Validating the Malmquist Productivity Index for Industry Change: The Case of Major League Baseball

# Abstract

Several articles use the Malmquist Productivity Index (MPI) to evaluate different industries performances over time. At issue is these articles use a very small set of DMUs over a short time frame to validate the usefulness of the MPI, as raised by their authors. Therefore, validating MPI properly requires a large and verified data set. To address the number DMU issues and validate the MPI, this article utilizes the Lehman’s Baseball Database, which spans baseballs rich history from 1871 to 2015. First validation happens by picking major events and verifying an index change at that particular time. The second validation takes index change and matches it to an event in baseball history.

# Introduction

The Malmquist index was introducing in a paper written by Professor Sten Malmquist in 1953 [1].

# DEA

Data envelopment analysis (DEA) is a non-parametric method in operations research and economics for the estimation of production frontiers[clarification needed]. It is used to empirically measure productive efficiency of decision making units (or DMUs). Although DEA has a strong link to production theory in economics, the tool is also used for benchmarking in operations management, where a set of measures is selected to benchmark the performance of manufacturing and service operations. In the circumstance of benchmarking, the efficient DMUs, as defined by DEA, may not necessarily form a “production frontier”, but rather lead to a “best-practice frontier” (Cook, Tone and Zhu, 2014). DEA is referred to as "balanced benchmarking" by Sherman and Zhu (2013).Non-parametric approaches have the benefit of not assuming a particular functional form/shape for the frontier, however they do not provide a general relationship (equation) relating output and input. There are also parametric approaches which are used for the estimation of production frontiers (see Lovell & Schmidt 1988 for an early survey). These require that the shape of the frontier be guessed beforehand by specifying a particular function relating output to input. One can also combine the relative strengths from each of these approaches in a hybrid method (Tofallis, 2001) where the frontier units are first identified by DEA and then a smooth surface is fitted to these. This allows a best-practice relationship between multiple outputs and multiple inputs to be estimated.

# Malquist

Before using the Malmquist Productivity Index, it must be validated in this use. Is it valid for locating changes in technical progress or technical efficiency, thus allowing us to identify the cause?

To validate Malmquist Productivity Index, a large amount of verified data is required. The data must be One data source that meets these requirements is Baseball Player performance data for 1871 through 2014 found in the Lahman Baseball Database. The advantages of using this database is · Long history of sophisticated methods applied to baseball · Can provide better understanding to the underlying models · Rich and very accurate historical data · Well understood history and easily understood application Using the database to validate the Malmquist Productivity Index, we can prove the following hypotheses · H1: Immature industries show highly variable performance · H2: Rule changes, etc. result in a significant frontier shift o In favor of hitters will result in a positive frontier shift in the the year that they first take effect o In favor of pitchers (against hitters) will result in a negative frontier shift in the the year that they first take effect · H3: New and “Exceptional” performers (ex. with new best practices) will result in: o A positive industrywide frontier shift in their their “first” year (P>1) o Decrease in efficiency as most people suffer in comparison to the star. (E<1) o Returns to stable if their performance stays at that level, later years (P, E=1) · H4: Years without changes should generally have little frontier shift o Both frontier shift and efficiency change are about 1.0 (P, E=1)

[1] R. Färe and A. Klevmarken, “In memoriam,” *Journal of Productivity Analysis*, vol. 23, no. 2, pp. 141–142, May 2005 [Online]. Available: <http://link.springer.com/article/10.1007/s11123-005-1325-8>. [Accessed: 23-Mar-2016]