# The hot hand in professional darts

Marius Ötting, Roland Langrock and Christian Deutscher

*Bielefeld University, Germany*

and Vianey Leos-Barajas

*North Carolina State University, Raleigh, USA*

**Summary.** We investigate the hot hand hypothesis in professional darts in a nearly ideal setting with minimal to no interaction between players. Considering almost 1 year of tournament data, corresponding to 167492 dart throws in total, we use state space models to investigate serial dependence in throwing performance. In our models, a latent state process serves as a proxy for a player's underlying form, and we use auto-regressive processes to model how this process evolves over time. Our results regarding the persistence of the latent process indicate a weak hot hand effect, but the evidence is inconclusive.

*Keywords*: Hidden Markov model; Hot hand; Sports statistics; State space model; Time series

## 1. Introduction

In sports, the concept of a 'hot hand' refers to the idea that athletes may enter a state in which they experience exceptional success. For example, in basketball, players are commonly referred to as being 'in the zone' or 'on fire' when they hit several shots in a row. Although empirical analyses of the hot hand phenomenon tend to focus on sports because the corresponding data are relatively easily accessible, the notion of the hot hand in fact applies to much more general settings in which streaks may occur, including human performance in general (Gilden and Wilson, 1995a), artistic, cultural and scientific careers (Liu *et al.*, 2018), the performance of hedge funds (Hendricks *et al.*, 1993; Edwards and Caglayan, 2001; Jagannathan *et al.*, 2010), enduring rivalries in international relations (Gartzke and Simon, 1999; Colaresi and Thompson, 2002) and even gambling activities, against all odds (Xu and Harvey, 2014). However, when perceiving such dynamics, people tend to overinterpret streaks of success and failure (Bar-Hillel and Wagenaar, 1991). This phenomenon has been studied intensively by behavioural economists and psychologists (see, for example, Tversky and Kahneman (1971, 1974)) and is regarded as a cognitive illusion that has been considered as a primary example for how humans form beliefs and expectations (Thaler and Sunstein, 2009; Kahneman and Egan, 2011). Especially in gambling settings it has been demonstrated that people strongly believe in the 'streakiness' of their performances, while at the same time also acting according to the gambler's fallacy, such that after a streak of identical outcomes an increase in betting volume against the streak is observed although an independent and identically distributed random process is generating the outcome (Croson and Sundali, 2005). Such apparent irrationality underlines the importance

*Address for correspondence*: Marius Ötting, Department of Business Administration and Economics, and Department of Psychology and Sports Science, Bielefeld University, Bielefeld 33615, Germany.
E-mail: marius.oetting@uni-bielefeld.de

of being able to quantify precisely a potential hot hand effect in settings where its existence is highly disputed, e.g. in professional sports. In general, a profound knowledge regarding the existence and magnitude of streakiness in performances can aid general decision making (Miller and Sanjurjo, 2018).

In their seminal paper, Gilovich *et al.* (1985) analysed basketball free-throw data but found no support for a hot hand, hence coining the notion of the 'hot hand fallacy'. The alleged fallacy has been attributed in particular to a potential memory bias, with notable streaks in performances being more memorable than outcomes that are perceived as random, but also to general misconceptions regarding chance, with laypeople expecting randomness to lead to performances that are more balanced in terms of successes and failures than is actually so. Since Gilovich *et al.* (1985), there has been mixed evidence regarding the hot hand in sports, with some papers claiming to have found indications of a hot hand phenomenon and others disputing its existence. Bar-Eli *et al.* (2006) reviewed the literature on the hot hand in sports, including analyses of data from basketball, baseball, golf, tennis, volleyball and bowling. They summarized 24 studies, from which only 11 studies provide evidence for a hot hand effect. Perhaps because of the availability of increasingly large data sets, most of the more recent studies have found evidence for a hot hand effect (see Raab *et al.* (2012), Shea (2014), Green and Zwiebel (2017) and Miller and Sanjurjo (2018)), whereas only some studies dispute its existence by providing mixed results (see Wetzels *et al.* (2016) and Elmore and Urbaczewski (2018)).

Two types of approach have been used to investigate such potential hot hand patterns, namely

  (a) analyses of the serial correlation of shot *outcomes* (see, for example, Gilovich *et al.* (1985), Dorsey-Palmateer and Smith (2004) and Miller and Sanjurjo (2014)) and
  (b) such that use of a latent variable to describe the form of a player (see, for example, Albert (1993), Sun (2004), Wetzels *et al.* (2016) and Green and Zwiebel (2017)), where the hot hand is understood as serial correlation in shot *probabilities*.

Although there is no consensus in the literature regarding the definition of the hot hand, Stone (2012) and Miller and Sanjurjo (2018) showed that direct analyses of the correlation in outcomes, as per (a) above, may vastly underestimate the correlation in shot probabilities. For example, a correlation of $\rho_p = 0.4$ in shot probabilities can co-occur with a very much lower correlation of $\rho_r = 0.057$ in shot realizations (Stone, 2012; Miller and Sanjurjo, 2014). In other words, if a genuine hot hand process is driven by auto-correlation in success probabilities (i.e. in players' forms)—which may very well be so—then this can easily go undetected if the focus lies on the (much weaker) serial correlation of outcomes. Stone (2012) and Arkes (2013) thus concluded that it is preferable to analyse correlation in shot probabilities, as per (b) above. Hence, in this paper, we focus on approach (b), which we believe is also more aligned with the way that terminology related to the hot hand concept (e.g. 'on fire' and 'in the zone') is commonly applied—as argued by Stone (2012), it seems most natural to measure players' form by their time-varying success probabilities, rather than (noisy) shot outcomes.

In addition to such conceptual issues regarding the representation of the hot hand in the data-generating process, Miller and Sanjurjo (2018) highlighted a subtle selection bias that may sneak into analyses of sequential data, which provides a further challenge to the findings of Gilovich *et al.* (1985). Aside from mathematical fallacies, which would already seem to explain many failed attempts to prove the existence of the hot hand, we note that many of the existing studies considered data, e.g. from baseball or basketball, which we believe are hardly suitable for analysing streakiness in performances. For example, when analysing hitting streaks of a batter

in baseball, other factors such as the performance of the pitcher are also important but difficult to account for. The same applies to basketball, as there are also several factors affecting the probability of a player to make a shot, e.g. the position (of a field goal attempt) or the effort of the defence. In particular, an adjustment of the defensive strategy to stronger focus on a player during a hot hand streak can conceal a possible hot hand phenomenon (Bocskocsky *et al.*, 2014).

To overcome these *caveats*, here we investigate whether there is a hot hand effect in professional darts: a setting with a high level of standardization of individual throws. In professional darts, well-trained players repeatedly throw at the dartboard from the exact same position and effectively without any interaction between competitors, making the course of play highly standardized. In the existing literature, there are very few contributions that consider darts data, and almost all of these are restricted to laboratory settings. For example, Van Raalte *et al.* (1995) analysed the effect of positive and negative self-talk on throwing performances, considering the throwing sequences of 60 individuals, each of length 15. The hot hand effect has previously been investigated by using darts data by Gilden and Wilson (1995b); analysing only 24 throwing sequences of eight volunteers, they found no evidence for a hot hand effect. Here we consider a much larger data set, with $n = 167492$ throws in total, which allows for comprehensive inference regarding the existence and the magnitude of the hot hand effect. Using state space models, we evaluate serial dependence in a latent state process, which can be interpreted as a player's varying form, in line with approaches of type (b) above.

## 2.  Data

A dartboard is divided into 20 numbered slices (1–20 points) and the centre of the board, the latter with an outer circle (the single bull; 25 points) and an inner circle (the bull's-eye; 50 points). Each of the 20 slices is further divided into three segments, the singles, doubles and triples, with the last two resulting in twice or triple the slice number being awarded as points. All matches in our analysis were played by two players in the *501-up* format. In this format, both players start with 501 points and take their turns one after another. Within each turn, a player throws three darts in quick succession, with the value of the segment hit by each dart being reduced from the current score. The first player to reach exactly 0 points wins a 'leg'. The last dart used to reduce the score to 0 must hit a double or the bull's-eye ('double out'). (In some matches, a player must win a prespecified number of 'sets', where each set is played as best of five legs. Whether the match is played in sets or in legs is not relevant for our analysis.) To win the match, a player must be the first to win a prespecified number of legs (typically between 7 and 15). If a player wins the match, he advances to the next round of the tournament.

Data were extracted from http://live.dartsdata.com/, covering all professional darts tournaments organized by the Professional Darts Corporation between April 2017 and January 2018. In our analysis, we include all players who played at least 50 legs during the time period considered. This leads to a total of 8310 legs and 167492 dart throws (from 833 matches played across 25 tournaments).

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
https://rss.onlinelibrary.wiley.com/hub/journal/1467985x/series-
a-datasets.
```

At the beginning of a leg, players consistently aim at high numbers to reduce their points quickly. The maximum score in a single throw is 60 as in a triple 20 (T20), which players

usually aim at, but the data indicate that the triples of the numbers 11–19 (T11–T19), and the bull's-eye (bull), are targeted in the initial phase of a leg as well. In fact, Tibshirani *et al.* (2011) showed that T20 is not necessarily the best segment to aim at, depending on the precision of a player's throws. In addition, when the score before the last throw of a turn is slightly above or around 180, then with that throw players commonly try to avoid 'bogey numbers', i.e. scores below 170 points which cannot be reduced to 0 within a player's turn. For example, if the score is 182 points before the third dart of a turn, then aiming at T20 but hitting the single 20 would reduce the score to 162 points, which is a bogey number. Hence, with 182 points left, instead of aiming for T20, players may aim for T12, since, if they fail and hit the single 12, the score reduces to 170 points which can still be checked out with three darts during the next turn. Thus, in the initial phase of a leg we regard any throw to land in the set $H = \{T11, T12, T13, T14, T15, T16, T17, T18, T19, T20, \text{bull}\}$ as success. The corresponding empirical proportions of throws hitting the elements of $H$ are displayed in Table 1. Since a leg is won once a player reaches exactly 0 points, players do not always target $H$ towards the end of legs, but rather numbers that make it easier for them to reduce to 0. To retain a high level of standardization and comparability across throws, we thus truncate our time series data, excluding throws where the remaining score was less than $c = 180$ points.

We consider binary time series $\{y_t^{p,l}\}_{t=1,\ldots,T_{p,l}}$, indicating the throwing success of player $p$ within his $l$th leg in the data set, with

$$y_t^{p,l} = \begin{cases} 1 & \text{if the $t$th throw lands in } H, \\ 0 & \text{otherwise,} \end{cases}$$

where the $T_{p,l}$th throw is the last throw of player $p$ in his $l$th leg with the player's remaining score still greater than or equal to $c = 180$. The final data set then comprises $n = 167492$ throws by $P = 73$ players. To illustrate the structure as well as typical patterns of the data, we display Gary Anderson's throwing success histories throughout his first 15 legs in the data:

001 011 011
111 110 0
000 111 101
010 000 101 01
000 110 101
111 000 010 0
110 100 101
100 010 010 00
101 010 000 1
110 100 101
101 101 1
001 011 010 0
000 010 010 11
000 001 000 110
000 111 100

Each row corresponds to one leg—truncated when the score fell below 180—and gaps between blocks of three successive dart throws indicate a break in Anderson's play when his opponent was taking his turn. Next we formulate a model that enables us potentially to reveal any unusual streakiness in the data, i.e. a possible hot hand effect.

**Table 1.** Absolute frequencies and proportions for the different outcomes of $H$ in our data set

| Outcome | Frequency | Proportion |
|---------|-----------|------------|
| T11 | 1 | 0.00001 |
| T12 | 3 | 0.00002 |
| T13 | 1 | 0.00001 |
| T14 | 0 | 0 |
| T15 | 6 | 0.00004 |
| T16 | 4 | 0.00002 |
| T17 | 546 | 0.003 |
| T18 | 1709 | 0.010 |
| T19 | 9897 | 0.059 |
| T20 | 53509 | 0.319 |
| Bull | 108 | 0.0006 |

## 3. Modelling the hot hand in darts

### 3.1. State space model of the hot hand

We aim at explicitly incorporating any potential hot hand phenomenon in a statistical model for throwing success. As discussed in Section 1, it seems desirable to focus on potential serial correlation in success probabilities. Conceptually, a corresponding hot hand phenomenon naturally translates into a latent, serially correlated state process, which for any player considered measures his varying form. For average values of the state process, we would observe normal throwing success, whereas, for high (or low) values of the state process, we would observe unusually high (or low) percentages of successful attempts. Figuratively speaking, the state process serves as a proxy for the player's 'hotness'—alternatively, it can simply be regarded as the player's varying form. The magnitude of the serial correlation in the state process then indicates the strength of any potential hot hand effect. A similar approach was indeed used by Wetzels *et al.* (2016) and by Green and Zwiebel (2017), who used discrete state hidden Markov models to measure the form. Although there is some appeal in a discrete state model formulation, most notably mathematical convenience and ease of interpretation (with cold *versus* normal *versus* hot states), we doubt that players traverse only finitely many states, and we advocate a continuously varying underlying state variable instead, thus allowing for gradual changes in a player's form. Specifically, dropping the superscripts $p$ and $l$ for notational simplicity, we consider models of the form

$$y_t \sim \text{Bern}(\pi_t), \qquad \text{logit}(\pi_t) = \eta_t(s_t), \quad s_t = h_t(s_{t-1}) + \epsilon_t, \tag{1}$$

where $\{y_t\}_{t=1,\dots,T}$ is the observed binary sequence indicating throwing success, and $\{s_t\}_{t=1,\dots,T}$ is the unobserved continuous valued state process indicating a player's varying form. We thus model throwing success by using a logistic regression model in which the predictor $\eta_t(s_t)$ for the success probability $\pi_t$ depends, among other things, on the current form as measured by $s_t$. The unobserved form process $\{s_t\}$ is modelled by using an auto-regressive process.

Model (1) is a special case of a state space model, with the dependence structure that is illustrated in Fig. 1. Under this model, the outcomes of the throws are *conditionally* independent, given the player's form process. However, if there is serial correlation in the form process $s_t$, then
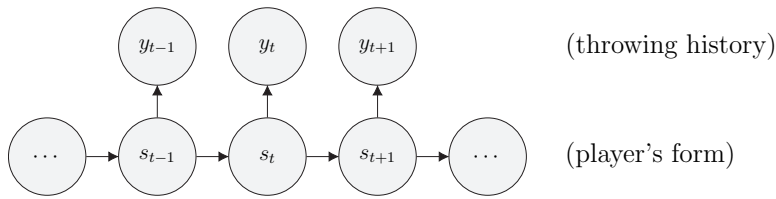
**Fig. 1.**    Dependence structure of the modelling framework used to investigate a potential hot hand effect: the throwing history $\{y_t\}$ is an observed binary sequence indicating whether or not individual throws were successful, whereas the player's form $\{s_t\}$ is an unobserved continuous valued variable which drives the player's time-varying success probabilities

this will induce serial correlation also in the throwing performance. This model formulation is thus in line with the suggestion by Stone (2012) to focus on the auto-correlation of success probabilities. Crucially, the model includes the possibility to be reduced to the nested special case of uncorrelated success probabilities, and hence independent observations, corresponding to the absence of a hot hand phenomenon. Before we discuss how to conduct maximum likelihood estimation within these types of model, we first present the models that were considered in the empirical analysis, i.e. the exact forms of $\eta_t(s_t)$ and of $h_t$ that we use to analyse the darts data.

### 3.2.  Model specifications
For our hot hand analysis, we formulate three models, which are all special cases of the general formulation stated in equation (1). We start with formulating a benchmark model (*model 1*) which corresponds to the absence of a hot hand effect. To account for differences between the players, this model includes player-specific intercepts $\beta_{0,p}$ in the predictor, which indicate the individual players' proportions of throwing success (on the logit scale). In principle, a player's overall performance level may change over his career but, since our data cover less than 1 year of tournaments, it is reasonable to assume that this quantity is constant for each player over the observation period that was considered. Furthermore, we include a categorical covariate $D_t$, $D_t \in \{1, 2, 3\}$, indicating the position of the dart thrown at time $t$ within the player's current turn (first, second or third). This categorical covariate is included since the relative frequency of hitting $H$, i.e. of throwing success in the early stages of a leg, in fact differs notably across the three throws within a player's turn, with the empirical proportions of hitting $H$ in our data found to be 0.355, 0.409 and 0.420 for the first, second and third throw respectively. The substantial improvement after the first throw within a player's turn is partly due to the necessary recalibration at the start of a turn (but see further discussion below). Hence, the predictor in model 1, a basic logistic regression model, is given by

$$\text{logit}(\pi_t) = \beta_{0,p} + \beta_1 I_{\{D_t=2\}} + \beta_2 I_{\{D_t=3\}},$$

with $I_{\{.\}}$ denoting the indicator function, and $\beta_{0,p}$ player $p$'s baseline level for the first dart within any given turn. In *model 2*, a state variable $\{s_t\}$ is included to account for potential serial correlation in success probabilities (i.e. a hot hand effect). We assume that $\{s_t\}$ follows an auto-regressive process of order 1:

$$\text{logit}(\pi_t) = \beta_{0,p} + \beta_1 I_{\{D_t=2\}} + \beta_2 I_{\{D_t=3\}} + s_t,$$

$$s_t = \phi s_{t-1} + \sigma \epsilon_t,$$

with $\epsilon_t \sim^{\mathrm{IID}} \mathcal{N}(0,1)$. Effectively this is a Bernoulli model for throwing success in which the success probability fluctuates around the players' baseline levels—$\beta_{0,p}$, $\beta_{0,p} + \beta_1$ and $\beta_{0,p} + \beta_2$ for within-turn throws 1, 2 and 3 respectively—according to the auto-regressive process $\{s_t\}$. The process $\{s_t\}$ can be interpreted as a player's time-varying form. Thus, at the time of an individual throw, the success probability of hitting $H$ depends on the general ability of the player, on the position of the throw within a turn, and on the current (underlying) form as modelled by the state process. Within this model formulation, a hot hand is defined as auto-correlation in the state process, i.e. in the current form. For $\phi = 0$, the model reduces to a model without auto-correlation in the underlying form (i.e. absence of a hot hand), which is similar to model 1 but includes an additional stochastic component in players' abilities. Otherwise, if there is auto-correlation in the underlying form, i.e. if $\phi > 0$, then this would provide evidence in favour of a hot hand effect in the form of positively correlated success probabilities. For the beginning of a new leg, we explicitly account for the result of the last leg, by assuming that $s_1 \sim \mathcal{N}(\mu_{\delta_{\mathrm{won}}}, \sigma_\delta)$ if the last leg was won, and $s_1 \sim \mathcal{N}(\mu_{\delta_{\mathrm{lost}}}, \sigma_\delta)$ if the last leg was lost, i.e. we assume that a player's form in a new leg may depend on the result of the last leg, as expressed by the mean of the initial distribution. It should be noted here that, by pooling the data of different players, the parameters $\beta_1$, $\beta_2$, $\phi$, $\sigma$, $\mu_{\delta_{\mathrm{won}}}$, $\mu_{\delta_{\mathrm{lost}}}$ and $\sigma_\delta$ are assumed to be equal across players, whereas the ability (as measured by $\beta_{0,p}$) varies between players. In Section 5 we discuss how these assumptions could be modified.

To consider the structure of a darts match in more detail, we further distinguish between transitions *within* a player's turn to throw three darts (e.g. between the first and second, or between the second and third throw) and those *across* the player's turns (e.g. between the third and fourth throw). This extension is considered in *model 3* and accounts for the fact that there is a short break in a player's action between his turns, with the next turn starting with an empty board, whereas within a single turn the three darts are thrown in very quick succession—it thus seems plausible that any possible hot hand effect may show time series dynamics within turns that are different from those across turns. Specifically, within model 3 we assume a periodic auto-regressive process of order 1 (PAR(1); Franses and Paap (2004)) to describe a player's time-varying form:

$$\mathrm{logit}(\pi_t) = \beta_{0,p} + \beta_1 I_{\{D_t=2\}} + \beta_2 I_{\{D_t=3\}} + s_t,$$

$$s_t = \begin{cases} \phi_a s_{t-1} + \sigma_a \epsilon_t & \text{if } t \bmod 3 = 1, \\ \phi_w s_{t-1} + \sigma_w \epsilon_t & \text{otherwise.} \end{cases}$$

Despite distinguishing between transitions within and across turns, this model still allows for longer-term hot hand effects that last through the whole leg.

Before we present the results of the various models in Section 4, in the next section we first discuss how to conduct maximum likelihood estimation within the general formulation given in equation (1).

### 3.3. Maximum likelihood estimation

The likelihood of a model as in equation (1) involves analytically intractable integration over the possible realizations of $s_t$, $t = 1, \ldots, T$. We use a combination of numerical integration and recursive computing, as first suggested by Kitagawa (1987), to obtain an arbitrarily fine approximation of this multiple integral. Specifically, we finely discretize the state space, defining a range of possible values $[b_0, b_m]$ and splitting this range into $m$ intervals $B_i = (b_{i-1}, b_i)$, $i = 1, \ldots, m$, of length $(b_m - b_0)/m$. The likelihood of a single throwing history can then be approximated as

$$L_T = \int \ldots \int p(y_1, \ldots, y_T, s_1, \ldots, s_T) \mathrm{d}s_T \ldots \mathrm{d}s_1$$

$$\approx \sum_{i_1=1}^{m} \ldots \sum_{i_T=1}^{m} \Pr(s_1 \in B_{i_1}) \Pr(y_1|s_1 = b_{i_1}^*) \prod_{t=2}^{T} \Pr(s_t \in B_{i_t}|s_{t-1} = b_{i_{t-1}}^*) \Pr(y_t|s_t = b_{i_t}^*), \qquad (2)$$

with $b_i^*$ denoting the midpoint of $B_i$. This is just one of several possible ways in which the multiple integral can be approximated (see, for example, Zucchini *et al.* (2016), chapter 11). In practice, we simply require that $m$ be sufficiently large. With the specification as a logistic regression model as in equation (1), we have that

$$\Pr(y_t|s_t = b_{i_t}^*) = [\mathrm{logit}^{-1}\{\eta_t(b_{i_t}^*)\}]^{y_t}[1 - \mathrm{logit}^{-1}\{\eta_t(b_{i_t}^*)\}]^{1-y_t}.$$

The approximate probability of the state process transitioning from interval $B_{i_{t-1}}$ to interval $B_{i_t}$, $\Pr(s_t \in B_{i_t}|s_{t-1} = b_{i_{t-1}}^*)$, follows immediately from the specification of $h_t$ and the distribution of the noise, $\epsilon_t$.

The computational cost of evaluating the right-hand side of equation (2) is of order $\mathcal{O}(Tm^T)$. However, the discretization of the state space effectively transforms the state space model into a hidden Markov model, with a large but finite number of states, such that we can apply the corresponding efficient machinery. In particular, for this approximating hidden Markov model, the forward algorithm can be applied to calculate its likelihood at a cost of order $\mathcal{O}(Tm^2)$ only (Zucchini *et al.* (2016), chapter 11). More specifically, defining $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ with $\delta_i = \Pr(s_1 \in B_i)$, $i = 1, \ldots, m$, the transition probability matrix (TPM) $\boldsymbol{\Gamma} = (\gamma_{ij})$ with $\gamma_{ij} = \Pr(s_t \in B_j|s_{t-1} = b_i^*)$, $i, j = 1, \ldots, m$, and $m \times m$ diagonal matrix $\mathbf{P}(y_t)$ with $i$th diagonal entry equal to $\Pr(y_t|s_t = b_i^*)$, the right-hand side of equation (2) can be calculated as

$$L_T \approx \boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Gamma}\mathbf{P}(y_2)\ldots\boldsymbol{\Gamma}\mathbf{P}(y_{T-1})\boldsymbol{\Gamma}\mathbf{P}(y_T)\mathbf{1}, \qquad (3)$$

with column vector $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^m$. Equation (3) applies to a single leg played by one player. Assuming independence of the individual leg's throwing histories, the likelihood of the full data set is obtained as

$$L = \prod_{p=1}^{73} \prod_{l_p=1}^{L_p} \boldsymbol{\delta}\mathbf{P}(y_1^{p,l_p})\boldsymbol{\Gamma}\mathbf{P}(y_2^{p,l_p})\ldots\boldsymbol{\Gamma}\mathbf{P}(y_{T_{p,l}}^{p,l_p})\mathbf{1}. \qquad (4)$$

We estimate the model parameters by numerically maximizing the approximate likelihood, subject to the usual technical issues as detailed in Zucchini *et al.* (2016).

## 4.  Results

For model 1, the benchmark model corresponding to the absence of a hot hand effect, the estimated player-specific baseline levels for the first within-turn throws, $\beta_{0,1}, \ldots, \beta_{0,73}$, range from $-1.021$ to $-0.295$, corresponding to throwing success probabilities ranging from 0.265 to 0.427. The coefficients $\beta_1$ and $\beta_2$, which correspond to the increase in throwing success probabilities after the first throw within a player's turn (on the logistic scale), are estimated as 0.228 and 0.276 respectively. From the first to the second within-turn throw, there is thus a strong increase in the chance of success, followed by a further, smaller increase from the second to the third throw.

Model 2, which unlike model 1 can capture a potential hot hand effect, was fitted by using $m = 150$ and $-b_0 = b_m = 2.5$ in the likelihood approximation, monitoring the likely ranges of the process $\{s_t\}$ to ensure that the range considered is sufficiently wide given the parameter estimates.

Table 2 displays the parameter estimates (except the player-specific intercepts) including 95% confidence intervals (CIs) based on the observed Fisher information. Crucially, the estimate $\hat{\phi} = 0.493$ seems to support the hot hand hypothesis, indicating considerable serial correlation in players' forms, with the associated CI not containing 0. The Akaike information criterion (AIC) clearly favours the state space formulation, model 2, over the benchmark model assuming independent throws, model 1 ($\Delta$AIC $= 550$). However, the state space model as it stands makes the implicit assumption that observations are regularly sampled, here such that the sampling unit is one throw. This fails to acknowledge the actual structure of a player's sequence of throws, with blocks of three darts being thrown in quick succession, with breaks of a couple of seconds between blocks (while the opponent takes his turn).

To reflect the grouping of darts better, model 3 distinguishes between transitions *within* a player's turn to throw three darts and those *across* the player's turns. For the (approximate) likelihood of model 3, the TPM $\mathbf{\Gamma}$ is then not constant across time anymore, but equal to either a within-turn TPM $\mathbf{\Gamma}^{(w)}$ or an across-turn TPM $\mathbf{\Gamma}^{(a)}$. For model 3, which is clearly favoured over model 2 by the AIC ($\Delta$AIC $= 242$), the parameter estimates as well as the associated CIs are displayed in Table 3. The estimate of the persistence parameter of the AR(1) process active within a player's turn, $\hat{\phi}_w = 0.727$ (95% CI [0.642; 0.811]), corresponds to fairly strong serial correlation. However, the estimate $\hat{\phi}_a = 0.058$ indicates only minimal persistence in the players' forms across turns. In fact, when at time $t$ a player begins a new set of three darts within a leg, then the underlying form variable is drawn from an $\mathcal{N}(0.058 s_{t-1}, 0.789^2)$ distribution, which is notably close to the two possible initial distributions of the AR(1) process ($\mathcal{N}(-0.025, 0.691^2)$

**Table 2.**   Parameter estimates with 95% CIs for model 2

| Parameter | Estimate | 95% CI |
|---|---|---|
| $\phi$ | 0.494 | [0.438; 0.550] |
| $\sigma$ | 0.659 | [0.565; 0.768] |
| $\beta_1$ | 0.248 | [0.221; 0.274] |
| $\beta_2$ | 0.297 | [0.269; 0.325] |
| $\mu_{\delta_{\text{won}}}$ | $-0.051$ | [$-0.102$; 0.001] |
| $\mu_{\delta_{\text{lost}}}$ | $-0.068$ | [$-0.117$; $-0.019$] |
| $\sigma_\delta$ | 0.701 | [0.659; 0.745] |

**Table 3.**   Parameter estimates with 95% CIs for model 3

| Parameter | Estimate | 95% CI |
|---|---|---|
| $\phi_w$ | 0.727 | [0.642; 0.811] |
| $\phi_a$ | 0.058 | [$-0.010$; 0.125] |
| $\sigma_w$ | 0.461 | [0.350; 0.607] |
| $\sigma_a$ | 0.789 | [0.699; 0.892] |
| $\beta_1$ | 0.270 | [0.242; 0.297] |
| $\beta_2$ | 0.331 | [0.302; 0.360] |
| $\mu_{\delta_{\text{won}}}$ | $-0.025$ | [$-0.069$; 0.019] |
| $\mu_{\delta_{\text{lost}}}$ | $-0.043$ | [$-0.084$; $-0.001$] |
| $\sigma_\delta$ | 0.691 | [0.648; 0.736] |

and $\mathcal{N}(-0.043, 0.691^2)$), which determine the form at the start of a leg. We cannot rule out that there may be a weak carry-over effect across turns—our results show no conclusive evidence in this regard, with the 95% CI for $\phi_a$ only just containing 0. We thus find strong evidence of serial correlation within turns, whereas the evidence regarding potential carry-over effects across turns is inconclusive—the relevance of these findings with regard to the hot hand effect will be discussed in Section 5.

Table 4 provides an overview of the three models fitted, detailing the number of parameters, the AIC values, the type of state process (if any) and a short description. To check the goodness of fit and the adequacy of our models, and also to obtain a more detailed picture of the (short-term) serial correlation that is implied by models 2 and 3, in Table 5 we compare the empirical relative frequencies of the eight possible throwing success sequences within players' turns—000, 001, 010, 011, 100, 101, 110 and 111—to the corresponding frequencies as expected under the three models that were fitted. We restricted this comparison to the first two turns of players within each leg and used Monte Carlo simulation to obtain the model-based frequencies of the eight sequences. The benchmark model (model 1), which corresponds to a complete absence of any hot hand pattern, clearly underestimates the proportion of 000 and 111 sequences, with deviations of up to 0.03. This indicates correlation in throwing performances within a turn. Model 2 better reflects the cumulation of 000 and 111 sequences, with a maximum deviation of 0.013. Finally, model 3, which is favoured by the AIC, almost perfectly captures the proportion of

**Table 4.** Overview of models 1–3

| Model | Number of parameters | AIC | State process | Description |
|---|---|---|---|---|
| 1 | 75 | 223200 | — | Player-specific intercepts and dummy variables for the throw within a player's turn |
| 2 | 79 | 222651 | AR(1) | Model 1 + AR(1) state process for the form |
| 3 | 81 | 222400 | PAR(1) | Model 1 + PAR(1) state process, distinguishing transitions within and across a player's turn |

**Table 5.** Relative frequencies of the eight possible throwing success histories within a player's turn†

| Sequence | Empirical proportion | Results for the following models: | | |
|---|---|---|---|---|
| | | *Model 1* | *Model 2* | *Model 3* |
| 0 0 0 | 0.252 | 0.222 | 0.239 | 0.250 |
| 0 0 1 | 0.151 | 0.159 | 0.153 | 0.150 |
| 0 1 0 | 0.130 | 0.152 | 0.139 | 0.136 |
| 0 1 1 | 0.114 | 0.111 | 0.113 | 0.110 |
| 1 0 0 | 0.103 | 0.121 | 0.114 | 0.108 |
| 1 0 1 | 0.080 | 0.088 | 0.082 | 0.082 |
| 1 1 0 | 0.086 | 0.084 | 0.083 | 0.082 |
| 1 1 1 | 0.084 | 0.063 | 0.075 | 0.083 |

†The second column gives the proportions found in the data, whereas the last three columns give the proportions as predicted under the various models fitted, for data structured exactly as the real data.

000 and 111 sequences, with the main mismatch in proportions (0.006) found for 010 sequences. Whereas Table 5 deals with outcomes *within* turns, Table 6 provides similar summary statistics corresponding to outcomes *across* turns. Considering the first two turns by a player within a leg, Table 6 displays the empirical and model-derived proportions of the 16 possible pairs $(a, b) \in \{0, 1, 2, 3\}^2$ corresponding to the number of successes within two consecutive turns; for example, the pair $(1, 3)$ indicates that the number of successful throws in the first and second turn are 1 and 3 respectively. We find that model 3 is clearly superior to model 2 in terms of capturing the observed proportions of pairs of turns with very different success (e.g. $(0, 3)$ and $(2, 0)$). This is because model 2 assumes that the parameter $\phi$ is constant across time, such that the corresponding estimate represents a compromise between within-turn and across-turn correlation found in the data. As a consequence, the former is underestimated, whereas the latter is overestimated, such that the model predicts that pairs of turns with very different success rates will occur less often than is actually so. In contrast, model 2 overestimates the magnitude of a potential hot hand effect across turns. In contrast, comparing the empirical proportions of the pairs $(0, 0)$ and $(3, 3)$ with the model-derived proportions, we find that model 1, representing the absence of a hot hand effect, underestimates persistence in success rates not only within turns, as shown in Table 5, but also across turns. Overall, model 3 thus comes closest to capturing both within-turn and across-turn correlation in performances. For the observed and expected frequencies that are shown in Table 6, a corresponding $\chi^2$ goodness-of-fit test rejects the null hypothesis at the 1% level for model 1 ($\chi^2 = 507.1$; degrees of freedom df $= 15$) and for model 2 ($\chi^2 = 130.2$; df $= 15$), whereas for model 3 the test fails to reject the null hypothesis ($\chi^2 = 29.80$; df $= 15$).

To investigate typical patterns of the hidden process $\{s_t\}$ further, we calculate, under model 3, the most likely trajectory of the latent state (i.e. form) for each player and leg. Specifically, again dropping the superscripts $p$ and $l$, we seek

**Table 6.**   Relative frequencies of the 16 possible combinations of numbers of success during the first two turns†

| Pair | Empirical proportion | Results for the following models: | | |
|------|---------------------|---------|---------|---------|
| | | *Model 1* | *Model 2* | *Model 3* |
| $(0, 0)$ | 0.067 | 0.051 | 0.063 | 0.064 |
| $(0, 1)$ | 0.097 | 0.096 | 0.100 | 0.098 |
| $(0, 2)$ | 0.072 | 0.062 | 0.063 | 0.067 |
| $(0, 3)$ | 0.022 | 0.013 | 0.016 | 0.021 |
| $(1, 0)$ | 0.092 | 0.096 | 0.098 | 0.100 |
| $(1, 1)$ | 0.150 | 0.186 | 0.166 | 0.155 |
| $(1, 2)$ | 0.110 | 0.122 | 0.114 | 0.109 |
| $(1, 3)$ | 0.032 | 0.027 | 0.031 | 0.034 |
| $(2, 0)$ | 0.067 | 0.062 | 0.060 | 0.066 |
| $(2, 1)$ | 0.106 | 0.122 | 0.110 | 0.105 |
| $(2, 2)$ | 0.079 | 0.082 | 0.081 | 0.076 |
| $(2, 3)$ | 0.026 | 0.018 | 0.024 | 0.025 |
| $(3, 0)$ | 0.020 | 0.013 | 0.015 | 0.019 |
| $(3, 1)$ | 0.031 | 0.027 | 0.029 | 0.030 |
| $(3, 2)$ | 0.023 | 0.018 | 0.023 | 0.023 |
| $(3, 3)$ | 0.007 | 0.004 | 0.007 | 0.007 |

†The second column gives the proportions found in the data, whereas the last three columns give the proportions as predicted under the various models fitted, for data structured exactly like the real data.
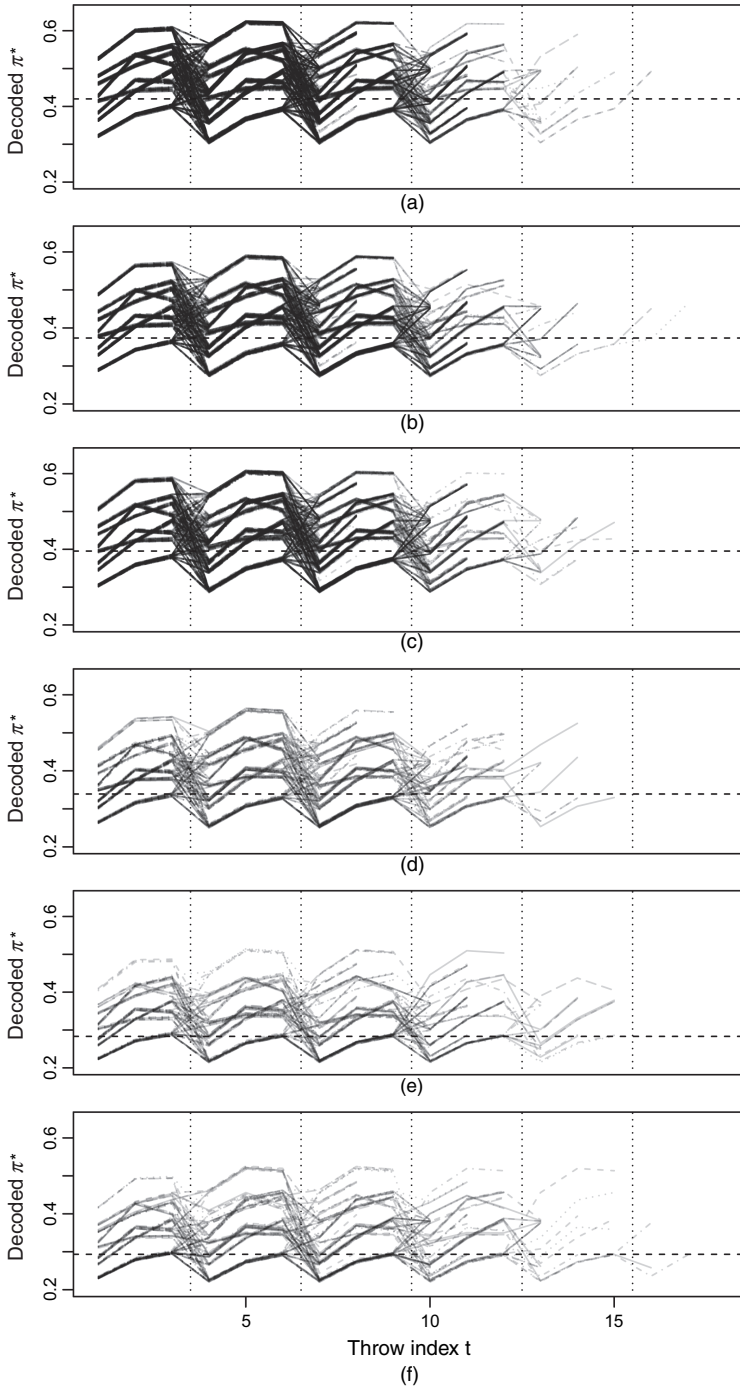
**Fig. 2.** Decoded most likely sequences of throwing success probabilities according to model 3, for more than 100 legs played by each of six players from the data set (- - - - -, player-specific intercepts for the respective player's within-turn throw 1; ⋮, transition between a player's turn of three darts each): (a) Michael van Gerwen; (b) Phil Taylor; (c) Gary Anderson; (d) Krzyszt of Ratajski; (e) Zoran Lerchbacher; (f) Peter Jacques

$$(s_1^*, \ldots, s_T^*) = \underset{s_1, \ldots, s_T}{\arg \max} \Pr(s_1, \ldots, s_T | y_1, \ldots, y_T),$$

i.e. the most likely state sequence, given the observations. After discretizing the state space into $m$ intervals, maximizing this probability is equivalent to finding the optimum of $m^T$ possible state sequences. This can be achieved at computational cost $\mathcal{O}(Tm^2)$ by using the Viterbi algorithm. We then calculate the corresponding trajectories $\pi_1^*, \ldots, \pi_T^*$ of the most likely success probabilities to have given rise to the observed throwing success histories, taking into account also the player-specific abilities and the dummy variables. Fig. 2 displays the decoded sequences for six players from the data set. Since there are only $2^3 = 8$ different possible sequences of observations within a player's turn, and since players start each turn almost unaffected by previous performances (compare $\hat{\phi}_a = 0.058$), there is only limited variation in the *most likely* sequences. The actual sequences may of course differ from these most likely sequences. The player-specific intercepts for within-turn throw 1 measure the difference in the players' abilities; in Fig. 2, the corresponding success probabilities range from 0.283 (Zoran Lerchbacher) to 0.420 (Michael van Gerwen). The probability of hitting $H$ increases after the first throw within a turn because of the two dummy variables. We also see confirmed that the form is not retained across turns.

## 5.  Discussion

Our results indicate that within a player's turn, involving three darts thrown in quick succession, there is strong correlation in the latent state process. However, short breaks, which in the given setting result from the opponent taking his turn, effectively result in a fresh start of the process describing the player's form. From a purely statistical point of view, if the hot hand phenomenon is understood as the presence of serial correlation in individuals' forms, then our findings would seem to provide strong evidence in favour of the hot hand. However, some strategic aspects in darts need to be considered when interpreting our results with regard to a potential hot hand effect. In particular, depending on the exact position of a turn's first dart within or close to a triple segment, this dart can potentially be used as a 'marker'. Darts 2 and 3 within the same turn can then be aimed at the marker and may be deflected into the target—the marker thus effectively increases the target area. The coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$, which indicate systematic differences in the success probabilities of the second and third throw within a turn, relative to the first throw, in fact indicate that the success probability for hitting $H$ increases during a player's turn (see Tables 2 and 3). However, the corresponding dummy variables do not incorporate any information on whether or not the first dart is a marker dart—if it is, then the success probability may *increase* even further than indicated by $\hat{\beta}_1$ and $\hat{\beta}_2$. Similarly, the first dart may also end up blocking a target area, thus *decreasing* the success probability of subsequent throws. Within our modelling approach, these effects seem to have been captured by the within-turn serial correlation that is induced by the latent auto-regressive process. In other words, although we conceptualized the latent state process as a proxy variable measuring a player's time-varying form, it seems likely that within turns this process actually accommodates other not directly observed effects, namely such as related to marker or blocker darts. The strong serial correlation within a player's turn, hence, may be partially tied to the effect of marker darts rather than a hot hand effect. Since our data do not contain corresponding information that would allow us to disentangle these two possible causes, there is no definitive conclusion whether the strong serial correlation within turns provides evidence of a hot hand effect, or rather is a consequence of marker darts. In any case, it is at least questionable whether serial correlation within a sequence of only three darts, thrown in quick succession, is what sports commentators, fans and athletes have in mind

when referring to the hot hand. In conclusion, although we find strong serial correlation for within-hand throws, the persistence parameter measuring potential correlation across turns was estimated to be very small, such that overall we do not find conclusive evidence for a hot hand.

Further research specifically on the hot hand in darts could focus on explicitly addressing player heterogeneity. In addition to the baseline level of $\pi_t$, the parameters $\phi_w$, $\phi_a$, $\sigma_w$ and $\sigma_a$, and hence the magnitude of the hot hand effect, may vary across players. This could reveal that for some players the hot hand effect lasts longer than for others, and potentially also across turns. Modelling this individual variability could be achieved by using covariates or, if no suitable covariates are available to explain the heterogeneity, via random effects. However, fitting state space models such as those presented here already involves a high computational cost, with the numerical maximization taking several days on a usual desktop computer. This would be further increased when incorporating random effects due to the required integration over possible values of the random effects. Possible parameter estimation approaches moving forward with random effects include

(a) using the Laplace approximation to evaluate the marginal likelihood via the template model builder package in R,
(b) an approximate Bayesian inference approach using an integrated nested Laplace approximation or
(c) using Markov chain Monte Carlo sampling in a Bayesian framework, where the realizations of the random effects are sampled alongside the other model parameters.

Although the focus of our analyses was on the hot hand hypothesis, darts in fact provide an excellent setting for studying several other performance-related hypotheses, because of the highly standardized actions and the absence of interactions between opponents. For example, darts data have recently been used to investigate how individuals perform in high-pressure situations (see Deutscher *et al.* (2018) and Klein Teeselink *et al.* (2018)).

The modelling framework that was developed in the present paper, with a continuous valued latent process representing a player's time-varying form, can easily be tailored to other sports—or in fact any sequential performance measures—for further investigations into the existence and magnitude of the hot hand. A *caveat* of the existing study is the binary nature of the observations, which corresponds to a rather noisy measure of the actual form. In sports such as archery and shooting, performance can be measured more precisely by considering the continuous valued distance between a shot and the middle of the target. Although these sports easily lend themselves to a time series analysis because of the structured way in which actions take place, there are many other sports of interest where time intervals between actions are irregular, including immensely popular sports such as football, American football, rugby or basketball. With the attention given to these, it would be of interest to transfer our modelling approach also to corresponding settings with irregular time intervals between actions. Conceptually, this can relatively easily be achieved by replacing the AR(1) process that was used in this work to represent the latent form of a player by its continuous time analogue, the Ornstein–Uhlenbeck process. Via discretizing the state space, the tools that are available for *continuous time* hidden Markov models can then be applied for making inference (Jackson *et al.*, 2003). However, given that evidence from the most recent literature points to a small hot hand effect (see, for example, Miller and Sanjurjo (2018) and Green and Zwiebel (2017)), we believe that sports settings with no direct interaction between opponents—e.g. archery, shooting, or free throws in basketball in an experimental setting as in Gilovich *et al.* (1985))—are best suited for analyses related to the hot hand, as otherwise it can be difficult to disentangle hot hand patterns from potential confounding factors.

## Acknowledgements

## References

Albert, J. (1993) A statistical analysis of hitting streaks in baseball: comment. *J. Am. Statist. Ass.*, **88**, 1184–1188.

Arkes, J. (2013) Misses in "hot hand" research. *J. Sprts Econ.*, **14**, 401–410.

Bar-Eli, M., Avugos, S. and Raab, M. (2006) Twenty years of "hot hand" research: review and critique. *Psychol. Sprt Exer.*, **7**, 525–553.

Bar-Hillel, M. and Wagenaar, W. A. (1991) The perception of randomness. *Adv. Appl. Math.*, **12**, 428–454.

Bocskocsky, A., Ezekowitz, J. and Stein, C. (2014) The hot hand: a new approach to an old "fallacy". *8th A. Massachusetts Institute of Technology Sloan Sports Analytics Conf.*

Colaresi, M. P. and Thompson, W. R. (2002) Hot spots or hot hands?: Serial crisis behavior, escalating risks, and rivalry. *J. Polit.*, **64**, 1175–1198.

Croson, R. and Sundali, J. (2005) The gambler's fallacy and the hot hand: empirical data from casinos. *J. Risk Uncertnty*, **30**, 195–209.

Deutscher, C., Ötting, M., Langrock, R., Gehrmann, S., Schneemann, S. and Scholten, H. (2018) Very highly skilled individuals do not choke under pressure: evidence from professional darts. *Preprint arXiv:1809.07659*. Bielefeld University, Bielefeld.

Dorsey-Palmateer, R. and Smith, G. (2004) Bowlers' hot hands. *Am. Statistn*, **58**, 38–45.

Edwards, F. R. and Caglayan, M. O. (2001) Hedge fund performance and manager skill. *J. Fut. Markts*, **21**, 1003–1028.

Elmore, R. and Urbaczewski, A. (2018) Hot and cold hands on the PGA tour: do they exist? *J. Sprts Analyt.*, **4**, 275–286.

Franses, P. H. and Paap, R. (2004) *Periodic Time Series Models*. Oxford: Oxford University Press.

Gartzke, E. and Simon, M. W. (1999) "Hot hand": a critical analysis of enduring rivalries. *J. Polit.*, **61**, 777–798.

Gilden, D. L. and Wilson, S. G. (1995a) On the nature of streaks in signal detection. *Cogn. Psychol.*, **28**, 17–64.

Gilden, D. L. and Wilson, S. G. (1995b) Streaks in skilled performance. *Psychon. Bull. Rev.*, **2**, 260–265.

Gilovich, T., Vallone, R. and Tversky, A. (1985) The hot hand in basketball: on the misperception of random sequences. *Cogn. Psychol.*, **17**, 295–314.

Green, B. and Zwiebel, J. (2017) The hot-hand fallacy: cognitive mistakes or equilibrium adjustments?: Evidence from Major League Baseball. *Mangment Sci.*, **64**, 5315–5348.

Hendricks, D., Patel, J. and Zeckhauser, R. (1993) Hot hands in mutual funds: short-run persistence of relative performance, 1974–1988. *J. Finan.*, **48**, 93–130.

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W. and Couto, E. (2003) Multistate Markov models for disease progression with classification error. *Statistician*, **52**, 193–209.

Jagannathan, R., Malakhov, A. and Novikov, D. (2010) Do hot hands exist among hedge fund managers?: An empirical evaluation. *J. Finan.*, **65**, 217–255.

Kahneman, D. and Egan, P. (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kitagawa, G. (1987) Non-Gaussian state-space modeling of nonstationary time series. *J. Am. Statist. Ass.*, **82**, 1032–1063.

Klein Teeselink, B., Potter van Loon, R., van den Assem, M. J. and van Dolder, D. (2018) Incentives, performance and choking in darts. VU Amsterdam, Amsterdam. (Available from SSRN 3304092.)

Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C. and Wang, D. (2018) Hot streaks in artistic, cultural, and scientific careers. *Nature*, **559**, 396–399.

Miller, J. B. and Sanjurjo, A. (2014) A cold shower for the hot hand fallacy. *Working Paper 518*. Innocenzo Gasparini Institute for Economic Research, Bocconi University, Milan.

Miller, J. B. and Sanjurjo, A. (2018) Surprised by the hot hand fallacy?: A truth in the law of small numbers. *Econometrica*, **86**, 2019–2047.

Raab, M., Gula, B. and Gigerenzer, G. (2012) The hot hand exists in volleyball and is used for allocation decisions. *J. Exptl Psychol. Appl.*, **18**, 81–89.

Shea, S. (2014) In support of a hot hand in professional basketball and baseball. *PsyCh J.*, **3**, 159–164.

Stone, D. F. (2012) Measurement error and the hot hand. *Am. Statistn*, **66**, 61–66.

Sun, Y. (2004) Detecting the hot hand: an alternative model. In *Proc. A. Meet. Cognitive Science Society*, vol. 26.

Thaler, R. H. and Sunstein, C. R. (2009) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin.

Tibshirani, R. J., Price, A. and Taylor, J. (2011) A statistician plays darts. *J. R. Statist. Soc.* A, **174**, 213–226.

Tversky, A. and Kahneman, D. (1971) Belief in the law of small numbers. *Psychol. Bull.*, **76**, 105–110.

Tversky, A. and Kahneman, D. (1974) Judgment under uncertainty: heuristics and biases. *Science*, **185**, 1124–1131.

Van Raalte, J. L., Brewer, B. W., Lewis, B. P., Linder, D. E., Wildman, G. and Kozimor, J. (1995) Cork!: The effects of positive and negative self-talk on dart throwing performance. *J. Sprt Behav.*, **18**, 50–57.

Wetzels, R., Tutschkow, D., Dolan, C., van der Sluis, S., Dutilh, G. and Wagenmakers, E.-J. (2016) A Bayesian test for the hot hand phenomenon. *J. Math. Psychol.*, **72**, 200–209.

Xu, J. and Harvey, N. (2014) Carry on winning: the gamblers' fallacy creates hot hand effects in online gambling. *Cognition*, **131**, 173–180.

Zucchini, W., MacDonald, I. L. and Langrock, R. (2016) *Hidden Markov Models for Time Series: an Introduction Using R*. Boca Raton: Chapman and Hall–CRC.