The Hot Hand Fallacy Re-examined: New Evidence from the English Premier League

S. Parsons and N. Rohde

No. 2014-03

Series Editor: Dr Nicholas Rohde and Dr Athula Naranpanawa

Copyright © 2014 by the author(s). No part of this paper may be reproduced in any form, or stored in a retrieval system, without prior permission of the author(s).
The Hot Hand Fallacy Re-examined: New Evidence from the English Premier League

S. Parsons* and N. Rohde+

Abstract

Previous studies have illustrated human misperceptions of randomness and resultant suboptimal decision making with reference to the “hot hand” or momentum effect in sport, the notion of serial dependency between outcomes. However, issues of omitted variables bias have plagued many due to a historical reliance on nonparametric techniques or basic regression models. This paper examines across-game and within-game momentum in the English Premier League football competition using fixed effects regressions to control for time-invariant heterogeneity in conjunction with traditional nonparametric techniques. Although the results show evidence of performance reversal following winning streaks, no such evidence is found for streaks of draws or losses or in goal scoring performance within games. This suggests momentum is better suited as a post hoc label of performance than a robust causal phenomenon.

Key words: Panel data; Hot hand; Momentum; Football; Soccer; Group dynamics

JEL Codes: D03; D83; D84; L83

Any errors are the authors’ responsibility.

* Department of Accounting, Finance and Economics, Griffith University, Gold Coast, Queensland, Australia. Email: stephanie.parsons@griffithuni.edu.au.

+ Corresponding author. Department of Accounting, Finance and Economics, Griffith University, Gold Coast, Queensland, Australia. Email: n.rohde@griffith.edu.au. Ph.: + 617 555 28243.
1. Introduction

In recent decades, insights from psychology have challenged economists’
conventional understanding of the human decision-making process. Behavioural
economists recognise this deeper understanding and incorporate behavioural
anomalies such as heuristics and biases into economic models, theories, and policy
development to more realistically describe human actions. The bias that forms the
focus of this study is misperceptions of randomness, the notion that humans falsely
attach causal explanations to random events. For instance, in cases where individuals
mistake small samples as representative of the population’s characteristics (Tversky
& Kahneman, 1971; Wagenaar, 1972), causal explanations may be sought for low
probability events such as sequences of identical outcomes that are falsely believed to
be the result of a non-random process (Gilovich et al., 1985; Oppenheimer & Monin,
2009; Rabin, 2002). Behaviour of this nature has been shown to lead to suboptimal
resource allocations and inefficiencies in many domains, from overreaction to
perceived trends in financial markets (Kahneman & Riepe, 1998; Rabin & Vayanos,
2010) to mispricing in gambling markets (Camerer, 1989). Consumers’
misperceptions of randomness have also been exploited in marketing (Johnson &
Tellis, 2005).

This study examines misperceptions of randomness in the context of momentum
in sporting competitions. As many casual observers would be aware, positive
momentum is the idea that a team’s or individual’s performance temporarily improves
(worsens) relative to their base rate immediately after experiencing a string of
successes (failures), whereas negative momentum assumes a negative association
between past and future performance (Arkes & Martinez, 2011; Koehler & Conley,
2003; Rabin & Vayanos, 2010). Positive momentum in success is synonymous with
the “hot hand” effect, a term predominantly used in relation to basketball to describe
the notion that a player’s probability of scoring with their next shot improves above
their base rate immediately following a run of successful baskets (Avugos et al., 2013;
Gilovich et al., 1985).

Several psychological models of momentum have been established in the
literature. For instance, Taylor and Demick (1994) developed a Multidimensional
Model of Momentum where a positive or negative past event (precipitating event) is
said to influence future performance through numerous channels, including initial
perceptions of past outcomes, physiological arousal, and cognitive changes, while
Vallerand et al. (1988) recognised that athletes and spectators often perceive
momentum differently. In contrast, Cornelius et al. (1997) proposed the Projected
Performance Model that suggests positive momentum is the result (as opposed to the
cause) of changes in athlete performance, and should be considered simply has a post
hoc label or description of performance. Cornelius et al. (1997) also drew on Silva et
al. (1988) to address the phenomenon of negative momentum with reference to two
theoretical constructs – positive inhibition and negative facilitation. Positive
inhibition suggests that a positive event or change in performance tends to be followed by a negative event or change in performance, because athletes become overconfident and less motivated when they have recently achieved successful outcomes (Burke & Houseworth, 1995; Silva et al., 1988). On the other hand, the negative facilitation construct proposes that athletes who experience a poor or disappointing result become more focused and highly motivated, causing their performance to improve in the future (Berger & Pope, 2011; Burke & Houseworth, 1995; Silva et al., 1988; Stanimirovic & Hanrahan, 2004).

There is a large quantity of survey evidence suggesting individuals believe in the existence of momentum in sport, particularly positive momentum (Burns, 2004; Gilovich et al., 1985). The empirical evidence of momentum, however, largely fails to support the theoretical literature and implies these beliefs are mostly spurious. ¹ Momentum among individual athletes has been most frequently examined in the context of basketball, with a famous study by Gilovich et al. (1985) concluding that the performance of selected National Basketball Association (NBA) players closely resembled a random sequence in both overall shooting and free-throw contests. These results were confirmed with a controlled shooting experiment using college players (Gilovich et al., 1985) as well as subsequent studies by other researchers (Avugos et al., 2013; Koehler & Conley, 2003). Minimal empirical evidence of momentum has been found for athlete outcomes in other sports both within and across games (Albright, 1993; Livingston, 2012), except where a repeated motor skill is involved, such as dart throwing (Gilden & Wilson, 1995; Klassen & Magnus, 2001).

Previous empirical studies have encountered omitted variable bias whereby the absence of adequate controls for athlete skill and situational factors potentially alters any apparent momentum effects (Arkes & Martinez, 2011; Leard & Doyle, 2011; Vergin, 2000). At the team level, this bias is also influenced by various theories regarding group dynamics and optimisation. Moral hazard problems, where teammates share credit or blame for sporting outcomes because individual effort decisions are unobservable, have been shown to alter athletes’ performance incentives and lead to free-riding behaviour (Drago & Turnbull, 1988; Groves, 1973; Holmström, 1982). For instance, when one player is performing at a significantly higher level than others in the team, their teammates may have an incentive to shirk as their poor performance will be compensated by above average performance elsewhere. The theory of social facilitation, on the other hand, suggests that observing another teammate’s above average performance can act as a positive externality for other athletes in the team (Aiello & Douthitt, 2001; Strauss, 2002). If these sources of

¹ The false inferences resulting from these mistaken beliefs are known as the hot hand fallacy (for positive momentum) or the gambler’s fallacy (for negative momentum) (Gilovich et al., 1985; Rabin & Vayanos, 2010).
heterogeneity among teams are assumed to be constant over some time frame, fixed effects regressions provide a means by which this major course of omitted variables bias can be eliminated from the analysis and more accurate momentum evidence collected.

This paper adds to the body of empirical literature by examining momentum on two different levels – within games and across games – using data obtained from the English Premier League (EPL) football (soccer) competition. By analysing momentum on multiple levels, this study attains a degree of comprehensiveness not achieved in the existing literature, and allows for the comparison of contextual differences between different levels of momentum, if they exist. This is particularly important in the case of team sports such as football where the team momentum dynamic has been less extensively analysed. Unlike previous studies, drawn outcomes are also explicitly incorporated when examining across-game momentum, introducing an extra dimension to the development of patterns in game-level data (Leard & Doyle, 2011; Vergin, 2000). Additionally, this study addresses methodological limitations identified in the existing literature by relying heavily on fixed effects regressions to control for time-invariant heterogeneity among teams.

The paper proceeds as follows. Section 2 introduces the data used in this investigation and defines the necessary variables. Methodology and results are addressed separately for each of the two levels of momentum analysed, with across-game momentum discussed in Section 3 and within-game momentum covered in Section 4. Section 5 provides a summary of the key results of the paper and offers some concluding remarks.

2. Data

Data were collected pertaining to historical match results in the EPL, a high profile men’s association football league in the United Kingdom, for the 2010-11, 2011-12, and 2012-13 seasons using the online portals Football-Data.co.uk and TransferMarkt.co.uk. Each season, teams play each other twice (home and away) between August and May in an approximately randomly generated sequence given various logistical factors. A total of 380 games are played each season under the current 20 team format, with each game consisting of two 45 minute halves (plus injury time). This produces a data set of 1,140 individual games or 2,280 team-level observations. Teams are awarded three points for a win, one point for a draw, and zero points for a loss, and the team that accumulates the greatest number of points over the season is deemed the champion.

In defining the panel, the cross-sectional units \( i \) reflect the 60 team-season combinations over the three seasons, each with an average of 19 chronologically ordered and randomly selected games or time series units \( t \). The methodology of Carmichael et al. (2000) and Leard and Doyle (2011) is followed whereby one of the two teams in each game is randomly assigned as the “observed” team, the team whose
perspective the match is analysed from. The other is deemed the “opponent”. This organisation ensures that each game only appears once in the data set so as not to introduce any unnecessary data dependency, with each of the 1,140 games reasonably assumed to be independent of others (Leard & Doyle, 2011). The panel is unbalanced due to the random assignment of the observed and opponent teams in each game.

Table 1 lists the variables generated from the raw data together with descriptive statistics including overall means, standard deviations, and minimum and maximum values for each variable. For the purpose of modelling across-game momentum, three binary variables are used where 1 is assigned for a given match outcome (win, draw, or loss) and 0 otherwise. Table 1 suggests the observed team won in 35.53% of the games analysed, drew in 27.37% of the games, and lost in 37.11% of the games. The independent (streak) variables measuring past performance across games reflect the number of consecutive accumulated outcomes (wins, draws, or losses) for the observed team since the most recent alternate outcome or the beginning of the season (whatever was more recent) leading into the current match (Leard & Doyle, 2011). Given the method adopted by Brown and Sauer (1993) and subsequently Leard and Doyle (2011), dummy variables are specified to allow for the potential nonlinear effect of past performance or streaks on the outcome of a match; equal to 1 if the observed team entered the current match with a streak of two or more match outcomes, and 0 otherwise.

Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Across-game Momentum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>=1 if observed team won</td>
<td>1,140</td>
<td>0.3553</td>
<td>0.4788</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>=1 if observed team drew</td>
<td>1,140</td>
<td>0.2737</td>
<td>0.4460</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>=1 if observed team lost</td>
<td>1,140</td>
<td>0.3711</td>
<td>0.4833</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Within-game Momentum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second half goals scored by the observed team</td>
<td>1,140</td>
<td>0.7798</td>
<td>0.9369</td>
<td>0.00</td>
<td>6.00</td>
</tr>
<tr>
<td><strong>Streak Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Across-game Momentum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed team’s win streak</td>
<td>1,111</td>
<td>0.5671</td>
<td>0.9877</td>
<td>0.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Observed team’s draw streak</td>
<td>1,111</td>
<td>0.3564</td>
<td>0.6595</td>
<td>0.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Observed team’s loss streak</td>
<td>1,111</td>
<td>0.6202</td>
<td>1.0619</td>
<td>0.00</td>
<td>8.00</td>
</tr>
<tr>
<td>=1 if observed team’s win streak is at least 2 games</td>
<td>1,079</td>
<td>0.1242</td>
<td>0.3300</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>=1 if observed team’s draw streak is at least 2 games</td>
<td>1,079</td>
<td>0.0658</td>
<td>0.2481</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>=1 if observed team’s loss streak is at least 2 games</td>
<td>1,079</td>
<td>0.1446</td>
<td>0.3518</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Within-game Momentum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First half goals scored by the observed team</td>
<td>1,140</td>
<td>0.6123</td>
<td>0.7921</td>
<td>0.00</td>
<td>4.00</td>
</tr>
</tbody>
</table>
Within-game team success is measured as the number of second half goals scored by the observed team, interpreted as an ordinal or count variable for the purpose of econometric modelling (Dixon & Robinson, 1998). As Table 1 shows, the number of second half goals scored by the observed team ranged from zero to six, with a mean of approximately 0.7798. Although the relative rarity of goals in football means this variable is still subject to some randomness, other measures of success such as possessions or shots on goal are inappropriate as momentum is ultimately concerned with outcomes that are unambiguously positive, neutral, or negative. 2 The independent variables measuring past performance reflect the first half goals scored by the observed team, averaging 0.6123. The number of goals scored by the opponent in the first half and second half are also recorded with mean values of 0.6202 and 0.7877, respectively.

Five control variables are specified for the econometric models to address time-varying heterogeneity between teams, guided by the previous literature where possible. Continuous control variables are used to measure the difference between the average age of the starting eleven players for the observed and opposition teams in each match (as a proxy for competitive experience) (Livingston, 2012), the difference in rest days between the observed and opposition teams since each team played their last EPL match (Arkes & Martinez, 2011; Reed & O'Donoghue, 2005), and the difference between the points accumulated by the observed team and the opponent at the end of the season under analysis excluding the points earned for the match in question to avoid reverse causality issues (Arkes & Martinez, 2011; Berger & Pope, 2011; Livingston, 2012). Dummy control variables are used to record whether teams have been recently promoted to the EPL (that season), with the difference between the value for the observed team and opponent included in the modelling. Finally, given evidence from Carmichael and Thomas (2005), Mizruchi (1985), and Yiannakis et al. 2

---

2 Although own goals can occur in football (where a player scores a goal that is registered against their own team), their frequency is minimal in terms of the total number of goals scored over the season, and in most cases own goals tend to be a reflection of persistent positive performance resulting from a piece of attacking play by the offensive team.
(2006) supporting the notion of home ground advantage in the EPL, a dummy variable is used to indicate whether the observed team was the home team for a given match.

3. Across-game Momentum

3.1 Methodology

The across-game momentum analysis uses two methodological approaches – nonparametric Wald-Wolfowitz runs tests and fixed effects logit models. Both approaches are discussed below.

3.2 Wald-Wolfowitz Runs Tests

Momentum across games is initially examined using the nonparametric Wald-Wolfowitz runs test to determine whether the dichotomous match outcome data follows a random process (Wald & Wolfowitz, 1940; Wolfowitz, 1943). The Wald-Wolfowitz test compares the actual number of runs (consecutive sequences of identical outcomes) in each season to the mean number of runs expected under the assumption that the dichotomous outcomes are independently and randomly drawn from the same distribution. Thus, the null \( H_0 \) and alternative \( H_1 \) hypotheses are

\[ H_0: \text{Sample values follow a random sequence} \]
\[ H_1: \text{Sample values follow a non-random sequence} \]

The test statistic, \( Z \), is asymptotically distributed as a standard normal under the null and defined as

\[ Z = \frac{R - \bar{R}}{s_R} \]  

where \( R \) is the number of observed runs, \( \bar{R} \) is the expected number of runs, and \( s_R \) is the standard deviation of the number of runs (Doane & Seward, 2007). Expressions for \( \bar{R} \) and \( s_R \) in terms of the number of each type of outcome \( (n_1 \text{ and } n_2) \) are presented below (Doane & Seward, 2007)

\[ \bar{R} = \frac{2n_1n_2}{n_1 + n_2} + 1 \]  

\[ s_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} \]
For a given level of significance, \( \alpha \), the Wald-Wolfowitz test rejects the null hypothesis (indicating non-randomness) if

\[
|Z| > Z_{1-\alpha/2}
\]  

(4)

Note this is a two-tailed test. For sufficiently large sample sizes (\( n \geq 20 \)), the test statistic can be compared to a critical value based on the standard normal distribution.

The level of significance, \( \alpha \), and related critical value for each individual hypothesis test controls for the probability of Type I error – rejecting the null hypothesis when it is true. However, multiple testing (multiplicity) can be an issue when a large number of hypotheses are examined, whereby true statistical significance becomes more complex to measure because there is a high probability of Type I error simply due to randomness (Shaffer, 1995; Vergin, 2000). Given the 180 Wald-Wolfowitz tests conducted, the probability of observing at least one significant result merely due to chance is 99.99% assuming a 5% independent significance level for individual tests and 83.62% for a 1% level of significance. There is no universally acceptable solution for dealing with multiple testing, but a simple Bonferroni correction can be made to the significance level by dividing by the number of tests performed (Dunn, 1961; Shaffer, 1995). Thus, adjusted significance levels of \( 2.7778 \times 10^{-4} \) instead of 0.05 and \( 5.5556 \times 10^{-5} \) instead of 0.01 are also considered when analysing the Wald-Wolfowitz test results.

Another way to account for multiplicity in hypothesis testing is to determine whether the calculated test statistics from the Wald-Wolfowitz runs test are distributed as a standard normal using the Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1933). Confirmation of this distributional assumption would imply that rejection or non-rejection of the null hypothesis is appropriate. The null and alternative hypotheses consider a sample of \( n \) observations with the cumulative distribution function \( F \) compared to a known cumulative distribution function \( F_0 \) (the standard normal)

\[
H_0: F(x) = F_0(x)
\]

\[
H_1: F(x) \neq F_0(x)
\]

The test statistic \( D_n \) determines the supremum (sup) of the distances between the empirical distribution function and the standard normal cumulative distribution, assuming independent and identically distributed observations

\[
D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}(x) - F_0(x) \right|
\]  

(5)

where \( \hat{F} \) is the empirical cumulative distribution function defined as
\[
\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x)
\]  

(6)

where \(I(x_i \leq x)\) is the indicator function which equals 1 if \(x_i \leq x\) and 0 otherwise. Note that as this is a nonparametric test, the distribution of \(D_n\) itself is not reliant on the underlying distribution being tested. The associated \(p\)-value can be compared against a given significance level (in this case, 5%) to determine whether the distribution of the test statistic differs significantly from a standard normal distribution.

### 3.3 Fixed Effects Logit Models

In order to fully utilise the available panel data set and control for omitted variable bias ignored by the Wald-Wolfowitz runs tests, panel data regression models are also used to analyse across-game momentum. Due to the binary specification of the dependent variables, the chosen regression model is the fixed effects logit model with parameter estimates obtained using maximum likelihood estimation (MLE) to constrain the predicted probabilities between 0 and 1. The fixed effects estimator is preferred over random effects as it is able to control for unobserved time-invariant heterogeneity over seasons for each team. This represents an improvement over many of the time series approaches of existing literature by eliminating omitted variable bias due to time-invariant factors. In addition, a fixed effects approach permits correlation between the regressors and the time-invariant fixed effects, a less restrictive assumption than the random effects model. Consistent parameter estimates with fixed effects can only be achieved using a logit specification, as the sufficient statistics required for conditional MLE and necessary to overcome the incidental parameters problem (Neyman & Scott, 1948) do not exist for the probit model.

The fixed effects logit model for a generic binary outcome variable \(y_{it}\) (\(i = 1, \ldots, N\) and \(t = 1, \ldots, T\)) can be expressed as follows

\[
\Pr(y_{it} = 1|x_{it}) = \Lambda(\alpha_i + x_{it}^\prime \beta)
\]  

(7)

where \(\Lambda(.)\) is the logistic cumulative distribution function with \(\Lambda(z) = e^z/(1 + e^z)\), \(\alpha_i\) is the team-season fixed effect, and \(x_{it}\) is a \(K \times 1\) vector of explanatory variables with associated parameters \(\beta\). Three models are subsequently estimated from this generic specification, where \(K\) ranges from 3 to 11. The first model considers only the observed team’s streak variables as independent variables, while the remaining two models account for time-varying differences between teams and situational factors in the form of the five control variables defined in Section 2. The third model also includes an allowance for nonlinearities by adding three dummy variables indicating whether a streak is at least two games long.
The assumption of independence imposed by conventional standard errors is no longer appropriate for this panel data set as correlation is expected within clusters or groups of observations. Cluster robust standard errors relax the assumption that observations must be independent and instead allow for intragroup correlation. That is, observations are not independent within groups (clusters), but are independent across groups. For this study, two-way cluster standard errors are necessary to cater for correlation not only within each team-season combination (such as Manchester United in 2011-12 or Arsenal in 2012-13) but also for each unique team (that is, Manchester United over all seasons or Arsenal over all seasons) (Cameron et al., 2011; Thompson, 2011).

4. Results and Discussions

4.1 Wald-Wolfowitz Runs Tests

The results of the 180 Wald-Wolfowitz runs tests performed are provided in Table 2. The null hypothesis of randomness can only be rejected eight times at the 5% significance level – once in the 2012-13 season, six times in the 2011-12 season, and once in the 2010-11 season – with no rejections made at the 1% significance level. Five of the eight rejections indicate that teams experienced more streaks of binary outcomes than expected under the assumption of independence as the calculated test statistic exceeds the positive 5% critical value (1.96). That is, more runs (of shorter length) imply a greater frequency of outcome reversals in line with the theory of negative momentum. The remaining three rejections of the null hypothesis imply a potential positive momentum effect because the calculated test statistic is less than the negative 5% critical value (-1.96). In these three cases, teams experienced fewer runs than expected under the assumption of randomness, implying the potential persistence of similar outcomes consistent with positive momentum.
### Table 2: Wald-Wolfowitz Runs Tests

<table>
<thead>
<tr>
<th>Team</th>
<th>2012-13</th>
<th>2011-12</th>
<th>2010-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arsenal</td>
<td>0.0700</td>
<td>0.1125</td>
<td>0.3228</td>
</tr>
<tr>
<td>Aston Villa</td>
<td>0.5400</td>
<td>1.3548</td>
<td>0.7355</td>
</tr>
<tr>
<td>Birmingham</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blackburn</td>
<td></td>
<td>0.6867</td>
<td>1.4379</td>
</tr>
<tr>
<td>Blackpool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bolton</td>
<td></td>
<td>0.1125</td>
<td>-0.7009</td>
</tr>
<tr>
<td>Chelsea</td>
<td>-0.1777</td>
<td>0.1210</td>
<td>1.4379</td>
</tr>
<tr>
<td>Everton</td>
<td>0.8350</td>
<td>-0.0544</td>
<td>-0.2348</td>
</tr>
<tr>
<td>Fulham</td>
<td>0.1482</td>
<td>0.5400</td>
<td>0.4028</td>
</tr>
<tr>
<td>Liverpool</td>
<td>0.8350</td>
<td>-0.4051</td>
<td>1.5006</td>
</tr>
<tr>
<td>Man City</td>
<td>0.9794</td>
<td>0.1210</td>
<td>0.5674</td>
</tr>
<tr>
<td>Man United</td>
<td>0.5400</td>
<td>0.2351</td>
<td>0.2351</td>
</tr>
<tr>
<td>Newcastle</td>
<td>1.3548</td>
<td>0.6867</td>
<td>1.3157</td>
</tr>
<tr>
<td>Norwich</td>
<td>-0.7425</td>
<td>-0.5964</td>
<td>-0.9505</td>
</tr>
<tr>
<td>QPR</td>
<td>-1.0609</td>
<td>-0.4051</td>
<td>-0.2627</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.0668</td>
<td>-0.7425</td>
<td>-1.8655</td>
</tr>
<tr>
<td>Southampton</td>
<td>-0.7987</td>
<td>-0.2423</td>
<td>-1.0883</td>
</tr>
<tr>
<td>Stoke</td>
<td>-0.7987</td>
<td>-0.0544</td>
<td>-0.5964</td>
</tr>
<tr>
<td>Sunderland</td>
<td>-0.7987</td>
<td>-1.3077</td>
<td>0.7355</td>
</tr>
<tr>
<td>Swansea</td>
<td>0.5504</td>
<td>0.3279</td>
<td>-0.9505</td>
</tr>
<tr>
<td>Tottenham</td>
<td>0.0700</td>
<td>-0.7987</td>
<td>-1.8224</td>
</tr>
<tr>
<td>West Brom</td>
<td>-0.2423</td>
<td>0.8804</td>
<td>-0.9281</td>
</tr>
<tr>
<td>West Ham</td>
<td>2.1325**</td>
<td>-0.3150</td>
<td>1.8477</td>
</tr>
<tr>
<td>Wigan</td>
<td>0.1210</td>
<td>0.5809</td>
<td>1.3368</td>
</tr>
<tr>
<td>Wolves</td>
<td>-1.2541</td>
<td>-0.3150</td>
<td>0.2902</td>
</tr>
</tbody>
</table>

Note: *** denotes significance at 1%, ** denotes significance at 5%, and * denotes significance at 10%.
To cater for multiplicity caused by the large number of hypothesis tests, the Bonferroni correction is applied to reduce the significance levels to $2.7778 \times 10^{-4}$ (from 0.05) and $5.5556 \times 10^{-5}$ (from 0.01). Subsequently, the null hypothesis cannot be rejected in any of the 180 cases. To test for the validity of the joint null hypothesis that the $Z$ scores are asymptotically distributed as a standard normal, Kolmogorov-Smirnov test statistics are produced for each series of $Z$ scores (containing 20 observations) and presented with their associated $p$-values (standard and corrected) in Table 3 below. As all of the $p$-values exceed 0.05, the null hypothesis cannot be rejected at a 5% level of significance and the Wald-Wolfowitz test statistics approximate a standard normal distribution.

Table 3: Kolmogorov-Smirnov Test for Standard Normality

<table>
<thead>
<tr>
<th>Season</th>
<th>Series</th>
<th>Combined Test Statistic ($D_n$)</th>
<th>Standard value</th>
<th>$p$-value</th>
<th>Corrected $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-13</td>
<td>Win/Non-win</td>
<td>0.1186</td>
<td>0.941</td>
<td>0.906</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Draw/Non-draw</td>
<td>0.0756</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loss/Non-loss</td>
<td>0.1170</td>
<td>0.947</td>
<td>0.914</td>
<td></td>
</tr>
<tr>
<td>2011-12</td>
<td>Win/Non-win</td>
<td>0.1816</td>
<td>0.524</td>
<td>0.425</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Draw/Non-draw</td>
<td>0.1026</td>
<td>0.984</td>
<td>0.972</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loss/Non-loss</td>
<td>0.1625</td>
<td>0.666</td>
<td>0.572</td>
<td></td>
</tr>
<tr>
<td>2010-11</td>
<td>Win/Non-win</td>
<td>0.1311</td>
<td>0.882</td>
<td>0.825</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Draw/Non-draw</td>
<td>0.1695</td>
<td>0.613</td>
<td>0.515</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loss/Non-loss</td>
<td>0.1449</td>
<td>0.795</td>
<td>0.717</td>
<td></td>
</tr>
</tbody>
</table>

Note: The standard $p$-values for the Kolmogorov-Smirnov test are based on asymptotic distributions derived by Smirnov (1933), with the corrected $p$-values applying the empirical continuity correction.

Overall, the results of the Wald-Wolfowitz tests suggest there is little difference between the actual number of streaks or runs and the number expected under the assumption of randomness in the EPL. Any rejections of the null hypothesis typically favour negative momentum over positive momentum, but the total number of rejections comprises less than 5% of all hypotheses. Not only does multiple testing imply that these rejections could simply be the result of randomness at the given significance level, the lack of control variables suggests that the observed negative momentum effect could also be the result of situational factors. For instance, a negative momentum effect may be due to the home and away sequence in which the matches are played rather than any systematic change in performance related to momentum. Thus, it is important to consider the fixed effects regressions that control for these confounding influences on performance.

4.2 Fixed Effects Logit Models

Table 4 presents the estimation results for the three fixed effects logit models based on Equation 7. These models control for time-invariant heterogeneity and, in specified cases, time-varying situational factors and team quality. As the coefficients do not directly reflect marginal effects, the exponential of each coefficient is taken to
allow for the interpretation of one unit changes in explanatory variables to occur in terms of odds ratios. Table 4 also reports two-way cluster robust standard errors using code obtained from Jaimovich and Kamuganga (2012).

Evidence for or against momentum can be analysed on the basis of the signs of the coefficients of the explanatory variables and their significance. For example, with the dependent variable indicating wins and non-wins (Model 1), a positive coefficient for the observed team’s win streak variable and negative coefficients for draw and loss streak variables would imply the persistence of similar outcomes over time in line with the theory of positive momentum. The coefficients of these streak variables would be expected to have the opposite signs in the presence of negative momentum indicating the reversal of performance outcomes over time. A more significant coefficient would suggest a more noteworthy momentum effect.
### Table 4: Fixed Effects Logit Models

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Win/Non-win</th>
<th>Draw/Non-draw</th>
<th>Loss/Non-loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Observed team’s win streak</td>
<td>-0.3322***</td>
<td>-0.2959***</td>
<td>-0.4492***</td>
</tr>
<tr>
<td></td>
<td>0.7174</td>
<td>0.7439</td>
<td>0.6381</td>
</tr>
<tr>
<td></td>
<td>(0.0721)</td>
<td>(0.0730)</td>
<td>(0.1782)</td>
</tr>
<tr>
<td>Observed team’s draw streak</td>
<td>0.0887</td>
<td>0.0881</td>
<td>0.0251</td>
</tr>
<tr>
<td></td>
<td>1.0928</td>
<td>1.0921</td>
<td>1.0254</td>
</tr>
<tr>
<td></td>
<td>(0.1345)</td>
<td>(0.1296)</td>
<td>(0.2638)</td>
</tr>
<tr>
<td>Observed team’s loss streak</td>
<td>-0.0514</td>
<td>-0.1210</td>
<td>-0.2853*</td>
</tr>
<tr>
<td></td>
<td>0.9499</td>
<td>0.8860</td>
<td>0.7518</td>
</tr>
<tr>
<td></td>
<td>(0.0888)</td>
<td>(0.0877)</td>
<td>(0.1532)</td>
</tr>
<tr>
<td>=1 if observed team’s win</td>
<td>0.4058</td>
<td></td>
<td>-0.1326</td>
</tr>
<tr>
<td>streak is at least 2 games</td>
<td>1.5005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4084)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>=1 if observed team’s</td>
<td>0.0464</td>
<td></td>
<td>-0.1468</td>
</tr>
<tr>
<td>draw streak is at least 2 games</td>
<td>1.0475</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5688)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>=1 if observed team’s loss</td>
<td>0.4674</td>
<td></td>
<td>-0.2392</td>
</tr>
<tr>
<td>streak is at least 2 games</td>
<td>1.5959</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2416)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control variables?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>1,111</td>
<td>1,109</td>
<td>1,079</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-568.6</td>
<td>-530.9</td>
<td>-510.2</td>
</tr>
<tr>
<td>McFadden (1974) Pseudo-$R^2$</td>
<td>0.0195</td>
<td>0.0818</td>
<td>0.0865</td>
</tr>
</tbody>
</table>
The estimation results in Table 4 suggest there is some evidence of a negative momentum effect, particularly with regards to the reversal of performance following a winning streak. When the dependent variable measures wins and non-wins, the observed team’s winning streak variable has a negative coefficient, indicating that having a longer winning streak upon entering the current match reduces the probability of the observed team winning. This is a reduction in the odds of winning by 0.7174 times in Model 1 (or 71.74%) for a one game increase in the winning streak for the observed team, holding all else constant. The negative momentum effect is slightly smaller in Model 2 (0.7439 times), with a larger effect noted in Model 3 (0.6381 times). These momentum effects are highly significant in all models (at the 1% level in Models 1 and 2 and the 5% level in the Model 3), but with increasing standard errors as more explanatory variables are added. The negative momentum effect in relation to winning streaks is also evident when the dependent variable measures draws and non-draws or losses and non-losses, given the positive effect of winning streaks on the probability of drawing or losing the next match. However, the effect is less significant compared to the models using wins and non-wins as the dependent variable, eventually becoming insignificant in Model 3 as standard errors increase.

The negative momentum effect associated with winning streaks is not consistently evident following streaks of draws and losses. With wins and non-wins as the dependent variable, the positive effect of draw streaks on the probability of winning the current game implies negative momentum, but the negative effect of losing streaks suggests a positive momentum effect. However, these momentum effects are insignificant with the exception of the negative influence of loss streaks on the probability of winning in Model 3 that is statistically significant at the 10% level. With draws and non-draws as the dependent variable, there is evidence of performance reversal following streaks of draws and losses, although only the coefficient of the loss streak variable in Models 1 and 3 is significant (at the 10% level). With losses and non-losses as the dependent variable, the positive coefficients of the draw streak variables also indicate performance reversal, but the positive coefficients of the loss streak variables in Models 2 and 3 imply positive momentum. The coefficients and odds ratios suggest these momentum effects are of small magnitude and not significant at any reasonable significance level. Adding dummy variables in Model 3 to allow for any nonlinear effects of streaks of wins, draws, or losses on the current match outcome produces insignificant coefficients in all model specifications. The coefficients vary in sign depending on the dependent variable used, with no clear mitigating or compounding influence of note. Thus, these results do not fully conform to the theoretical description of either positive or negative momentum, but do imply some inconsistent effect of past performance on future match outcomes.

The finding of a significant but inconsistent negative momentum effect is in contrast to the majority of the existing empirical literature that has predominantly rejected the notion of across-game momentum entirely (Gilovich et al., 1985; Goddard & Asimakopoulos, 2004; Vergin, 2000) or found evidence of a small positive momentum effect (Arkes & Martinez, 2011; Leard & Doyle, 2011). As discussed in Section 1, the phenomenon of negative momentum is best explained by Cornelius et al. (1997) and Silva et al. (1988) with reference
to the theoretical constructs of positive inhibition and negative facilitation. The empirical
evidence of this study supports the positive inhibition construct by suggesting there is a shift
in performance attributes away from optimality following victory in the EPL, but does not
support the negative facilitation construct as poor performance does not tend to be
systematically related with future success.

5. Within-game Momentum

5.1 Methodology

Similar to the across-game momentum analysis, fixed effects regressions are also used to
determine the direction and significance of any momentum effects within EPL games. The
dependent variable in these models, the number of goals scored by the observed team in the
second half of a match, is a form of count data and as such Poisson and negative binomial
variations of the fixed effects regression model are used to cater for different dispersion
assumptions. Maher (1982) noted that early attempts to model goal scoring patterns in
football favoured the negative binomial model due to over dispersion in the data, though the
Poisson model has been shown to be a reasonably strong fit for English Premier League goal
scoring data in more recent studies (Dixon & Robinson, 1998; Heuer et al., 2010). The fixed
effects Poisson model is also popular due to its consistency under weaker distributional
assumptions than the negative binomial model (Cameron & Trivedi, 2005).

For a given a generic count variable $z_{it}$ ($i = 1, ..., N$ and $t = 1, ..., T$), the fixed effects
Poisson model imposes a Poisson distribution such that

$$z_{it} \sim \text{Poisson}[\exp(\alpha_i + w_{it}'\phi)]$$  \hspace{1cm} (8)

where $\alpha_i$ is the team-season fixed effect, and $w_{it}$ is a $K \times 1$ vector of explanatory
variables with associated parameters $\phi$. The expected count of $z_{it}$, defined as $\mu_{it}$, is linked to
the strictly exogenous explanatory variables by way of the log-linear function

$$\ln \mu_{it} = \alpha_i + w_{it}'\phi$$  \hspace{1cm} (9)

Similar to the fixed effects logit model, a conditional maximum likelihood approach can
be used to derive the fixed effects Poisson estimator, where the fixed effects are eliminated
by conditioning on a sufficient statistic for $\alpha_i$ (Hausman et al., 1984; Palmgren, 1981). Panel-
robust standard errors are used for statistical inference to avoid significantly underestimating
the true standard errors in the presence of over dispersion (Allison & Waterman, 2002).3

---

3 The data appear to be slightly over dispersed based on Table 1 as the overall mean number of second half goals scored by
the observed team (0.7798) is lower than the variance (0.8777).
An alternative modelling approach for count data assumes that $z_{it}$ follows a negative binomial distribution, a modified version of the Poisson distribution that permits the variance to exceed the mean (overdispersion) by adding an additional parameter (Hausman et al., 1984). There are multiple ways in which the negative binomial distribution and, subsequently, the fixed effects negative binomial model can be parameterised, the most common of which is the NB2 parameterisation proposed by Hausman et al. (1984). Under this parameterisation, the mass function is

$$f(z_{it} | \lambda_{it}, \theta_i) = \frac{\Gamma(\lambda_{it} + z_{it})}{\Gamma(\lambda_{it})\Gamma(z_{it} + 1)} \left( \frac{\theta_i}{1 + \theta_i} \right)^{z_{it}} \left( \frac{1}{1 + \theta_i} \right)^{\lambda_{it}}$$

(10)

where $\theta_i$ is the dispersion parameter, $\Gamma(\cdot)$ is the gamma function, and $\lambda_{it}$ is the expected count of $z_{it}$ and depends on a log-linear specification of explanatory variables similar to the Poisson model

$$\ln \lambda_{it} = \alpha_i + w_{it}^t \phi$$

(11)

where $\alpha_i$ is the team-season fixed effect, and $w_{it}$ is a $K \times 1$ vector of explanatory variables with associated parameters $\phi$ (Cameron & Trivedi, 1998).

Using the Hausman et al. (1984) approach produces a negative binomial model where the dispersion (ratio of the variance over mean) equates to $1 + \theta_i$ and is permitted to vary over cross-sectional units (Cameron & Trivedi, 1998; Hausman et al., 1984). The negative binomial model has the advantage of directly accounting for over dispersion in the model specification, rather than through post-estimation adjustments to standard errors. Statistical inferences may, therefore, be more accurate with the negative binomial model and can use conventional observed information matrix standard errors. However, Allison and Waterman (2002) note that the conditional fixed effects negative binomial estimator is not a true fixed effects estimator and does not control for all stable predictors. An intercept term is included in the computer generated estimate as a result (unlike in the Poisson model), and it is also possible to estimate time-invariant explanatory variables (Allison, 2009). For this reason, the negative binomial model is not considered as the sole model of within-game momentum.

Three fixed effects Poisson and negative binomial models are estimated to examine within-game momentum in the EPL, all with second half goals scored by the observed team as the dependent variable. The initial model uses only the first half goals scored by the observed team as the independent variable, with the second model adding the first and second half goals scored by the opponent team. The third model incorporates the same five control

---

4 $\theta_i$ can take any value because it eventually drops out of the conditional likelihood, but is usually defined as $\alpha_i/\phi_i$ where $\alpha_i$ is the individual-specific fixed effect and $\phi_i$ is an overdispersion parameter (Cameron & Trivedi, 1998).
variables used in the across-game momentum analysis. To avoid confusion with the across-game momentum analysis, the three models are referred to as Models 4, 5, and 6.

5.2 Results and Discussion

The fixed effects Poisson and negative binomial estimation results for Models 4, 5, and 6 are provided in Table 5, including coefficients, incidence rate ratios, and standard errors (panel-robust for the Poisson models and conventional observed information matrix for the negative binomial models). The coefficients of the regression models measure the effect of a one unit change in the explanatory variable on the difference of the natural logarithm of the expected counts (or the incidence rate ratio), holding all other variables constant (McKenzie, 2013). When the coefficients are transformed exponentially, the interpretation of a one unit change in the explanatory variable relates more simply to incidence rate ratios.

The Poisson regression results in Table 5 indicate no significant within-game momentum effect in the EPL. According to Model 4, scoring one additional goal in the first half increases the incidence rate ratio for the observed team’s second half goals by a factor of 1.0445. This factor falls to 1.0428 in Model 5 and falls below 1 (indicating a negative coefficient) in Model 6 to 0.9956. Note that these momentum effects are small in magnitude and the standard errors increase in size as more covariates are added to the models. While Model 5 suggests that if the opponent scores more goals in the first half the observed team scores fewer goals in the second half, the reverse is true in Model 6. Both Models 5 and 6 show that an increase in the number of goals scored by the opponent in the second half also increases the second half goals scored by the observed team.

The results of the negative binomial regressions suggest that the conclusion of no significant within-game momentum is robust to changes in methodology, with no variations in the signs of the coefficients in any of the three models estimated. Only one coefficient is significant, that being for the variable measuring goals scored in the second half by the opponent where the coefficient is significantly different from zero at a 10% significance level. This conclusion largely supports the existing empirical literature that rejects the notion of momentum within games for individual athletes.
### Table 5: Fixed Effects Poisson and Negative Binomial Models

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Poisson Models</th>
<th>Negative Binomial Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed Team’s Second Half Goals</td>
<td>Observed Team’s Second Half Goals</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>First half goals scored by the observed team</td>
<td>0.0435</td>
<td>0.0419</td>
</tr>
<tr>
<td></td>
<td>1.0445</td>
<td>1.0428</td>
</tr>
<tr>
<td></td>
<td>(0.0396)</td>
<td>(0.0402)</td>
</tr>
<tr>
<td>First half goals scored by the opponent</td>
<td>-0.0311</td>
<td>0.0176</td>
</tr>
<tr>
<td></td>
<td>0.9694</td>
<td>1.0177</td>
</tr>
<tr>
<td></td>
<td>(0.0472)</td>
<td>(0.0458)</td>
</tr>
<tr>
<td>Second half goals scored by the opponent</td>
<td>0.0167</td>
<td>0.0680</td>
</tr>
<tr>
<td></td>
<td>1.0169</td>
<td>1.0704</td>
</tr>
<tr>
<td></td>
<td>(0.0395)</td>
<td>(0.0415)</td>
</tr>
<tr>
<td>Control variables?</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>1,140</td>
<td>1,140</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-1171</td>
<td>-1171</td>
</tr>
</tbody>
</table>

Note: For each variable, coefficients are reported first, followed by incidence rate ratios and then standard errors in parentheses (panel-robust for the Poisson models and conventional observed information matrix for the negative binomial models). *** denotes significance at 1%, ** denotes significance at 5%, and * denotes significance at 10%. The coefficients of the control variables had the expected signs in most cases.
6. Conclusion

This paper has examined the presence of across-game and within-game momentum in the context of the EPL. Initially, nonparametric Wald-Wolfowitz runs tests similar to those employed in the existing literature produced minimal evidence of momentum across games. Controlling for time-invariant heterogeneity among teams and across seasons, a series of fixed effects logit models found evidence of negative momentum in terms of the reversal of performance following a winning streak, consistent with the theory of positive inhibition (Cornelius et al., 1997; Silva et al., 1988). However, negative momentum was not consistently associated with streaks of draws and losses and was rarely significant. No evidence was found to support the conjecture of a nonlinear effect of streaks on performance. The fixed effects regression results for the within-game momentum analysis were more unanimous and robust to methodological changes, with both Poisson and negative binomial models indicating no significant momentum effect.

It is not possible from the evidence presented here to entirely reject or accept momentum as a real phenomenon. Rather, the results suggest that evidence of momentum is context-dependent but largely inconsistent with the ubiquitous popular belief in a positive association between past and future sporting outcomes. As such, it would appear that momentum, at least in the EPL, is better suited as a post hoc label of performance in the vein of the Projected Performance Model rather than a robust causal phenomenon (Cornelius et al., 1997). Further research should consider investigating the general implications of misperceptions of randomness for probabilistic reasoning and public policy development. A broader understanding of how these beliefs are constructed could lead to behavioural biases being incorporated into public policy formation such as through the development of advertising regulations for financial products and gambling.
References


