JOURNAL OF APPLIED STATISTICS
2021, VOL. 48, NO. 16, 3116–3129
https://doi.org/10.1080/02664763.2020.1795819

Taylor & Francis
Taylor & Francis Group

Identifying key factors in momentum in basketball games

Tao Chen[a,b,c], Qingliang Fan [d*], Kai Liu[b,e] and Lingshan Le[f]

[a]Department of Economics, University of Waterloo, Ontario, Canada; [b]Big Data Research Lab, University of Waterloo, Ontario, Canada; [c]Senior Research Fellowship of Labor and Worklife Program, Harvard University, Cambridge, MA, USA; [d]Department of Economics, The Chinese University of Hong Kong, Hong Kong, People's Republic of China; [e]School of Mathematical and Computational Sciences, University of Prince Edward Island, Prince Edward Island, Canada; [f]School of Economics, Xiamen University, Xiamen, People's Republic of China

ABSTRACT

Momentum as elaborated under a recent novel definition has been shown quantitatively to have a significant impact on basketball game outcomes. This paper makes two contributions to the analytical literature on sports momentum: (1) two aspects of the new definition are operationalized so that its practicality becomes evident; and (2) through a dimension-reduction technique (elastic net), key factors associated with momentum are identified. Both technical variables such as field goals, assists, rebounds, etc. and environmental variables such as the spectator attendance rate and player salary dispersion are considered, and the potential for useful real-time analyzes is illustrated.

ARTICLE HISTORY
Received 18 July 2019
Accepted 5 July 2020

KEYWORDS
Basketball; momentum; elastic net; attendance rate; Gini coefficient

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
62-07; 62P99

1. Introduction

Professional sports have always been a fiercely competitive industry on and off the playing field. Obtaining an upper hand can have critical effects, which usually lead to individual success and team victories [13]. To explore every detailed aspect of competition, statistical models are widely utilized to enhance player performance by learning athletes' tendencies and predicting what they will do under any given circumstances [2,3,14].

Momentum study is one of the most challenging topics in sports analytics because people can certainly feel the 'hot trend' when it is happening, but articulating it is nevertheless difficult. Sports psychologists view momentum as an intrinsic motivation that affects an individual's psychological and physical performance [8]; however, this approach does not offer researchers empirical tools when they need to analyze the momentum phenomenon quantitatively.

Mace *et al.* [12] are pioneers who brought data mining to sports momentum analysis and categorized momentum triggers into three types: reinforcers, adversities and response to adversities. Arkes and Martinez [1] studied whether there is a momentum

CONTACT Qingliang Fan michaelqfan@gmail.com Department of Economics, The Chinese University of Hong Kong, 10/F, Esther Lee Building, Shatin, N.T., Hong Kong, People's Republic of China
*Part of the research was done while Fan was a faculty member at Xiamen University.

effect in the National Basketball Association (NBA) by examining how success over the past few games affects the probability of winning the next game. Focusing on key plays such as fourth down conversions/stops, turnovers and scores allowed, Fry and Shukairy [6] analyzed momentum in the National Football League (NFL). Lehman and Hahn [11] examined how momentum shapes organizational risk-taking across quarters of NFL games. Kniffin and Mihalek [9] used hockey game data to study cross-game momentum by building a probabilistic model to verify that in hockey games, leads from the first game do not imply wins in the second game of a two-game series in a statistically significant way.

This paper adopts the recently proposed definition of momentum from Chen and Fan [4, CF hereafter]; we offer an intuitive recapitulation of it before we proceed. Prior to CF, an influential work by Stern [16] had modeled the score difference process through Brownian motion, which is known for its independent increments. CF's perspective was as follows: momentum should be a representation of something 'larger or quicker' than expected changes in the score difference process. Appendix 3 in CF verified that in the presence of momentum events, Brownian motion is not fast enough to capture the score difference changes. This motivated CF to define momentum in terms of the score difference changes above a prespecified threshold.

As a direct extension of CF, the current study makes two contributions: (1) two aspects (explosiveness and duration; see the formal definition in the following Section 2.1) of their momentum definition are operationalized so that its practicality becomes evident; and (2) through a dimension-reduction technique (elastic net, [18]), key factors associated with momentum are identified. The first contribution adds an extra tool to the empirical researchers' toolkit, and the second contribution provides some insights for practitioners, including coaches, team managers and sports data analysts.

The remainder of the paper is organized as follows: Section 2 presents the definition of the variables and our statistical model. Section 3 introduces the data set, which we obtain from play-by-play NBA game logs. The empirical results are collected and discussed in Section 4. Section 5 concludes the paper.

2. Variable definition and statistical model

We employ regression models to explore what happens during periods of momentum from the perspective of measurable game-related statistics. This is the first formal application of the CF momentum definition in a regression model setting; therefore, a detailed explanation of variable construction is necessary. This is followed by detailed algorithms to compute momentum in a game.

2.1. Variable definition

Regression models have both a response variable and a set of regressors. We start with the response variable, the 'size' of momentum, and then list the regressors, which include both game log-type and environmental variables.

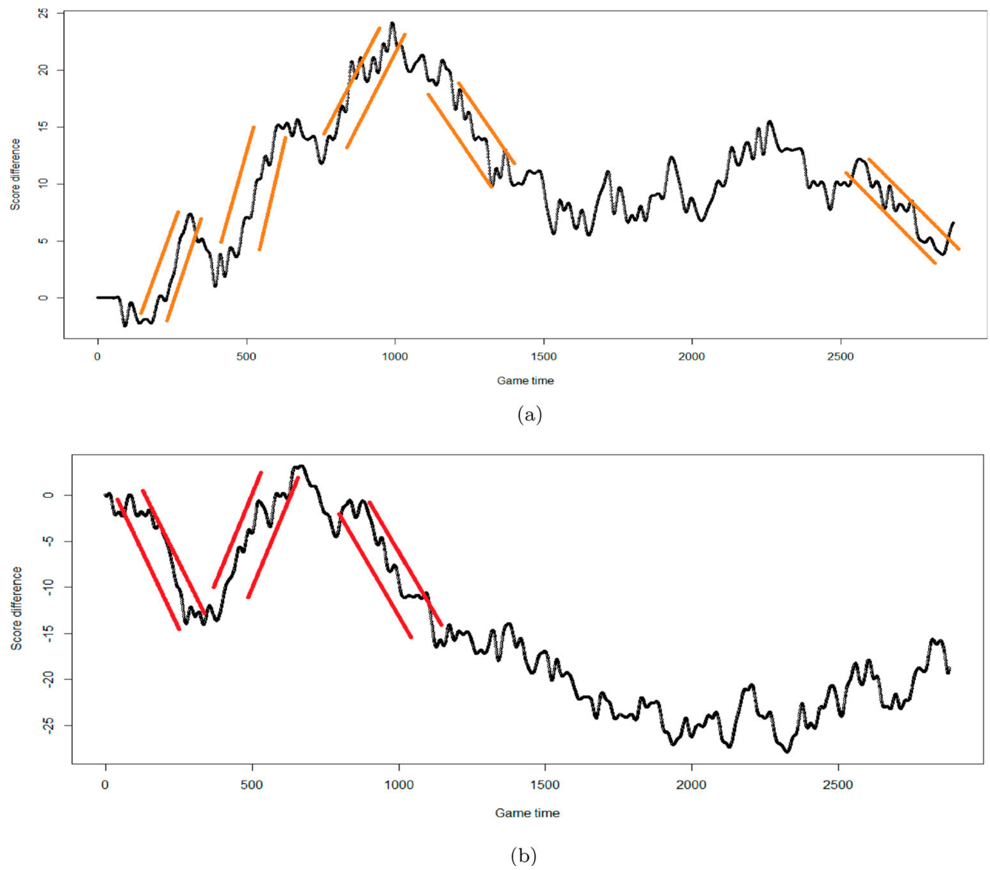

Figure 1. Examples of CF's momentum. (a) Game between WAS and PHI on Nov 16, 2016 and (b) Game between DAL and HOU on Dec 27, 2016.

2.1.1. Response variable

Recall that the momentum of the home team in CF is defined as

$$M(s,t,\gamma) = \begin{cases} y(t+s) - y(t) & \text{if } y(t+s) - y(t) > \gamma \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where $y(t)$ represents the score difference between the home and visiting team at time t, s denotes an increment of game time and γ represents the threshold value of momentum. If we set the mathematical equation aside, the intuition is quite simple: if one team outscores its opponent by a large margin in a relatively short period, this is defined as momentum of size γ (a positive number) over time s. To be flexible, the CF model allows researchers to choose the values of γ and s, such that the smaller s is and the greater γ is, the stronger the momentum will be. In the context of a real basketball game, CF recommend using an s of at least 90 seconds but less than 360 seconds (half a quarter in an NBA game). Similarly, the momentum of the visiting team is

$$M(s,t,\gamma) = \begin{cases} y(t+s) - y(t) & \text{if } y(t+s) - y(t) < -\gamma \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Figure 1 illustrates the selected games' progress, where we can observe momentum by CF's definition. Graph-(a) shows the game between the Washington Wizards (WAS) and Philadelphia 76ers (PHI) on Nov 16, 2016, and Graph-(b) shows the game between the Dallas Mavericks (DAL) and Houston Rockets (HOU) on Dec 27, 2016. The vertical axis is the score difference between the home and away teams. The horizontal axis is time (in a regular NBA game, there are 4 quarters with 12 minutes each or 2880 seconds in total). It is common to use either the standardized time [0, 1] or the original high-frequency time measurement (in seconds), with no change to the qualitative results.

From these two graphs, we can observe the two key features of momentum: (1) momentum occurs in the 'steep' part of the graph, which coincides with large score differences in a relatively short time; (2) the duration of each momentum episode can vary, with some lasting more and some lasting less time. These intriguing features motivate our studies on (1) what key factors are associated with the explosiveness and duration of the momentum and (2) what are the quantitative contributions of the key factors to the two aspects of the momentum. To facilitate the investigation of those questions, we use a standardized version of CF's momentum defined as follows. We start with the standardized home team's momentum:

$$M\left(s' \in [t, s], \mu\right) := \frac{y(s') - y(t)}{s' - t} \geq \mu, \tag{3}$$

where the positive real number μ is a slope threshold. Using slopes to capture the changes provides the geometric perspective and is compatible with the momentum definition of Equation (1). This naturally leads us to focus on two aspects of this slope threshold:

- From the collection of moving fixed-time intervals in $[t, s]$, what are the largest possible values of μ? We call this maxima the *explosiveness* of momentum or explosiveness for short.
- For a given μ, how far does s' extend so that within $[t, s']$, all the slopes do not fall below μ? We call this the *duration* of momentum or duration for short.

Similarly, the standardized momentum of the visiting teams is defined as:

$$M\left(s' \in [t, s], \mu\right) := \frac{y(s') - y(t)}{s' - t} \leq -\mu.$$

Following CF, we set $|\mu| = 0.06$, $s = t + 360$ and $s' - t \geq 90$. We set the lower bound of momentum duration to 90 seconds, which is approximately 6 possessions,[1] mainly to reduce noise in the analysis. The reason we do not consider a time span longer than 6 minutes is that momentum episodes longer than 6 minutes are rarely found in our data, and thus setting a longer time span would not have much practical meaning. A detailed description of the momentum calculation algorithm based on explosiveness is as follows:

Algorithm 1 Momentum Calculation Algorithm Based on Explosiveness

Input: $Y = \{\boldsymbol{y_k} | k = 1, 2, \ldots, K\}$, where K is the total number of games and y_k is the data for the n^{th} game.

Output: momentum dataset: $\boldsymbol{d} = \{d_1, d_2, \ldots, d_N\}$, where N is the total number of momentum episodes, i.e. the number of response variables.

```
for k = 1 to K do
    Let y_k = {t_i | i = 1, 2, ..., I}, where t_i represents the starting time of the i^th turning point, 0 ≤ t_i ≤ 2880 (seconds), and I is the total number of turning points in the k^th game.
    for i = 1 to I do
        Find all turning points s'_i in the interval [t_i + 90, t_i + 360] and determine whether the explosiveness from t_i to s'_i is greater than 0.06.
        if |explosiveness|_(t_i, s'_i) ≥ 0.06 then
            construct the momentum data, i.e. d_k = {game = k, start = t_i, end = s'_i, explosiveness}.
        else
            repeat
    if the explosiveness episodes overlap in the data then
        remove the episodes with smaller values and then update d_k.
```

Next, we present a detailed description of the momentum calculation algorithm based on duration:

Algorithm 2 Momentum Calculation Algorithm Based on Duration

Input: $Y = \{\boldsymbol{y_k} | k = 1, 2, \ldots, K\}$, where K is the total number of games and y_n is the data for the k^{th} game.

Output: momentum dataset: $\boldsymbol{d} = \{d_1, d_2, \ldots, d_N\}$, where N is the total number of momentum episodes.

```
for k = 1 to K do
    Let y_k = {t_i | i = 1, 2, ..., I}, where t_i represents the starting time of the i^th turning point, 0 ≤ t_i ≤ 2880 (seconds), and I is the total number of turning points in the k^th game.
    for i = 1 to I do
        Find all turning points s'_i in the interval [t_i + 90, t_i + 360] and determine whether the explosiveness from t_i to t'_i is greater than 0.06.
        if |explosiveness|_(t_i, s'_i) ≥ 0.06 then
            construct the momentum data, i.e. d_k = {game = k, start = t_i, end = s'_i, duration = s'_i − t_i}.
        else
            repeat
    if the episode durations overlap in the data then
        remove the episodes with smaller values and then update d_k.
```

Remark 1: In the above two algorithms, the search is in chronological order on a one-second grid. Each $t_i \in [0, 2880]$ is considered the potential turning point, that is, the start or end of a momentum episode. The first turning point is the origin; we fix this time point, and then we keep searching for the next turning point within the range of 90 to 360 seconds from the first turning point. Once a momentum episode is found, then the corresponding t_i is defined as the momentum episode starting point. The search for the next momentum episodes starts at the end of the previous momentum. If a momentum episode is not found within the range of 90 to 360 seconds from the given turning point, we add one unit to t_i and repeat the same procedure using $t_i + 1$ as the potential turning point.

2.1.2. Explanatory variables

In principle, we should consider all measurable game-related variables in investigating what is going on during momentum episodes. Specifically, we include the following regressors in the model: $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ is a vector of basketball game-related variables, including (for both the home and visiting teams) 20-second timeouts, assists, blocks, bonuses, defensive/offensive fouls, defensive/offensive rebounds, flagrant fouls, full timeouts, free throws made/two-point field goals made/three-point field goals made, field goals missed, free throws missed, steals, turnovers, the attendance rate (AR) and the Gini coefficient. The model is parameterized by a p-dimensional vector $\boldsymbol{\beta}$. Here, p is 37, with the first 34 being technical variables, virtually all at the player level, obtained directly from game logs. Since the elastic net, to be presented in the next subsection, is capable of screening out statistically nonsignificant variables while taking into account the correlations between the regressors, we allow ourselves to include all candidate variables into it without subjective decisions *a priori*.

We call the game-log variables such as assists, blocks, field goals, etc., technical variables, and the AR and Gini coefficient environmental variables, which are not directly determined by the players. It is conceivable that players' performance rides on the emotion of the crowd; hence, we use the AR, which is the percentage of occupied seats for home teams only, to represent the audience effect. The results from both Smith and Groetzinger [15] and La [10] justify the inclusion of the AR.

The Gini coefficient[2] measures salary dispersion within a team. Although we are unaware of any published results discussing its impact on game outcomes, it will be interesting to find out whether teams led by stars (high-paid players) or more balanced teams tend to generate more momentum.

2.2. Statistical model

The most widely used ordinary least squares (OLS) estimator minimizes the total sum of squares (SST). However, when there are many explanatory variables, OLS leads to overfitting and is weak in interoperability and prediction. To overcome this issue, a penalized OLS with additional constraints is proposed. For example, ridge regression [7] obtains the coefficients by minimizing the SST subject to a size restriction on the L_2 norm of the estimated $\boldsymbol{\beta}$. Although ridge regression is superior to traditional OLS in predictive performance, it is silent on screening out irrelevant variables. LASSO, a penalized least squares estimator invented by Tibshirani [17], serves as a good variable selection mechanism by applying

L_1 penalty to $\boldsymbol{\beta}$; however, this estimator is not able to handle correlation problems among variables.

The elastic net model combines the advantages of both ridge regression and LASSO. Therefore, it can select important variables while dealing with multicollinearity between variables. Suppose $\boldsymbol{d} = (d_1, \dots, d_N)$ is the response variable and $\boldsymbol{X} = (\boldsymbol{x}_1^\top, \dots, \boldsymbol{x}_N^\top)^\top$ is the $N \times p$ design matrix. Note that we normalize all the $\boldsymbol{x}_i$s (divided by their own standard deviation) so that a direct comparison of their coefficients becomes possible.

Our objective function is:

$$L(\lambda_1, \lambda_2, \boldsymbol{\beta}) = \|\boldsymbol{d} - \boldsymbol{X}\boldsymbol{\beta}^\top\|_2 + \lambda_2\|\boldsymbol{\beta}\|_2 + \lambda_1\|\boldsymbol{\beta}\|_1,$$

where $\|\boldsymbol{a}\|_2 = \sqrt{\sum_{j=1}^{J}(\boldsymbol{a}^{(j)})^2}$ and $\|\boldsymbol{a}\|_1 = \sum_{j=1}^{J}|\boldsymbol{a}^{(j)}|$ for any J-dimensional vector $\boldsymbol{a}$, with $\boldsymbol{a}^{(j)}$ being the jth coordinator; specifically, we consider $\boldsymbol{a} = \boldsymbol{d} - \boldsymbol{X}\boldsymbol{\beta}^\top$ or $\boldsymbol{a} = \boldsymbol{\beta}$. λ_1 and λ_2 are the tuning parameters. The estimated value of the elastic net $\hat{\boldsymbol{\beta}}$ is obtained by the following minimization:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, L(\lambda_1, \lambda_2, \boldsymbol{\beta}).$$

Let $\alpha = \lambda_1/(\lambda_1 + \lambda_2)$, $\lambda = \lambda_1 + \lambda_2$; then, this minimization problem can be rewritten as

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \|\boldsymbol{d} - \boldsymbol{X}\boldsymbol{\beta}^\top\|_2 + \lambda\left[(1-\alpha)\|\boldsymbol{\beta}\|_2 + \alpha\|\boldsymbol{\beta}\|_1\right]. \tag{4}$$

When $\alpha = 0$, the elastic net becomes a simple ridge regression, and when $\alpha = 1$, the elastic net reduces to LASSO. In fact, both ridge regression and LASSO are special cases of the elastic net. Our estimation scheme follows the recipe by Friedman *et al.* [5]: first, we fix α and use two-thirds of the data to compute $\hat{\boldsymbol{\beta}}(\alpha)$ defined in Equation (4) and the corresponding $\lambda(\alpha)$ through cross-validation. For the remaining third, we calculate $\mathrm{SST}(\alpha)$ based on the first stage $(\hat{\boldsymbol{\beta}}(\alpha), \lambda(\alpha))$. The optimal α, denoted by α^*, minimizes $\mathrm{SST}(\alpha)$ by sweeping through 0 to 1 with a step size 0.01. Then, an updated $\hat{\boldsymbol{\beta}}(\alpha^*)$ is obtained by using the full sample with (α, λ) fixed at $(\alpha^*, \lambda(\alpha^*))$. Next, a simple t-test is conducted to screen out nonsignificant variables from the model. Finally, OLS is applied to the remaining variables.[3]

3. Data description

In the NBA, each team plays 82 games in the regular season every year. Therefore, the 30 teams in the league generate 1230 games annually. We scraped data from the ESPN gamelog website for all NBA regular season games covering the 2005–06 season to the 2016–17 season. We excluded games with overtime and converted the rest to high-frequency data in 2880 seconds. Game ARs and players' salary data are fetched from the ESPN website as well.

Based on the total 12,845 games, 12,263 of them have momentum episodes. Among those, 38,330 and 37,929 momentum episodes were identified following the explosiveness and duration definitions, respectively. The summary charts of the home and visiting teams are presented in the Figure 2:

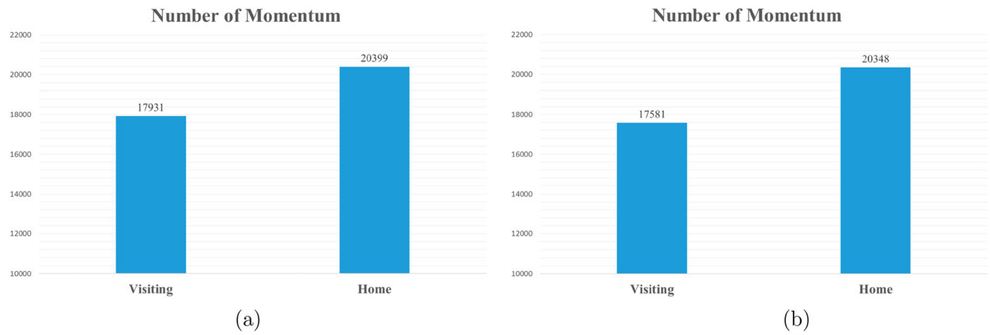

Figure 2. The number of momentum episodes with two different definitions. (a) Explosiveness definition and (b) Duration definition.

Table 1. Number of wins – explosiveness definition.

	number = 0[a]		number = 1		number = 2		number = 3		number > 3	
Number of wins	Home	Visiting	Home	Visiting	Home	Visiting	Home	Visiting	Home	Visiting
number = 0[b]	330	252	335	480	158	437	30	215	8	94
number = 1	702	272	919	668	454	603	125	360	15	163
number = 2	700	115	942	333	568	375	167	229	33	145
number = 3	342	30	571	104	364	105	125	73	30	53
number >3	184	7	253	19	202	28	83	16	18	11

[a]This row shows the number of momentum episodes of the visiting team in a game.[b]This column shows the number of momentum episodes of the home team in a game.

Table 2. Number of wins – duration definition.

	number = 0[a]		number = 1		number = 2		number = 3		number > 3	
Number of wins	Home	Visiting	Home	Visiting	Home	Visiting	Home	Visiting	Home	Visiting
number = 0[b]	330	252	333	480	159	445	31	204	7	83
number = 1	716	275	929	680	456	609	116	353	14	148
number = 2	707	119	963	358	563	380	161	229	21	127
number = 3	342	30	571	107	368	110	117	62	26	52
number > 3	184	8	258	19	194	27	72	17	17	8

[a]This row shows the number of momentum episodes of the visiting team in a game.[b]This column shows the number of momentum episodes of the home team in a game.

Table 1 compares the overall wins between home and visiting teams based on the different combinations of momentum episodes that each team has. For example, when the number of momentum episodes is 0 for the home team and 1 for the visiting team, we observed 335 wins for home teams and 480 for away teams. It becomes obvious that the larger the number of momentum episodes is, the higher the probability of winning. When the momentum values are the same for both the home and visiting teams, the home team has a higher probability of winning, which implies a possible home court advantage in NBA games.

Table 2 is the counterpart of Table 1 with the duration definition.

As an illustration, the Table 3 lists the number of momentum episodes generated in the regular seasons for all NBA teams from 2005 to 2017 under the explosiveness definition:

Table 3. Number of momentum episodes per game.

Team	Average	Std	Min	Max
ATL	1.4643	1.0679	0	5
BRK	1.6157	1.1008	0	5
BOS	1.4058	1.0639	0	7
CHA	1.6204	1.1262	0	6
CHI	1.4133	1.0722	0	5
CLE	1.5165	1.1211	0	7
DAL	1.4106	1.0539	0	6
DEN	1.6311	1.1922	0	8
DET	1.4753	1.1037	0	5
GSW	1.7033	1.1821	0	6
HOU	1.4947	1.1211	0	5
IND	1.4981	1.0973	0	7
LAC	1.5000	1.0364	0	4
LAL	1.5376	1.1646	0	6
MEM	1.4808	1.1244	0	6
MIA	1.3717	1.0904	0	7
MIL	1.5740	1.1000	0	7
MIN	1.6620	1.1621	0	6
NOP	1.5132	1.0757	0	5
NYK	1.5609	1.1441	0	8
OKC	1.5154	1.1049	0	6
ORL	1.4494	1.0922	0	5
PHI	1.6326	1.1761	0	5
PHO	1.6033	1.1189	0	6
POR	1.4282	1.0554	0	5
SAC	1.7766	1.1594	0	6
SAS	1.2387	1.0575	0	6
TOR	1.5523	1.1250	0	8
UTA	1.4050	1.0796	0	5
WAS	1.5512	1.1275	0	6

4. What happens during momentum episodes?

Now we are ready to explore what the key factors are behind a basketball momentum episode. We study this question using two response variables: explosiveness and duration of the momentum episode. The explanatory variables are the same for the two response variables.

4.1. *Explosiveness*

The α^* obtained from Equation (4) is 0.04, and the selected regressors are marked with checkmarks in Table 4. Roughly one-third of the regressors are selected. While we do not expect missing free throws or committing flagrant fouls to contribute to momentum, it is insightful to see that both short and long timeouts, blocks, field goal misses, offensive rebounds and steals are not momentum triggers, either.

After the variable selection, let us quantify the contributions from the remaining regressors. This postselection OLS helps reduce the estimation bias of the elastic net in one step. The estimated coefficients are listed in Table 5 together with t-test statistics. We see that all of the coefficients are statistically significant, and the signs of all the estimates are consistent with our intuition (positive and negative signs mean that the variable is a positive factor for home and away team momentum, respectively).

Table 4. Variable selection.

Variable name	Home team	Visiting team
Sec_timeout	–	–
Full_timeout	–	–
Assist	✓	✓
Blocks	–	–
Penalty	–	–
Defensive_foul	✓	–
Defensive_rebound	✓	✓
Flagrant_foul	–	–
Makes_free_throw	–	✓
Makes_three_point	✓	✓
Makes_two_point	–	✓
Miss	–	–
Misses_free_throw	–	–
Offensive_foul	–	–
Offensive_rebound	–	–
Steals	–	–
Turnover	✓	–
AR	✓	N/A
Gini	✓	✓

Note: '–' denotes the variables that were not selected by the elastic net.

Table 5. Estimated coefficients.

Variable name	Estimate	*t*-value
Home_assists	0.0002	2.0
Visiting_assist	−0.0003	−3.4
Home_defensive_foul	−0.0003	−3.8
Home_defensive_rebound	0.0004	5.5
Visiting_defensive_rebound	−0.0012	−16.2
Visiting_makes_free_throw	−0.0011	−12.7
Home_makes_three_point	0.0010	16.0
Visiting_makes_three_point	−0.0033	−40.2
Visiting_makes_two_point	−0.0017	−19.7
Home_turnover	−0.0011	−17.2
AR	0.0035	77.0
Home_gini	0.0027	55.5
Visiting_gini	−0.0032	−71.3

When we look at the magnitude of estimates, an immediate finding is that environmental variables play a more important role overall than many of the technical variables do. The two Gini coefficients combined tell us that a roster of higher-paid star players and a lower-paid supporting cast is more likely to initiate momentum than is a more balanced team (salary-wise). The AR is also a crucial factor, the magnitude of which is larger than that of any other variable. It is well known in sports that home crowds can be a key factor in the game. Another possible reason for the large crowd attendance effect is that fans are more enthusiastic in their support of great teams, which tend to have more momentum. This result shows us that though the duty of a franchise's general manager (GM) is mostly off-court,[4] its impact on the court surpasses everything else in terms of game momentum.

With regard to the technical regressors, we can basically group them into three tiers according to their marginal contribution to momentum. The regressor standing out is three-point field goals from the visiting team, which is comparable to the AR. This finding

Table 6. Variable selection.

Variable name	Home team	Visiting team
Sec_timeout	–	–
Full_timeout	–	–
Assist	✓	✓
Blocks	–	–
Penalty	–	–
Defensive_foul	✓	✓
Defensive_rebound	✓	✓
Flagrant_foul	–	–
Makes_free_throw	✓	✓
Makes_three_point	✓	✓
Makes_two_point	✓	✓
Miss	✓	✓
Misses_free_throw	–	✓
Offensive_foul	–	–
Offensive_rebound	✓	✓
Steals	✓	✓
Turnover	✓	✓
AR	✓	N/A
Gini	✓	✓

Note: '–' denotes the variables that were not selected by the elastic net.

can be a very valuable piece of information for coaches. The second tier includes home-team three-point field goals, defensive rebounds and two-point field goals from the visiting team and home-teams turnovers. The three factors with negative coefficients are meaningful signals for the home teams to be extra cautious. The rest, such as assists, home defensive rebounds and fouls, are smaller in magnitude.

4.2. Duration

This subsection focuses on duration as the response variable. In this case, $\alpha^* = 0.22$. Table 6 has more selected regressors than the explosiveness result at first glance. This makes sense because given μ, we try to extend the momentum time window as much as possible, which naturally leads to the inclusion of more regressors. Nonetheless, both short and long timeouts, blocks, penalties, and flagrant and offensive fouls are excluded in the duration model, and all of those regressors were also screened out in the explosiveness case. Therefore, our results are robust to the two practical definitions (aspects) of momentum.

Table 7 is the counterpart of Table 5 in the previous subsection. Similar to Table 5, all the signs of the selected variables are consistent with our intuition. Furthermore, all the selected regressors in Table 5 remain significant in Table 7 with the exception of assists by the visiting team. Given the influence of both two-point and three-point field goals by the visiting team, the deviation regarding assists should not negatively affect the consistency of the results.

A noticeable pattern that distinguishes the results in Tables 5 and 7 is the reversed relationship between the environmental and technical variables. In Table 7, turnovers become the leading factor, followed by rebounds. This is a very interesting finding: for teams that want to extend their momentum, it is not only about scoring more points directly; less eye-catching moments, such as getting the rebounds and minimizing turnovers, are actually more important.

Table 7. Estimated coefficients.

Variable name	Estimate	*t*-value
Home_assists	0.7958	4.1
Visiting_assists	−0.2592	−1.4
Home_defensive_foul	−2.2136	−10.6
Visiting_defensive_foul	1.0804	5.7
Home_defensive_rebound	9.0485	31.8
Visiting_defensive_rebound	−9.3360	−35.2
Home_makes_free_throw	1.9629	9.5
Visiting_makes_free_throw	−0.9709	−4.7
Home_makes_three_point	8.6159	43.6
Visiting_makes_three_point	−7.8863	−40.7
Home_makes_two_point	4.4631	19.3
Visiting_makes_two_point	−2.9691	−13.2
Home_miss	−7.1362	−25.2
Visiting_miss	9.8773	32.3
Visiting_misses_free_throw	1.9195	13.9
Home_offensive_rebound	4.0518	23.0
Visiting_offensive_rebound	−4.7777	−26.6
Home_steals	0.9910	5.7
Visiting_steals	−1.7797	−10.0
Home_turnover	−9.4116	−42.8
Visiting_turnover	11.2140	50.5
AR	3.4395	33.9
Home_gini	2.8300	26.5
Visiting_gini	−3.2886	−33.4

4.3. A brief discussion of real-time analysis

In addition to identifying the important factors for momentum, our approach provides insights into real-time analysis in basketball games. In a real game, the team composition and way of playing can be highly varied; additionally, making adjustments to the opponent team is crucial. The computation time for our model is presented in endnote 2, which implies that real-time application is feasible using a laptop computer. Moreover, it would be even faster if it were run by more powerful computers and if the R-script were replaced with a lower level of machine language such as C++.

From the last two subsections, we find that calling timeouts, whether full or short, neither enhances nor slows down momentum. This result supports the strategy we often see in real games: let the players play it out.[5] What we are more interested in is what the coach should tell the players on the court, either during a timeout or in real time. Let us consider one hypothetical scenario as an illustration, using the results from Table 5. From the perspective of the explosiveness definition, if the visiting team is having a run and has the next possession, the coach of the home team should draw up plays emphasizing the following points: (1) do not let the opponents make an uncontested three-pointer and use the defense scheme to guard the three-point shooters; (2) it is fine to be aggressive on the defense, even at the cost of potential defensive fouls when the visiting team is not in the active shooting mode; and (3) make sure to contest defensive rebounds. This would be the most effective way to stop the opponent's momentum.

5. Conclusion

Adopting the momentum definition proposed by CF, this paper shows how to operationalize their definition through two aspects and identifies momentum triggers. The quantitative results are meaningful for general managers, head coaches and players and are potentially even useful in live games. An interesting extension of this line of research is to assess whether there are key players who are more closely associated with momentum generation than others. We leave this topic to future projects.

Notes

1. According to the play-by-play data, each team averages approximately 100 possessions per game, that is, 15 seconds per possession.
2. For the k^{th} game, $k = 1, \ldots, K$ with K being the total number of games

$$\text{Gini}_k^{\mathbb{I}} = \frac{\sum_{i=1}^{k(\mathbb{I})} \sum_{j=1}^{k(\mathbb{I})} |\text{inc}_i - \text{inc}_j|}{2k(\mathbb{I}) \sum_{i=1}^{k(\mathbb{I})} \text{inc}_i},$$

   where inc_i is the salary of player i, $\mathbb{I} = 1$ or 0 denotes the home or visiting team, respectively, and $k(\mathbb{I})$ means the number of players, including those on the bench.
3. The results in this section are run on a 2.60 GHz dual core CPU and R Software Version 3.6.2. The total run-time is approximately 1.20 minutes.
4. The responsibility of the general manager often includes player transactions, contract negotiations, and hiring and firing of the coaching staff. Some GMs also take responsibility for team operations, including the promotion of games (attracting more fans).
5. This strategy is famously used by some coaches such as Phil Jackson, who was one of the best basketball coaches in the NBA. The coaching strategy has received much criticism and drawn controversy in some games, but that is not the focus here.

Acknowledgments

The authors are grateful to the anonymous referees and Jie Chen (the Editor-in-Chief) for helpful comments, and would like to thank S. Fu, P. Oyer, S. Rigdon and participants in 2018 CAAS conference on high dimensional statistics, 2017 WISE econometrics/statistics group brown bag seminars, etc., for their helpful comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Fan's research, in part, was supported by the National Natural Science Foundation of China Grants 71671149, 71801183 and 71631004 (Key Project) and the Science Foundation of the Ministry of Education of China (18YJC790073).

ORCID

Qingliang Fan http://orcid.org/0000-0001-9560-3311

References

[1] J. Arkes and J. Martinez, *Finally, evidence for a momentum effect in the NBA*, J. Quant. Anal. Sports7 (2011), pp. 12–21.
[2] R.P. Bunker and F. Thabtah, *A machine learning framework for sport result prediction*, Appl. Comput. Inform. 15 (2019), pp. 27–33.

[3] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, *A multiresolution stochastic process model for predicting basketball possession outcomes*, J. Am. Stat. Assoc. 111 (2016), pp. 585–599.
[4] T. Chen and Q. Fan, *A functional data analysis approach to model score difference in professional basketball games*, J. Appl. Stat. 45 (2018), pp. 112–127.
[5] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, J. Stat. Softw. 33 (2008), pp. 1–22.
[6] M. Fry and F. Shukairy, *Searching for momentum in the NFL*, J. Quant. Anal. Sports 8 (2012), pp. 32–47.
[7] A.E. Hoerl and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics 12 (1970), pp. 80–86.
[8] S.E. Iso-Ahola and K. Mobily, *"Psychological momentum": A phenomenon and an empirical (unobtrusive) validation of its influence in a competitive sport tournament*, J. Quant. Anal. Sports 46 (1980), pp. 391–401.
[9] K.M. Kniffin and W. Mihalek, *Within-series momentum in hockey: No returns for running up the score*, Econ. Lett. 122 (2014), pp. 400–402.
[10] V. La, *Home team advantage in the NBA: The effect of fan attendance on performance*, MPRA Paper 54579, University Library of Munich, Germany, 2014.
[11] D.W. Lehman and J. Hahn, *Momentum and organizational risk taking: Evidence from the national football league*, Manage. Sci. 59 (2013), pp. 852–868.
[12] F.C. Mace, J. Lalli, M.C. Shea, and J.A. Nevin, *Behavioral momentum in college basketball*, J. Quant. Anal. Sports. 25 (1992), pp. 657–663.
[13] S. Mellalieu and S. Hanton, *Advances in Applied Sport Psychology: A Review*, 1st ed., New York, NY: Routledge, 2010.
[14] J. Sampaio, T. McGarry, J. Calleja-González, S. Jiménez Sáiz, X. Schelling i del Alcázar, and M. Balciunas, *Exploring game performance in the national basketball association using player tracking data*, PLoS ONE 10 (2015), e0132894. doi:10.1371/journal.pone.0132894.
[15] E.E. Smith and J.D. Groetzinger, *Do fans matter? The effect of attendance on the outcomes of major league baseball games*, J. Quant. Anal. Sports. 6 (2010), Article 4.
[16] H.S. Stern, *A Brownian motion model for the progress of sports scores*, J. Amer. Statist. Assoc. 89 (1994), pp. 1128–1134.
[17] R. Tibshirani, *Regression shrinkage and selection via LASSO*, J. R. Stat. Soc. Ser. B 25 (1996), pp. 267–288.
[18] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, J. R. Stat. Soc. Ser. B67 (2005), pp. 301–320.