Roll, R. (1978). "Ambiguity when Performance is Measured by the Securities Market Line," *Journal of Finance* 33(4), 1051-69.

Schell, M. (1999). *Baseball's All-Time Best Hitters*, Princeton, NJ: Princeton University Press.

Schmidt, Martin B. (2001). "Competitive Balance in Baseball," *Applied Economics Letters* 8:1, 21-26.

Terviö, Marko (2009). "Superstars and Mediocrities: Market Failure in the Discovery of Talent," *Review of Economic Studies* 72(2), 829-850.

Table 1: Batting Average: Absolute Standard vs. Benchmark

	Player	Year	SD above the absolute mean	Rank	Player	Year	SD above the season mean
1	Levi Meyerle	1871	5.659514	1	Bob Hazle	1957	3.86
2	Hugh Duffy	1894	4.368691	2	Manny Ramirez	2008	3.75
3	Tip O'Neill	1887	4.258268	3	Ted Williams	1941	3.69
4	Ross Barnes	1872	4.187384	4	George Brett	1980	3.68
5	Cal McVey	1871	4.164273	5	Tip O'Neill	1887	3.65
6	Ross Barnes	1876	4.095538	6	Tony Gwynn	1994	3.59
7	Nap Lajoie	1901	4.043987	7	Oscar Gamble	1979	3.57
8	Ross Barnes	1873	4.019332	8	Tris Speaker	1916	3.54
9	Willie Keeler	1897	3.977446	9	David Dellucci	1999	3.54
10	Roger Hornsby	1924	3.971277	10	Jack Glasscock	1884	3.51

Table 2: Slugging Percentage: Absolute Standard vs. Benchmark

Rank	Player	Year	SD above the absolute mean	Rank	Player	Year	SD above the season mean
1	Barry Bonds	2001	5.65	1	Babe Ruth	1920	5.77
2	Babe Ruth	1920	5.49	2	Babe Ruth	1921	5.21
3	Babe Ruth	1921	5.45	3	Barry Bonds	2001	5.03
4	Barry Bonds	2004	5.06	4	Barry Bonds	2004	4.91
5	Barry Bonds	2002	4.90	5	Barry Bonds	2002	4.79
6	Babe Ruth	1927	4.59	6	Babe Ruth	1927	4.57
7	Lou Gehrig	1927	4.51	7	Babe Ruth	1926	4.50
8	Babe Ruth	1923	4.50	8	Lou Gehrig	1927	4.49
9	Rogers Hornsby	1925	4.40	9	Ted Williams	1941	4.36
10	Mark McGwire	1998	4.36	10	Babe Ruth	1924	4.35

Table 3: Home Runs: Absolute Standard vs. Benchmark

Rank	Player	Year	SD above the absolute mean	Rank	Player	Year	SD above the season mean
1	Barry Bonds	2001	7.37	1	Babe Ruth	1920	10.58
2	Mark McGwire	1997	6.53	2	Babe Ruth	1921	8.07
3	Mark McGwire	1998	6.51	3	Babe Ruth	1919	7.26
4	Mark McGwire	2000	6.40	4	Babe Ruth	1927	7.04
5	Mark McGwire	1999	5.81	5	Ned Williamson	1884	7.01
6	Mark McGwire	1995	5.71	6	Babe Ruth	1926	6.83
7	Mark McGwire	1996	5.71	7	Babe Ruth	1926	6.50
8	Hill Glenallen	2000	5.62	8	Buck Freeman	1899	6.41
9	Barry Bonds	2004	5.58	9	Babe Ruth	1928	6.11
10	Babe Ruth	1920	5.45	10	Gavvy Cravath	1915	6.08
11	Barry Bonds	2003	5.30	13	Barry Bonds	2001	5.85
12	Frank Thomas	2005	5.24	19	Mark McGwire	1998	5.42
13	Barry Bonds	2002	5.23	20	Mark McGwire	1997	5.41

Table 4: RBIs: Absolute Standard vs. Benchmark

Rank	Player	Year	SD above the absolute mean	Rank	Player	Year	SD above the season mean
1	Reb Russell	1922	5.04	1	Reb Russell	1922	4.93
2	Hack Wilson	1930	4.71	2	Cap Anson	1886	4.74
3	Sam Thompson	1894	4.62	3	Babe Ruth	1920	4.65
4	Charlie Ferguson	1887	4.61	4	Babe Ruth	1919	4.21
5	Rynie Wolters	1871	4.54	5	Babe Ruth	1921	4.21
6	Babe Ruth	1921	4.49	6	Babe Ruth	1926	4.20
7	Manny Ramirez	1999	4.47	7	Charlie Furguson	1887	4.17
8	Jimmie Foxx	1938	4.33	8	Gavvy Cravath	1913	4.05
9	Mark McGwire	2000	4.32	9	Joe Wood	1921	4.04
10	Joe Wood	1921	4.32	10	Babe Ruth	1932	4.03