

Descriptive and Graphical Analysis of the NBA Data

Carlo Favero

The NBA teams database

Our first database is made of 40 seasons (from 1979-1980 to 2018-2019) for all NBA teams, we obtain it at the following link:

https://www.basketball-reference.com/leagues/NBA_2018.html#all_opponent-stats-base.
Files contained in the **data_NBAtteams.zip** allow you to reconstruct the database.

The final version of the database is named **teams_overall2020.csv**.

Building the database

There is no API for basketball reference so the procedure to download the data has been slightly cumbersome.

- gone to www.basketball-reference.com and select season by season the summary page.
- scrolled down the screen to reach the tables called "Team Stats", "Opponent Stats" and "Miscellaneous Stats"
- Saved them as Excel files with the names `data_team_xx.xlsx`, `data_opp_xx.xlsx` and `data_misc_xx.xlsx` where `xx=1...40`
- the data base can be updated year by year just by adding three files for the new season
- run the R programme **dataset.R** that combines all data and produces a csv file called `Teams_overall2020.csv`

The relevant dimensions of the data

- There are two relevant dimensions in our data set
 - cross-section (in each year we observed data for all the different teams)
 - time-series (for each team we have 40 seasons of data)
- In general, we shall define $X_{i,t}$ as the statistics observed at time t for team i .
 - the t index captures the time-series dimension
 - the i index captures the cross-sectional dimension

Data Transformation

- After importing the data in the statistical package, the first step in the analysis is data transformation and organization.
- In R data are imported in a data-frame
- We can use the data-frame features to transform the data and organize them (for example, take subsets or sort them)

Descriptive Analysis

- Descriptive analysis can be univariate or multivariate
- Analysis of the marginal distribution of a variable
- Correlation analysis

- scatter-plots
- time-series graphics
- multiple graphs
- density estimates (histograms)

The idea is to plot in a standard Cartesian reference graph:

- the quantiles of the series under consideration, X_t , against the quantiles of any given distribution. If the returns were truly normal, then the graph should look like a straight line with a 45-degree angle.
 - first, sort all (standardized) returns in ascending order, and call the i th sorted value x_i ;
 - second, compute the empirical probability of getting a value below the actual as $(i - 0.5)/T$, where T is number of observations available in the sample.
 - Finally, we calculate the quantiles of the benchmark distribution quantiles as $\Phi^{-1}((i - 0.5)/T)$, where $\Phi^{-1}(\cdot)$ denotes the inverse of the benchmark density.
 - Represent on a scatter plot the (standardized) returns and sort the data on the Y-axis against the standard distribution quantiles on the X-axis.

Matrix Representation of the data

A matrix is a double array of i rows and j columns, whose generic element can be written as a_{ij} , it is a convenient way of collecting simultaneously information on the time-series and the cross-section of returns:

$$A = \begin{bmatrix} a_{11} & \cdot & \cdot & a_{1j} \\ & & & \\ a_{i1} & & & a_{ij} \\ & & & \end{bmatrix}, 0 = \begin{bmatrix} 0 & \cdot & \cdot & 0 \\ & & & \\ 0 & & & 0 \\ & & & \end{bmatrix}$$
$$I = \begin{bmatrix} 1 & \cdot & \cdot & 0 \\ & & & \\ 0 & & & 1 \\ & & & \end{bmatrix}$$

Matrix Operations

- Transposition $a'_{ij} = a_{ji}$
- Addition: For A and B $n \times m$ $(a + b)_{ij} = a_{ij} + b_{ij}$
- Multiplication: For A $n \times m$ and B $m \times p$ $(ab)_{ij} = \sum_{k=1}^m a_{ik}b_{kj}$
- Inversion for non-singular A $n \times n$, A^{-1} satisfies $A^{-1}A = AA^{-1} = I$

An Illustration with NBA data

- Construct a time-series plot of GSW pace over the seasons
- Did the share of three points/field court shots taken by Chicago Bulls increase over time ?
- Did the relative efficiency of three points and two points taken by Chicago Bulls increase over time ?