# Performance Persistence in Baseball Free Agency

WH299/399 Data and Methodology

*Chris Hua*

*Advisor: Linda Zhao*

*24 October, 2016*

## Introduction

This paper aims to determine what causal impact the publication of Michael Lewis's *Moneyball* (Lewis 2003) and the popularization of statistical techniques to determine player worth had on the business of sports.

In general, most people accept that there have been fundamental changes to how teams evaluate talent, as well as how they price players. There is no shortage of popular press giving anecdotal evidence that some 'Moneyball effect' exists (see e.g. Futterman (2011)), but relatively little work has been done to quantify this effect.

We will focus on the question of 'performance persistence' in salary allocations.

*Moneyball* posited that certain statistics were undervalued by most baseball teams when considering contract negotiations, and that teams which properly valued those characteristics were able to reap outsized benefits. Previous research by Hakes and Sauer (2006) found that there was an inefficiency for the particular statistics that the Oakland A's used before 2004, and that after the publication of *Moneyball* and the increased transparency around those statistics, the market inefficiency was closed. Other teams adapted to what the Oakland A's used and increased how they valued those statistics.

However, even if these particular inefficiencies were played out, it is likely that other inefficiencies exist. Broadly, we would expect that MLB teams get a better sense of which features to focus on when evaluating players over time. Lewis focused on on base percentage and slugging, and Hakes and Sauer (2006) shows that these underpriced attributes were quickly brought to an efficient position quickly. We would expect that teams have continued to search out other under-priced attributes of players, and that as time goes on, fewer underpriced features to remain untapped. The diminishing marginal returns of quantitative analysis to

professional teams is quite possibly a concern, but more to the point of this paper, it would be valuable to explore which characteristics were discovered and when.

As a corrolary, it's likely that certain teams have 'better' (or luckier) quantitative analysts in their employ. In this case, we would expect different teams to take advantage of different pricing inefficiencies at different times; that is, a more quantitative team would exploit a pricing inefficiency before a less advanced team. By performing a temporal analysis, we could discover which teams have the most impactful analytics departments.

## Data

We identify a couple sources of data.

- Historical player performance data is readily available from reference sites, such as Baseball Reference or FanGraphs, which both also offer their own Wins above Replacement metric.
- Various R packages offer connectivity with baseball data. Sean Lahman's baseball package offers an easy store of historical performance data, without requiring webscraping.
- SpotTrac offers contract data for various sports, including Major League Baseball.
- Other sources for contract data include Baseball Prospectus or ESPN.

## Methodology

We will examine the contracts signed during each offseason between 1996 and 2012, by free agents who switched teams. We may split the dataset to include or exclude pitchers. Our methodology borrows from Ahlstrom, Si, and Kenelley (1999), who reviewed free agent contracts and performance, and who in turn examined the specific methodology performed by previous authors.

From this subset of contracts, for each year, we will estimate a logistic regression model based on past player performance, for some set of statistics, and compare the change in the coefficients over time. We will also perform statistical tests to compare the differences in the coefficient estimates. Our motivation here is to identify what features were relevant in any particular year to the overall market.

Futhermore, we will estimate a measure of player performance relative to signing their contracts. Ahlstrom, Si, and Kenelley (1999) present a comparison of equity and expectation theory in the MLB free agent market. We will perform comparisons between player performance in a number of years before signing their contract,

and again in a number of years after the contract. For player performance, we can use traditional statistics (e.g. batting average, home runs) or more advanced statistics such as Wins Above Replacement. Alternatively, we can take the post-contract years, and given the regression model that we estimated for the year they were signed, see if the player has performed at a reasonable level given their contract expectations. This can give us an idea of which free agent signings were effective.

## References


Ahlstrom, David, Steven Si, and James Kenelley. 1999. "Free Agent Performance in Major League Baseball." *Journal of Sport Management* 13: 181–96. http://fitnessforlife.org/AcuCustom/Sitename/Documents/DocumentItem/492.pdf.

Futterman, Matthew. 2011. "Baseball After Moneyball." http://www.wsj.com/articles/SB10001424053111903791504576584691

Hakes, Jahn K., and Raymond D. Sauer. 2006. "An Economic Evaluation of the Moneyball Hypothesis." *Journal of Economic Perspectives* 20 (3): 173–85. doi:10.1257/jep.20.3.173.

Lewis, Michael. 2003. *Moneyball: The Art of Winning an Unfair Game*. New York: W. W. Norton.