The Hot Hand Fallacy in Call of Duty Search and Destroy

Tony ElHabr

anthonyelhabr@gmail.com

Andrew ElHabr
School of Industrial and Systems Engineering (ISyE)
Georgia Institute of Technology
Atlanta, Georgia, USA
andrewelhabr@gatech.edu

August 25, 2022

Abstract

Our research investigates patterns in round win percentages in professional Search and Destroy (SnD) matches of the popular first-person shooter game Call of Duty (CoD). First, we find evidence supporting the hypothesis that round win probability can be modeled as a constant across rounds in the series, although not at the naive 50%. Second, we examine the proportion of round wins immediately following a win streak. Given streak length, series length, and series winner, we find no evidence that post-streak win proportion is significantly different than expected, suggesting that the perception of momentum at the team level is deceiving. Wald-Wolfowitz run tests also fail to provide evidence for the hot hand phenomenon.

Keywords esports · Call of Duty · hot hand · runs test

1 Introduction

1.1 Call of Duty Search and Destroy

Call of Duty (CoD), first released in 2003, is one of the most popular first-person shooter (FPS) video game franchises of all-time. The most popular mode in the competitive scene is "Search and Destroy" (SnD), in which one team tries to destroy one of two designated bomb sites on the map.[1]

In professional CoD SnD, the two teams[2] take turns playing offense and defense every round. The first to six round wins (best-of-11 round format) is declared the series winner.[3] A round can end in one of four ways:

1. One team eliminates all members of the other team prior to a bomb plant. (Eliminating team wins.)
2. The offensive team eliminates all members of the defensive team after a bomb plant. (Offense wins.)
3. The defense defuses the bomb after a bomb plant.[4] (Defense wins.)

---

[1]SnD bears resemblance to "Bomb Defusal" in Counter-Strike and "Plant/Defuse" in Valorant, two other FPS games played in more popular professional leagues.

[2]In the 2020 season, teams played with five players each; in the 2021 and 2022 seasons, teams played with four players each.

[3]A short list of rules that govern timing follows.

- Each round has a 90 second time limit, not counting the potential extension granted if a bomb is planted
- A bomb plant takes five seconds. The timer resets if the player stops planting site prior to completing it.
- A bomb defuse takes seven seconds. The timer resets if the player "drops" the bomb.
- Once the bomb is planted, the round timer is set to 45 seconds.

[4]Often the defensive team will try to eliminate all team members prior to making the defuse, but in some cases, they may try to "ninja" defuse.

Table 1: Offensive round win percentage for the upcoming round, given both the offensive and defensive team's prior number of round wins. Numbers in parentheses indicate sample sizes.

	Offense round wins					
Defense round wins	0	1	2	3	4	5
0	47.8%(852)	46.6%(408)	43.1%(216)	43.5%(115)	43.3%(67)	40.5%(37)
1	48.6%(444)	49.3%(418)	51.5%(309)	43.4%(205)	43.3%(120)	39.4%(99)
2	52.8%(218)	48.9%(305)	48.9%(315)	46.6%(262)	48.7%(189)	42.1%(133)
3	54.5%(123)	46.0%(200)	49.6%(250)	45.6%(248)	44.4%(214)	44.8%(174)
4	56.9%(65)	54.5%(145)	47.2%(193)	44.7%(228)	55.2%(221)	50.5%(208)
5	47.4%(38)	49.4%(83)	47.1%(136)	50.9%(175)	45.2%(177)	46.0%(202)

    4. The offense does not make a plant by the time the round timer ends. (Defense wins.)

We adopt the terminology "series" to refer to a single best-of-11 matchup, so as to mirror the terminology of playoff series in professional leagues like the National Basketball Association (NBA), National Hockey League (NHL), and Major League Baseball (MLB). A "game" or a "match" in such leagues is analogous to a "round" of CoD SnD for the purposes of this paper.

We note that SnD is not the only game mode played in competitive CoD.[5] Teams play in a head-to-head, best-of-five format, where SnD is always played as the second and fifth game modes. The best-of-five matchup could also be called a "series", but since we analyze only the SnD games, we refer to the SnD games as series.

1.2 Data

CoD has roughly gone through three eras of professional gaming: (1) Major League Gaming (MLG) tournaments prior to 2016; (2) the CoD World League (CWL), initiated in 2016; and (3) the 12-franchise CoD League (CDL), operating since 2020. The CDL has completed three year-long "seasons" as of August 2022.[6]

The data set consists of all SnD matches played in major tournaments and qualifiers during the CDL era, totaling 7,792 rounds across 852 series.[7] Data has been collected in spreadsheets by community member "IOUTurtle".[8]

6-0 sweeps make up 4.7% of all series, while 6-5 series make up 23.7% of series. The observed offensive round win percentage across all rounds is $\bar{\tau}_O = 47.8\%$.[9] Table 1 shows offensive round win percentages by series "state", i.e. the number of round wins by each team prior to an upcoming round. Offensive round win percentage is not quite constant, although never veers more than 10% from $\bar{\tau}_O$.

At tournaments, higher-seeded ("better") teams choose whether they want to start on offense in the SnD series, played as the second or fifth game mode of the best-of-five matchup. Their opponent is assigned to start on offense in the SnD slot not chosen.[10] Outside of tournaments, teams all play each other twice throughout the season, constituting qualifiers. Each team gets to play the higher-seeded role once in their head-to-head matchups with a given team, despite their win-loss records. Overall, we can argue that potential

---

[5]Hardpoint has been played as the first and fourth maps in a matchup. The variant for the third map was Domination in the 2020 season, and Control in the 2021 and 2022 seasons.

[6]CoD is fairly unique compared to other esports in that it runs on an annual lifecycle, with releases coming in the late fall. A new game under the same brand—Call of Duty—is published every year by a rotating set of developers. Each new game bears resemblance to past ones, often introducing relatively small variations ("improvements") to graphics, game modes, and other facets of gameplay. During the CDL era, the games released have been Modern Warfare (2020), Cold War (2021) and Vanguard (2022).

[7]288 of the series occur in major tournaments, which are considered to be more competitive than the qualifiers since they are played in person (COVID-permitting), eliminating randomness due to online latency and ping. The qualifiers are played online.

[8]Raw data: https://linktr.ee/CDLArchive. Author: https://twitter.com/IOUTurtle. Our clean data can be found at https://github.com/tonyelhabr/fps_round_games/blob/master/data/cod_rounds.csv.

[9]Offensive round win percentage has been nearly constant across the three games during the CDL era: 1. 47.2% in MW (2020) 2. 47.9% in Cold War (2021) 3. 48.1% in Vanguard (2022)

[10]Anecdotally, the higher-seeded teams tend to choose to play defense in the first round of the SnD series played as the second game mode in the best-of-five matchup, although this choice is not consistent across teams, or even with the same team over time.

bias due to stronger teams playing defense—or, offense, if offense happens to be generally advantageous for a given map—is very minimal.

2 Literature review

There have been a handful of studies of the distribution of games played in a series of a professional sport. Most assume a constant probability ϕ of a given team winning a game in the series, regardless of the series state. Mosteller (1952) observed that the American League had dominated the National League in MLB World Series matchups through 1952, implying that games should not be modeled with a constant $\phi_0 = 0.5$. Mosteller proposed three approaches for identifying the optimal constant probability value of the stronger team in the World Series, finding $\hat{\phi} \approx 0.65$ in each case: (1) solve for ϕ from the observed average number of games won by the loser of the series, which he called the "method of moments" approach; (2) maximize the likelihood that the sample would have been drawn from a population in which the probability of a team winning a game is constant across the series (i.e. maximum likelihood), and (3) minimize the chi-squared goodness of fit statistic χ^2 as a function of ϕ.

Chance (2020) re-examined the constant probability notion in the MLB World Series (1923–2018), the NBA Finals (1951–2018), and the NHL Stanley Cup (1939–2018). Chance found strong evidence against the null hypothesis of $\phi_0 = 0.5$ in the MLB and NHL championship series when applying Mosteller's first and second methods.

The concept of momentum goes hand-in-hand with a discussion of the nature of series outcomes.[11] Two opposing fallacies are observed in the context of momentum: the "gambler's fallacy" (negative recency) and the "hot hand fallacy" (positive recency). Per Ayton et al. (2004), negative recency is "the belief that, for random events, runs of a particular outcome . . . will be balanced by a tendency for the opposite outcome", while positive recency is the expectation of observing future results that match recent results.

Studying both player streaks and team streaks in basketball, in both observational and controlled settings, Gilovich et al. (1985), henceforth GVT, did not find evidence for the hot hand phenomenon. However, Miller and Sanjurjo (2018), henceforth MS, provided a framework for quantifying streak selection bias, which effectively works in the manner posited by the gambler's fallacy. Specifically, MS stated that a "bias exists in a common measure of the conditional dependence of present outcomes on streaks of past outcomes in sequential data" implying that, under i.i.d. conditions, "the proportion of successes among the outcomes that immediately follow a streak of consecutive successes is expected to be strictly less than the underlying (conditional) probability of success". When applying streak selection bias to GVT's data, MS came to the opposite conclusions as GVT.

Other research has borrowed techniques from the field of quality control, such as identifying unlikely sequences of events with the Wald-Wolfowitz runs test (Peel and Clauset 2015, Steeger et al. 2021). Peel and Clauset found no evidence for unlikely sequences of scoring events in the NHL, College Football (CFB), and National Football League (NFL), but did in the NBA. As a check on their entropy approach to momentum identification, Steeger et al. found several NHL teams with sequence of wins in the 2018-2019 regular season that violated the Wald-Wolfowitz null hypothesis.

2.1 Our contribution

Despite the plethora of existing research on games played in a series and momentum in sports, these topics have yet to be investigated heavily, if at all, in esports. Work has been done to examine intra-round, second-by-second win probability in other FPS titles such as Counter-Strike (Xenopoulos, Freeman, and Silva 2022) and Valorant (DeRover 2021), both of which are round-based like CoD SnD. However, considering Counter-Strike and Valorant specifically, research on round-level trends seems non-existent, perhaps for one of several reasons:

1. Both have economic aspects that can create clear advantages for one team in a round, given how prior rounds played out. CoD has no such equivalent, except for perhaps "score streaks", which infrequently occur.
2. Teams play either offense or defense for many consecutive rounds. (The former has a 15-15-1 format and the latter uses a 12-12-1 format.) On the other hand, teams in CoD SnD rotate roles every round,

---

[11]We often use "streaks" and momentum interchangeably, but as Steeger et al. ((2021)) note, momentum implies dependence between events, whereas streaking does not.

analogous to a 1-1-1-1-1-1-1 format for home advantage in best-of-seven series for professional sports like the MLB, NBA, and NHL.[12] While theoretically one might be able to account for inter-round win percentage relationships in any kind of format, such as a 5-5-1, the rotation of team sides every round is convenient for convincing ourselves that rounds could reasonably be modeled as i.i.d. Bernoulli trials.

3. Both Counter-Strike and Valorant have overtime rules—team must win by two rounds—which can make end-of-series sequences difficult to model properly. CoD SnD does not have overtime rules.

To our knowledge, there is no existing public statistical research on the CoD SnD format, beyond descriptive analysis on social media. While intra-round trends may be more directly applicable to teams looking for an advantage on their competition, broader investigation of a concept like the hot-hand fallacy in a sport where it has not yet been invetigated—particularly one that is less subjected to factors that may be difficult to control for, e.g. weather—should be useful as a precedent or supplement for future researchers.

3 Methodology

3.1 Distribution of rounds played

In a best-of-s format, assuming a constant round win probability ϕ[13] (where $0 \leq \phi \leq 1$) for one team[14], the expected proportion of series ending in r rounds ($r \leq s$) is given by Equation 2.

$$m = \frac{s+1}{2}. \tag{1}$$

$$\hat{\Phi}(r) = \frac{(r-1)!}{(m-1)!(s-r)!}(\phi^m(1-\phi)^{r-m} + \phi^{r-m}(1-\phi)^m). \tag{2}$$

For example, assuming $\phi = 0.5$, the probability of a series ending in nine rounds in CoD SnD, where $s = 11$, is

$$\hat{\Phi}(9) = \frac{(9-1)!}{(6-1)!(11-9)!}(0.5^6(1-0.5)^{9-6} + 0.5^{9-6}(1-0.5)^6) = 56(0.5^9 + 0.5^9) = 0.21875.$$

To evaluate the constant round win probability hypothesis, we can compute the test statistic

$$\chi^2 = \sum_{r \in R} \frac{(\Phi(r) - \hat{\Phi}(r))^2}{\hat{\Phi}(r)} \tag{3}$$

in which $\hat{\Phi}(r)$ and $\Phi(r)$ represent the expected and observed round win proportions respectively, and $R = [6, 7, 8, 9, 10, 11]$ for CoD SnD. The null hypothesis is that the test statistic is chi-squared distributed, where the number of degrees of freedom is $|R| - 1 = 5$ for Cod SnD. We reject the null hypothesis if $\chi^2 > \chi^2_{1-\alpha,|R|-1}$, where $\chi^2_{1-\alpha,|R|-1}$ is the chi-square critical value with $|R| - 1$ degree of freedom and siginficance level α.

3.2 Momentum

3.2.1 MS post-streak probability

Let us now consider round win proportions immediately following a streak of k wins, given that the series lasts r rounds. Removing our knowledge of a streak, we might model the Bernoulli round win probability as

---

[12]1-1-1-1-1-1-1 is not used today in these leagues, but it was at least once in each.

[13]If $\phi > 0.5$, we can say that this is the stronger team.

[14]ϕ and other symbols are selected so as to reserve symbols like p for usage in other contexts without causing confusion for the reader. Upper-case symbol, e.g. Φ, are consistently used in this paper to represent proportions, while lower-case symbols corresponding to upper-case symbols, e.g. ϕ, are used to represent Bernoulli trial success rates. "Hats", e.g. $\hat{\Phi}$, are used to convey expectations, while bare symbols symbols convey observational data.

$$p_0(\text{win}|r) = p_0^{+|r} = \begin{cases} \frac{m}{r} & \text{team wins series,} \\ \frac{r-m}{r} & \text{team loses series,} \end{cases} \tag{4}$$

using m from Equation 1.

As implied by MS's Theorem 1, we should expect the proportion of round wins immediately following a streak of k rounds wins for a series lasting r rounds, $\hat{P}_{MS}^{+|k,r}$, to be strictly less than $p^{+|r}$.[15]

Although one might be tempted to mirror the equal proportions hypothesis testing performed by MS, our context is fundamentally different from that of MS, who focus on longitudinal data in controlled settings, unaffected by opposition.[16] The number of trials is fixed in their experimental designs, but in CoD SnD, the number of rounds played is determined as a function of the max number of possible rounds (s) and whether or not the team wins the series. The Bernoulli trial success probability, i.e. single round win percentage in CoD SnD, is not independent of the opponent. Consequently, a statistical test of the difference in $\hat{P}^{+|k,r}$ and $\hat{P}^{-|kr}$, as performed by MS to evaluate their hypothesis regarding post-streak success rate, is not completely appropriate, although useful as a reference.

3.2.2 "Notional" post-streak probability

One can consider another form of the expected proportion of rounds won immediately after a streak of k round wins in a best-of-s series given the length of the series (r rounds), the "notional" proportion $\hat{P}_0^{+|k,r}$. The Bernoulli round win probability underlying $\hat{P}_0^{+|k,r}$ is

$$p_0(\text{win}|k,r) = p_0^{+|k,r} = \begin{cases} \frac{m-k}{r-k} & \text{team wins series,} \\ \frac{s-m-k}{r-k} & \text{team loses series,} \end{cases} \tag{5}$$

using m from Equation 1.

We can perform a two-tailed binomial test to evaluate the null hypothesis $\omega = \omega_0$ for the observed probability of success ω and a user-specified ω_0 (where $0 \leq \omega_0 \leq 1$). If there are r^+ observed successes in a sample of r trials and we expect that there should be $r * \omega_0$, the probability of arriving at this expected number of successes is

$$\Pr(r^w) = \binom{r}{r^+} \omega^{r^+} (1 - \omega)^{r - r^+}. \tag{6}$$

Treating the notional proportion $\hat{P}_0^{+|k,r}$ as the null ω_0 and plugging in the observed proportion $P^{+|k,r}$ for ω in Equation 6 (treating proportions as probabilities), we can evaluate the null hypothesis that the observed proportion of post-streak round wins is equal to the notional proportion, given streak length k and series length r. If we can reject this null hypothesis, then we can consider team momentum, represented by post-streak success, plausible.

As a reference, we can perform the same binomial test for the MS's streak-selection-adjusted proportion, $\hat{P}_{MS}^{+|k,r}$. However, given the caveats mentioned before regarding applying MS's theorem to our setting, the results of such binomial tests should be heeded with caution.

We can further decompose Equation 5 by the round i (where $i \leq r$) in which the streak of length k carries into.

---

[15]Unfortunately, there does not exist a closed form representation for the expected value of $\hat{P}_{MS}^{+|k,r}$ for $k > 1$. Nonetheless, one may run simulations to estimate the expected value, as we choose to do. We have adapted the code from Vafa (2017), which implements MS's framework (2018).

[16]GVT also perform statistical tests on shots from players in live games, i.e. "observational" data, but they note that their findings are likely affected player shot selection in the face of defensive strategy by the opposing team.

$$p_0(\text{win}|k,r,i) = p_0^{+|k,r,i} = \begin{cases} \frac{m-k}{r-k}, & \text{team wins series}, i \neq r, \\ 1, & \text{team wins series}, i = r, \\ \frac{s-m-k}{r-k}, & \text{team loses series}, i \neq r, \\ 0, & \text{team loses series}, i = r. \end{cases} \tag{7}$$

Again, we can apply a binomial test to evaluate the hypothesis that the expected proportions, $\hat{P}_0^{+|k,r,i}$ and $\hat{P}_{MS}^{+|k,r,i}$ separately, are equal to the observed proportion $P^{+|k,r}$.

3.2.3 Wald-Wolfowitz runs test

Stepping back from the (one-sided) perspective of a single team's round win probability when streaking, one can attempt to detect the hot hand phenomenon more broadly with a Wald-Wolfowitz runs test. Under the null hypothesis, the number of runs in a sequence of r trials is a random variable that can take on values $+$ or $-$ and arrive at r^+ successes ($r^- = r - r^+$ failures), with the following mean μ and variance σ^2:

$$\mu = \frac{2r^+r^-}{r} + 1, \sigma^2 = \frac{(\mu-1)(\mu-2)}{r-1}. \tag{8}$$

To incorporate our findings regarding constant round win probability, We can specify that $\Pr(r^+) = \phi$ (and, conversely, that $\Pr(r^-) = 1 - \phi$).

One can subset the observed series sequences, $\zeta(r)$, to those that violate the null hypothesis for the runs test and perform a test of equal proportions, where the null is that the observed proportion of a sequence relative to all possible sequences, $P(\zeta) = P^\zeta$, is equal to the expected proportion of the sequence relative to all possible sequences, $\hat{P}^\zeta$. The test statistic is

$$z = \frac{P^\zeta - \hat{P}^\zeta}{\sqrt{P_\delta^\zeta(1 - P_\delta^\zeta)(1/N^\zeta + 1/\hat{N}^\zeta)}} \tag{9}$$

where

$$P_\delta^\zeta = \frac{P^\zeta - \hat{P}^\zeta}{N^\zeta - \hat{N}^\zeta}$$

and where N^ζ and $\hat{N}^\zeta$ are the observed and expected number of sequences respectively. The null hypothesis is that z is normally distributed. If $z > +z_{(1-\alpha)/2}$ or $z < -z_{(1-\alpha)/2}$ for critical values at a significance level α, we can reject the null hypothesis for such sequences and build an argument in support of the hot hand effect.

4 Results

4.1 Distribution of rounds played

Using Equation 3, we find that $\chi^2 = 16.0$ (p-value of 0.0068) for $\phi_0 = 0.5$. Thus, we can comfortably reject the constant probability hypothesis for the null $\phi_0 = 0.5$, even at a confidence level of $\alpha = 0.01$.

Table 2 shows the expected series lasting r rounds, the expected proportion of series given $\phi_0 = 0.5$, $\hat{\Phi}_0(r)$, and the observed proportions, $\Phi(r)$.

Table 3 shows the alternate values for the constant round win probability that we find when applying the three methods suggested by Mosteller (1952). Each is approximately equal to 0.575. When applying Equation 3, each results in a χ^2 value for which we cannot reject the constant probability null hypothesis. Note that rejection of the null hypothesis is the baseline and failure to reject the null is the non-trivial result, contrary to most statistical analysis.

Table 2: The expected proportion of CoD SnD series lasting r rounds, $\hat{\Phi}_0(r)$, under the assumption that each team has a constant round win probability $\phi_0 = 0.5$. Additionally, the observed frequencies for CoD SnD shown as a count $N(r)$ and as a proportion $\Phi(r)$ of all series ($\sum_{r \in R} N(r)$, where $r \in R = [6, 7, 8, 9, 10, 11]$).

r	$N(r)$	$\Phi(r)$	$\hat{\Phi}_0(r)$
6	40	4.7%	3.1%
7	101	11.9%	9.4%
8	141	16.5%	16.4%
9	185	21.7%	21.9%
10	183	21.5%	24.6%
11	202	23.7%	24.6%

Table 3: Alternate estimates of the constant probability (ϕ) for winning a given round in a CoD SnD, applying the three methods suggested by Mosteller (1952), in addition to the naive ($\phi_0 = 0.5$).

Method	ϕ	χ^2 (p-value)
0. Naive	0.5000	16.0 (≤ 0.01)
1. Method of moments	0.5725	3.6 (0.6)
2. Maximum likelihood	0.5750	3.5 (0.62)
3. Minimum (χ^2)	0.5775	3.5 (0.62)

Table 4 shows the new $\hat{\Phi}(r)$ when re-applying Equation 2 for the maximum likelihood estimate $\phi_2 = 0.575$, resulting in a new set of expected proportions of series lasting r rounds $\hat{\Phi}_2(r)$.[17] We observe that $\hat{\Phi}_2(r)$ is larger than $\hat{\Phi}_0(r)$ for $r \in [6, 7]$, more closely matching $\Phi(r)$. $\hat{\Phi}_2(r)$ is also closer to the observed $\Phi(r)$ for $r \in [9, 10]$, although not for $r \in [8, 11]$.

Observing that the null hypothesis acceptance shown in Table 3, along with the observation that $\hat{\Phi}_2(r)$ reasonably matches $\Phi(r)$ (in comparison to $\hat{\Phi}_0(r)$), we can say that the constant round win probability assumption is valid in CoD SnD series with the appropriate choice of ϕ (≈ 0.575).

4.2 Momentum

4.2.1 Post-streak probability

Given that people typically perceive streaks as beginning after the third success (or failure) at minimum (Carlson and Shu 2007), we focus on streaks of three round wins.[18] Table 5 compares the notional and MS expected proportions, $\hat{P}_0^{+|k,r}$ and $\hat{P}_{MS}^{+|k,r}$ respectively, with the observed round win proportion, $P^{+|kr}$, following streaks of $k = 3$ round wins given the series outcome.

---

[17]The method of moments and minimum χ^2 estimates for ϕ are omitted simply because the results would be nearly identical to those for the maximum likelihood estimate of ϕ (since they are all ≈ 0.575).

[18]Three happens to also be a reasonable number for series that last at maximum 11 rounds.

Table 4: The observed proportion of CoD SnD series, $\Phi(r)$, ending in r rounds, compared to the expected proportion, $\hat{\Phi}_0(r)$, under the naive assumption $\phi_0 = 0.5$ and the expected proportion, $\hat{\Phi}_2(r)$, under the maximum likelihood estimate, $\phi_2 = 0.575$, for constant round win probability.

r	$\Phi(r)$	$\hat{\Phi}_0(r) = 0.5$	$\hat{\Phi}_2(r) = 0.575$
6	4.7%	3.1%	4.2%
7	11.9%	9.4%	11.2%
8	16.5%	16.4%	17.8%
9	21.7%	21.9%	21.8%
10	21.5%	24.6%	23.0%
11	23.7%	24.6%	22.0%

Table 5: Given the round win streak $k = 3$, the length of the series (r rounds), and the series winner, the observed count of rounds wins, $r^{+|k=3,r}$, and proportion of round wins, $P^{+|k=3,r}$, among $N^{k=3,r}$ instances where a team could win after the streak ($P^{+|k=3,r} = \frac{r^{+|k=3,r}}{N^{k=3,r}}$). Additionally, the notional and MS expected proportions, $\hat{P}_0^{+|k=3,r}$ and $\hat{P}_{MS}^{+|k=3,r}$ respectively.

r	Win series?	$r^{+|k=3,r}$	$N^{k=3,r}$	$P^{+|k=3,r}$	$\hat{P}_0^{+|k=3,r}$	$\hat{P}_{MS}^{+|=k=3,r}$
7	yes	156	209	74.6%	75.0%	75.7%
8	yes	130	209	62.2%	60.0%	61.9%
9	yes	100	193	51.8%	50.0%	52.7%
10	no	8	60	13.3%	14.3%	26.7%
10	yes	66	151	43.7%	42.9%	44.7%
11	no	31	129	24.0%	25.0%	30.8%
11	yes	60	150	40.0%	37.5%	38.8%

With the exception of $\hat{P}_{MS}^{+|k,r}$ when $r = 10$ and the eventual series loser is the team that wins after a streak of three round wins, all un-adjusted binomial test p-values are greater than the $\alpha = 0.05$ confidence level, implying that we cannot reject the null hypothesis that the expected notional and MS post-streak round win proportions are different than the observed proportion. However, after adjusting the p-values with the Benjamini and Yekutieli (BY) correction (2001), the null hypothesis cannot be rejected for any case.[19]

When performing the same tests for streaks of two, four, and five, there is no case in which we can reject the binomial null hypothesis for the expected notional proportion $\hat{P}_0^{+|k,r}$ (even before the BY p-value correction). The null hypothesis can only be rejected for the expected MS proportion $\hat{P}_{MS}^{+|k,r}$ when $k = 5, r = 7$ after applying the BY p-value correction.[20]

Now let us explicitly consider the round, i. As shown in Table 6, the notional win proportion in the round i immediately following a streak of k round wins in a series lasting r rounds, $\hat{P}+|k,r,i_0$, is statistically different than the observed $P^{+|k,r,i}$ in several cases, notably when $r = i$ for each of $r \in [9, 10, 11]$.[21] Interestingly, the observed rate is greater than the expected rate in each case where we can reject the null hypothesis, whereas the observed rate is less than the expected rate in all other cases, as well in a strong majority of cases not shown.

4.2.2 Wald-Wolfowitz runs test

In Table 7, the observed and expected count of sequences, $N^{\zeta r}$ and $\hat{N}^{\zeta r}$ respectively, are shown for sequences, $\zeta(r)$, for which one can reject the Wald-Wolfowitz null hypothesis, given probability of Bernoulli trial success $\phi = 0.575$,[22] at a confidence level of $\alpha = 0.05$, for $r \in [8, 9, 10]$. In addition to the 13 sequences shown, there are 12 additional sequences for $r = 11$.

The relative frequency of the expected proportions are based on 10,000 simulations using the constant round probability $\phi_2 = 0.575$. A test for the difference between the observed and expected proportions of all sequences, $P^{\zeta r}$ and $\hat{P}^{\zeta r}$ respectively, indicates that the null hypothesis—that the two proportions are equal—cannot be rejected for any of the 25 significant sequences.

5 Discussion

Anecdotally, most observers swear by the existence of momentum in CoD SnD series, to the extent that vernacular has been developed to describe such phenomenon. Viewers have come to embrace the "5-3" phenomenon, where teams win three consecutive rounds after facing a 5-3 deficit to win 6-5. There is even a term for the rare 0-5 comeback—a "full sail".

---

[19]As noted before, the MS expectations may be unreliable for CoD SnD, so one is inclined to prefer the results of the notional binomial tests.

[20]At 36, the sample size for $k = 5, r = 7$ is not as large as most other combinations of k and r.

[21]The null can also be rejected for $r \in [7, 8]$, although these are not shown, and the sample sizes are smaller.

[22]If we used $\phi = 0.5$, we would get the same exact seqences.

Table 6: Given the round win streak $k = 3$, the index of the round immediately following the streak i, the length of the series (r rounds), and the series winner, the observed count of rounds wins, $r^{+|k=3,r,i}$, and proportion of round wins, $P^{+|k=3,r,i}$, among $N^{k=3,r,i}$ instances where a team could win after the streak. Additionally, the notional proportion, $\hat{P}_0^{+|k=3,r,i}$. Table restricted to $r \in [10, 11]$, $i < r$, and $N^{k=3,r,i} \geq 10$ for brevity.

r	i	Win series?	$r^{+|k=3,r,i}$	$N^{k=3,r,i}$	$P^{+|k=3,r,i}$	$\hat{P}_0^{+|k=3,r,i}$
10	7	no	2	13	15.4%	14.3%
10	7	yes	11	13	84.6%	42.9%
10	8	no	1	10	10.0%	14.3%
10	8	yes	9	10	90.0%	42.9%
11	5	no	9	13	69.2%	25.0%
11	5	yes	4	13	30.8%	37.5%
11	6	no	4	13	30.8%	25.0%
11	6	yes	9	13	69.2%	37.5%
11	7	no	3	10	30.0%	25.0%
11	7	yes	7	10	70.0%	37.5%
11	9	no	5	15	33.3%	25.0%
11	9	yes	10	15	66.7%	37.5%
11	10	no	5	10	50.0%	25.0%
11	10	yes	5	10	50.0%	37.5%

Table 7: Sequences ($\zeta(r)$) for which we can reject Wald-Wolfowitz null hypothesis for $r \in [8, 9, 10]$, where round wins and losses are denoted with $+$ and $-$ respectively. The observed count, $N^{\zeta r}$, and proportion, $P^{\zeta r}$, of all possible sequences, as well as the expected count, $\hat{N}^{\zeta r}$, and proportion, $\hat{P}^{\zeta r}$. $\hat{N}^{\zeta r}$ is scaled to the observed number of series played, hence its non-integer value.

r	$\zeta(r)$	$N^{\zeta r}$	$P^{\zeta r}$	$\hat{N}^{\zeta r}$	$\hat{P}^{\zeta r}$
8	- - + + + + + +	4	0.47%	7.67	0.90%
9	- - - + + + + + +	4	0.47%	3.75	0.44%
10	+ + + + - - - - + +	1	0.12%	2.39	0.28%
10	+ - - - - + + + + +	0	0.00%	2.04	0.24%
10	+ - + - + + - + - +	3	0.35%	2.04	0.24%
10	+ + + - - - - + + +	2	0.23%	2.04	0.24%
10	- - - - + + + + + +	2	0.23%	1.79	0.21%
10	+ - + - + - + + - +	3	0.35%	1.70	0.20%
10	+ - + + - + - + - +	0	0.00%	1.70	0.20%
10	+ + - - - - + + + +	1	0.12%	1.19	0.14%
10	+ - + - + - + - + +	3	0.35%	1.11	0.13%
10	+ + - + - + - + - +	0	0.00%	1.11	0.13%
10	+ + + + + - - - - +	2	0.23%	1.02	0.12%

However, our results do not provide evidence for momentum on several fronts: (1) the evidence supporting the constant round probability; (2) the failure to reject the binomial null hypothesis for post-streak win proportions; and (3) the failure to reject the equal proportions null hypothesis for streaky round sequences identified by the Wald-Wolfowitz runs test.

Perhaps it is not surprising that we did not find evidence in favor of the hot hand effect given the small "skill gap" in the CoD relative to other esports. (Most professional esports players who have played CoD, including CoD players themselves, would not hesitate to state this.) A small skill gap fosters randomness in outcomes, implying that a given team is less likely to enjoy streaks of success in SnD.

For the sake of brevity, we did not account for offensive or defensive role when considering win proportions during streaks. For example, it would be interesting to look at whether the round win percentage is higher after a streaks of three round wins when the team starts the streak as the offensive team. This would mean that, in the round immediately following the latest streak win, the streaking team would be playing defense, where teams are slightly more likely to win on average. On the other hand, streaks of three starting with an offensive round win—theoretically, knowing that $\bar{\tau}_O = 47.7\%$, the expected frequency is $\bar{\tau}_O^2(1 - \bar{\tau}_O) \approx 11.9\%$—are likely to occur less frequently than such streaks starting with a defensive win—$\bar{\tau}_O(1 - \bar{\tau}_O)^2 \approx 13.0\%$—so we would want to account for sample size differences.

In the future, we could account for the quality of the teams, both in general and on specific maps. Even if doing so does not change the results, we could gain additional insight into why we do not observe statistically significant streakiness.

Further regarding future work, we have contacted the Twitter user "R11stats", who privately tracks in-round player engagements.[23] R11stats expressed intent on making the data public, which would allow for research into player-specific momentum. While it seems there is no team-level hot hand effect in CoD SnD, perhaps there is with eliminations performed by individual players.

References

Ayton, Peter, and Ilan Fischer. 2004. "The Hot Hand Fallacy and the Gambler's Fallacy: Two Faces of Subjective Randomness?" *Memory & Cognition* 32 (8): 13691378.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *Annals of Statistics*, 11651188.

Carlson, Kurt A., and Suzanne B. Shu. 2007. "The Rule of Three: How the Third Event Signals the Emergence of a Streak." *Organizational Behavior and Human Decision Processes* 104 (1): 113121.

Chance, Don. 2020. "Conditional Probability and the Length of a Championship Series in Baseball, Basketball, and Hockey." *Journal of Sports Analytics* 6 (2): 111–27. https://doi.org/10.3233/JSA-200422.

DeRover, DeMars. 2021. "Round Win Probabilities Based on Who's Alive + Time." https://www.reddit.com/r/VALORANT/comments/n3lpoo/round_win_probabilities_based_on_whos_alive_time/.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. "The Hot Hand in Basketball: On the Misperception of Random Sequences." *Cognitive Psychology* 17 (3): 295314.

Miller, Joshua B., and Adam Sanjurjo. 2018. "Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers." *Econometrica* 86 (6): 20192047.

Mosteller, Frederick. 1952. "The World Series Competition." *Journal of the American Statistical Association* 47 (259): 355380.

Steeger, Gregory M., Johnathon L. Dulin, and Gerardo O. Gonzalez. 2021. "Winning and Losing Streaks in the National Hockey League: Are Teams Experiencing Momentum or Are Games a Sequence of Random Events?" *Journal of Quantitative Analysis in Sports* 17 (3): 155170.

Vafa, Keyon. 2017. "Is the Hot Hand Fallacy a Fallacy?" https://keyonvafa.github.io/hot-hand/.

Xenopoulos, Peter, William Robert Freeman, and Claudio Silva. 2022. "Analyzing the Differences Between Professional and Amateur Esports Through Win Probability." In, 34183427.

---

[23] https://twitter.com/R11stats