LS88: Sports Analytics

Dean Oliver's
Four Factor Basketball Model

Dean Oliver's Four Factor Model

An empirically driven model that explains basketball performance

It aims to answer "How do teams win basketball games?"

It features four fundamental components of the sport

- ➔ Score efficiently
- ➔ Protect the basketball on offense
- ➔ Grab as many rebounds as possible
- ➔ Get to the foul line and make FTs as often as possible

Dean Oliver's Four Factor Model

The Model

$$
\begin{aligned}
\text{Team Performance} = &\ .4 \times Z(eFG\% - eFG\%_{\text{Opp}}) \\
&- .25 \times Z(\text{Turnover\%} - \text{Turnover\%}_{\text{Opp}}) \\
&+ .2 \times Z(OREB\% - OREB\%_{\text{Opp}}) \\
&+ .15 \times Z(\text{FT Rate} - \text{FT Rate}_{\text{Opp}})
\end{aligned}
$$

What are the *Z*s? What are the weights?

Dean Oliver’s Four Factor Model

The outline

- ➔ What is a factor model?
- ➔ How are they used? They’re ubiquitous in finance
- ➔ How do we compute the Four Factor Model?
- ➔ What does it tell us about basketball?

Factor Models

First, what is a “factor model”?

Basic idea:

- ➔ You have have observed outcomes (stock returns, team performance, etc)
- ➔ There are “factors” or features or just generally quantities that drive performance of your outcomes
- ➔ Typically you take the factor model as linear

$$\text{Outcome} = \beta_1 \cdot \text{Factor 1} + \cdots + \beta_K \cdot \text{Factor K}$$

- ➔ Common in Pyschology and Finance (I’ll talk about finance)

Factor Models

Quick aside:

LWTS was a linear model but not a factor model

We were just estimating the run values of events

Factors instead would be things like:

- ➔ Measures of getting on-base or power
- ➔ Measures of baserunning aggressiveness
- ➔ Measures of pitching or defense

Factor Models in Finance

We use a factor model to explain market phenomena

For example asset prices/returns, or correlation of returns

The most important factor: the “market”

- ➔ The notion of the market is a bit abstract
- ➔ But we can substitute something for the market though. For example DJIA or S&P500 indices
- ➔ A model for stock returns (the famed CAPM)

$$\text{Stock Return} = \text{Expected Return} + \beta \cdot \text{Market Return} + \text{Idiosyncratic Return}$$

Factor Models in Finance

Model for stock returns (the famed CAPM)

$$\text{Stock Return} = \text{Expected Return} + \beta \cdot \text{Market Return} + \text{Idiosyncratic Return}$$

The “market” explain/drives the variation in returns you observe (and after that, it’s all just noise)

Stocks go up/down with the market so every stock is correlated

Factor Models in Finance

Model for stock returns (the famed CAPM)

$$\text{Stock Return} = \text{Expected Return} + \beta \cdot \text{Market Return} + \text{Idiosyncratic Return}$$

Magnitude of idiosyncratic return and beta: fraction of variation due to the market

Beta > 1: stock is extra reactive to market
Beta < 1: stock not as reactive
Beta = 0: stock not affected by market
Beta < 0: stock anti-correlated with market

Factor Models in Finance

Another (potential) factor: momentum

Momentum is recent historical performance. Trending up? Trending down?

What do you do with the factors? Try to build portfolios based on them:

- ➔ Try to match the market return but minimize risk
- ➔ Place a bet on momentum
- ➔ Anything you want: you look at the “betas” and decide how you want the portfolio to exposed to the risk of the factors

Factor Models in Finance

So what can be a factor?

Potentially anything you want. But you might concoct a bogus factor which doesn't explain anything about the market (and potentially lose a lot of money!!)

What else?

Factor Models in Finance

- Macroeconomic factors: economy level data like employment or interest rates
- “Fundamental” factors: things like value (price/earnings ratio) or momentum
  - Anything related to the company
- Statistical: not firm or directly observable
  - Observed in data but nothing fundamental/macro explains it

Factor Models in Finance

What about the weights? How do we determine the importance of factors?

Ultimately, we have to figure out on our own how to weight the factors for each stock

- ➔ Market factor is the most important for stocks (not bonds). But how important?

The short answer: regression

Dean Oliver's Four Factor Model

So what are Dean Oliver's Four Factors? We know the intuitive ideas from basketball:

- ➔ Efficient shooting
- ➔ Turnovers
- ➔ Rebounding
- ➔ Free Throws

These will be fundamental factors: compute values from team observables

Okay, we know the concepts, but what should we actually compute?

Dean Oliver's Four Factor Model

Shooting:

$$eFG\% = \frac{FG + .5 \cdot 3FG}{FGA}$$

Rebounding:

$$\text{OREB\%} = \frac{\text{Off Reb}}{\text{Off Reb} + \text{Opposition Def Reb}}$$

Turnovers:

$$\text{Turnover Rate} = \frac{\text{Turnovers}}{\text{Possessions}}$$

Free Throws:

$$\text{FT Rate} = \frac{FT}{FGA}$$

or $$\text{FT Rate} = \frac{FTA}{FGA}$$

Dean Oliver’s Four Factor Model

Okay, so now what?

- ➔ We want to explain team performance
- ➔ Dean Oliver suggests weighting the factors by importance: 40, 25, 20, 15

Okay… so we just multiply each of those values by those numbers and add them up?

None of those numbers are related at all: different avg level, different variation

Z-Scores and Standardizing Data

A bit of an important digression for a very useful and great tool for standardizing data

Suppose you have some variables on different levels and different variations

We do. In fact, this happens basically all the time.

The goal is to standardize the variables

Remove average level, inherent variation, and units

Z-Scores and Standardizing Data

You can quantify effects in terms of standard units like *standard deviation*

- ➔ An example: two variables see a change of .01 and 100, respectively. Two totally different levels of change
- ➔ But if the standard deviation for those variables are .01 and 100, then it's really the same level of change

Z-Scores

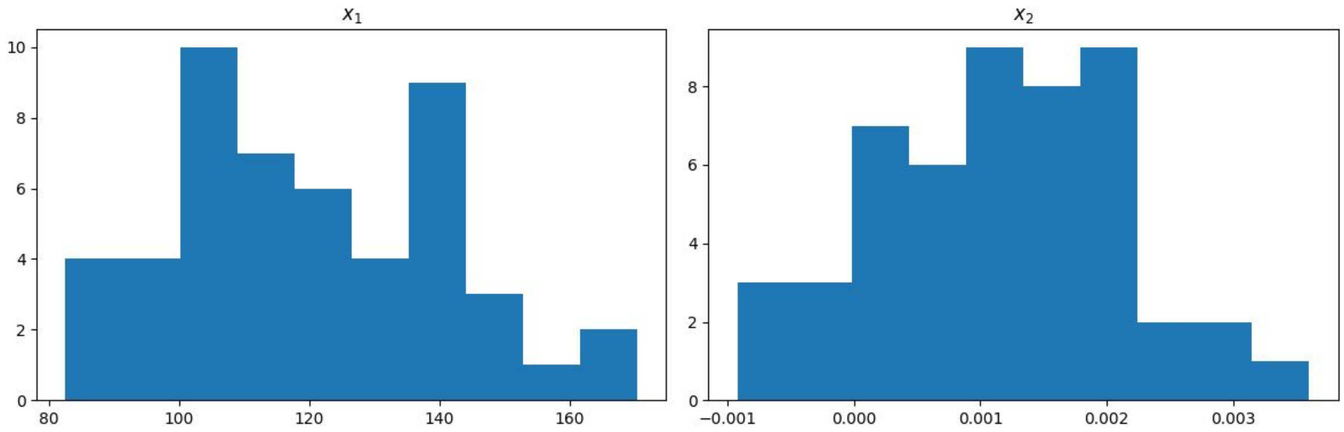
For data observations X_1, ... X_N, the Z-Scores are given by

$$Z_i = \frac{X_i - \text{Avg of } X_1, \ldots, X_N}{\text{Std Dev of } X_1, \ldots, X_N}$$

The mechanism is to de-mean the data and then scale it so it has Std Dev = 1

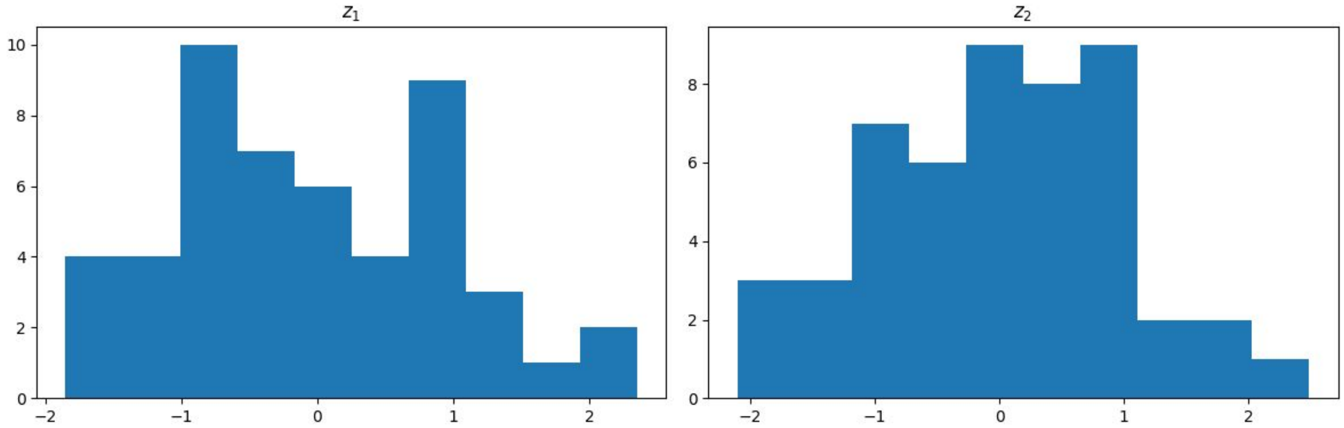
Z-Scores

Non-standardized data:



Z-Scores

Standardized data:



Histograms have the same shape: data hasn't fundamentally change

Just shifted and scaled

Z-Scores

Okay, so now everything is on the same scale

➔ 1.1 means the same thing for each factor

Coefficients have meaning: they now indicate a relative importance

And now we can just use Dean Oliver’s suggested weightings

Dean Oliver's Four Factor Model

But wait, one more issue: there are 8 values for the team and its opposition

We take the raw components, compute the difference, and then compute Z-scores

$$\text{Shooting Factor} = Z(eFG\% - eFG\%_{\text{Opp}})$$
$$\text{Turnover Factor} = Z(\text{Turnover}\% - \text{Turnover}\%_{\text{Opp}})$$
$$\text{Rebounding Factor} = Z(OREB\% - OREB\%_{\text{Opp}})$$
$$\text{Free Throw Factor} = Z(\text{FT Rate} - \text{FT Rate}_{\text{Opp}})$$

Dean Oliver's Four Factor Model

The final model

$$\begin{aligned}\text{Team Performance} = {} & .4 \times Z(eFG\% - eFG\%_{\text{Opp}}) \\ & - .25 \times Z(\text{Turnover\%} - \text{Turnover\%}_{\text{Opp}}) \\ & + .2 \times Z(OREB\% - OREB\%_{\text{Opp}}) \\ & + .15 \times Z(\text{FT Rate} - \text{FT Rate}_{\text{Opp}})\end{aligned}$$

Turnovers are bad hence the negative sign

Dean Oliver's Four Factor Model

So where did those weightings come from?

Honestly, I don't know and I haven't seen anyone offer an explanation

However, I have seen evidence that shooting should be weighted more heavily

We will revisit this when doing statistical inference

Anyhow, let's look at the data and how this all works

Player Efficiency Rating

Player Efficiency Rating

Player performance is about efficient production using possessions

Our goal should be to rate players comprehensively

We know how to evaluate their scoring efficiency

What about the rest of the play?

Player Efficiency Rating

Enter PER: Player Efficiency Rating

- ➔ Developed by John Hollinger in the late 90s
- ➔ Based entirely on well known box score stats: FGs, FTs, Steals, Rebs, etc
- ➔ What PER tries to do:

A comprehensive, model-based rating that counts positive and negative contributions and rates them on a per possession basis.

Player Efficiency Rating

$$uPER = \frac{1}{min} \times \Big($$

$$\begin{aligned}
&\text{Three Pointers Made} \\
&+ \text{Contributions from Assists} \\
&+ \text{Contributions from FGs} \\
&+ \text{Contributions from FTs} \\
&- \text{Contributions from TOs} \\
&- \text{Contributions from Missed FGs} \\
&- \text{Contributions from Missed FTs} \\
&+ \text{Contributions from Def Rebounds} \\
&+ \text{Contributions from Off Rebounds} \\
&+ \text{Contributions from Steals} \\
&+ \text{Contributions from Blocks} \\
&- \text{Contributions from Fouls}
\end{aligned}$$

$$\Big)$$

Player Efficiency Rating

- ➔ Raw *uPER* is a per-minute efficiency calculation
- ➔ We compute the final PER by first adjusting for the pace of play of teams
- ➔ And finally adjusting to league average

  Performance is put relative to the league average

Player Efficiency Rating

We first need to understand and deconstruct PER

Then we can get into the pros and cons and finding potentially better approaches

Let's study PER hands-on in a lab...