THE HOT HAND FALLACY IN CALL OF DUTY SEARCH AND DESTROY

A PREPRINT

Person One

personone@domain.com

Person Two

persontwo@domain.com

August 29, 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.

***K*eywords** 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:

- 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.

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

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.

Defense round wins	Offense 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)

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. Note that SnD is not the only game mode played in competitive CoD.[4] 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

The data set consists of all SnD matches played in major tournaments and qualifiers during the past three years[5], totaling 7,792 rounds across 852 series.[6] Data has been collected in spreadsheets by community member "IOUTurtle".[7]

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 season), "Cold War" (2021 season) and "Vanguard" (2022 season).

Sweeps (6-0 series) make up 4.7% of all series, while 6-5 series make up 23.7% of series. The observed offensive round win percentage[8] across all rounds is $\bar{\tau} = 47.8\%$.[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 $\bar{\tau}$.

The professional competition format balances the frequency with which teams play offense or defense to start a series, to the extent possible. Thus, win percentages in certain states are subjected to minimal selection bias, i.e. bias due to better teams playing defense more frequently.[10]

---

[4]Hardpoint has been played as the first and fourth game modes in a matchup. The third game mode was Domination in the 2020 season, and Control in the 2021 and 2022 seasons.

[5]Since the release of "Call of Duty: Modern Warfare" in fall 2019, professional CoD has been orchestrated by the 12-franchise CoD League (CDL). The CDL has completed three year-long "seasons" as of August 2022.

[6]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.

[7]Raw data: `https://linktr.ee/CDLArchive`. Cleaned data: `https://github.com/%7Bauthor%7D/%7Brepo%7D/blob/master/data/%7Bfile%7D.csv`.

[8]We use the terms "percentage", "probability", and "proportion" synonymously throughout this paper when discussing round win rates. For the most part, we restrict "percentage" usage to observed values and "probability" usage to expected values, and we use "proportion" to refer to either observed or expected values.

[9]Offensive round win percentage has been nearly constant across the three games during the CDL era: 47.2% in Modern Warfare; 47.9% in Cold War; and 48.1% in Vanguard

[10]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. Anecdotally, the better team tends to choose to play defense in the first round of the SnD

Outside of tournaments, teams 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.

This format has minimized the range of rates at which teams start series playing offense. The highest rate at which a team has started their series playing defense across all tournament and qualifier SnD series in a single season is 66% (31 of 47). The lowest is 38% (16 of 42).

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, and National Football League, although did in the NBA. As a check on their entropy approach to momentum identification, Steeger et al. found several NHL teams with sequences 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:

---

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.

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

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. (Counter-Strike has a 15-15-1 format and Valorant has a 12-12-1 format.) On the other hand, teams in CoD SnD rotate attacking and defending roles every round, minimizing the effect of map asymmetry on streaks of wins.[12]
3. Both Counter-Strike and Valorant have overtime rules—a 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 investigated—particularly one that is less subjected to factors that may be difficult to control for, e.g. weather—should be useful as an inspiration for future researchers.

3 Methodology

3.1 Distribution of rounds played

In a best-of-s format, assuming a constant round win probability ϕ (where $0 \leq \phi \leq 1$) for a given team, the expected proportion of series ending in r rounds (where $r \leq s$) 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 exactly nine rounds in CoD SnD, where $s = 11$ (implying $m = 6$), 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

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

where $\hat{\Phi}(r)$ and $\Phi(r)$ represent the expected and observed round win proportions respectively, and where $R = [6, 7, 8, 9, 10, 11]$ 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 use a significance level of $\alpha = 0.05$ for this test and all others in this paper.

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 expect the win proportion for a round to be

$$\hat{P}_0(\text{win}|r) = \hat{P}_0^{+|r} := \begin{cases} \frac{m}{r} & \text{team wins series,} \\ \frac{r-m}{r} & \text{team loses series,} \end{cases} \tag{4}$$

---

[12]Professional CoD has a set of maps (at least three, depending on the season) on which teams choose to play SnD. Most are asymmetric, introducing slight advantages for the offense or defense that do not exist on other maps. Counter-Strike and Valorant also have their own sets of maps for their analogous SnD modes.

for $r \in R$ and 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 ending in r rounds, $\hat{P}_{MS}^{+|k,r}$, to be strictly less than the observed proportion of round wins given the series length, but not given knowledge of the potential existence of a streak, $P^{+|r}$, i.e. the analogue of the expected proportion in Equation 4 for the observed data. However, our context is fundamentally different from that of MS, who evaluate data in controlled settings, unaffected by opposition.[13]

The number of trials is fixed in their experimental designs; in CoD SnD, however, 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 single round win percentage in CoD SnD—analogous to a player's shooting percentage in MS's analysis of basketball players—is not independent of the opponent. Consequently, the i.i.d. assumption of MS's theorem regarding post-streak success may or may not be valid. Nonetheless, we still leverage $\hat{P}_{MS}^{+|k,r}$ as a reference for our results.

3.2.2 "Notional" post-streak probability

Let us consider another form of the expected proportion of rounds won immediately after a streak of k round wins in a best-of-s series ending in r rounds: the "notional" proportion

$$\hat{P}_0(\text{win}|k,r) = \hat{P}_0^{+|k,r} := \begin{cases} \frac{m-k}{r-k} & \text{team wins series}, r > m; \\ 1 & \text{team wins series}, r = m; \\ \frac{s-m-k}{r-k} & \text{team loses series}, r > m + k; \\ 0 & \text{team loses series}, r = m + k, \end{cases} \tag{5}$$

where $r \in R$ and m from Equation 1.

We can perform a two-tailed binomial test to evaluate the null hypothesis $\omega = \omega_0$ for the observed proportion of successes, ω, and a null proportion, ω_0 (where $0 \le \omega_0 \le 1$). The probability of exactly r^+ successes (and $r^- = r - r^+$ failures) in a sample of r trials is

$$\binom{r}{r^+} \omega^{r^+} (1-\omega)^{r^-}, \tag{6}$$

where $\omega = \frac{r^+}{r}$ is the observed proportion of successes. If we can reject the null hypothesis—treating $P^{+|k,r}$ and $\hat{P}_0^{+|k,r}$ as ω and ω_0 respectively—then we might consider team momentum plausible. Since we perform this test for many combinations of k and r, we adjust the p-values with the Benjamini and Yekutieli (BY) false discovery procedure (2001).

As a reference, we can perform the same binomial test for MS's proportion, $\hat{P}_{MS}^{+|k,r}$, which adjusts for streak selection bias. However, given the caveats mentioned before regarding the application of 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 $k < i \le r$) in which the streak of length k carries into.

$$\hat{P}_0(\text{win}|k,r,i) = \hat{P}_0^{+|k,r,i} := \begin{cases} \frac{m-k}{r-k} & \text{team wins series}, i \ne r; \\ 1 & \text{team wins series}, i = r; \\ \frac{s-m-k}{r-k} & \text{team loses series}, i \ne r, r > m + k; \\ 0 & \text{team loses series}, i = r. \end{cases} \tag{7}$$

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

---

[13]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 by player shot selection influenced by the opposing team's defensive strategy. This is similar to our setting.

3.2.3 Wald-Wolfowitz runs test

We can attempt to detect the hot hand phenomenon more broadly by focusing on non-random sequences identified by the Wald-Wolfowitz runs test. Note that the probability of success and failure need not be equal; thus, we can incorporate our findings regarding constant round win probability, specifying that the probability of success is $\phi_2 = 0.575$.

Under the null hypothesis for the Wald-Wolfowitz runs test, 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 (and r^- failures), with the following mean μ_r [14] and variance σ_r^2:

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

The null hypothesis is that the test statistic,

$$z_r = \frac{r^* - \mu_r}{\sigma_r}, \tag{9}$$

where r^* is the number of runs, is normally distributed.

We can subset the observed series sequences, ζ_j, to those that violate the null hypothesis for the runs test and perform a test of equal proportions, where the null hypothesis is that the observed proportion of a sequence relative to all possible sequences, P_j^ζ ($\frac{N_j^\zeta}{\sum_j N^\zeta}$), is equal to the expected proportion of the sequence relative to all possible sequences, $\hat{P}_j^\zeta$. The test statistic is

$$z_j^\zeta = \frac{P_j^\zeta - \hat{P}_j^\zeta}{\sqrt{P_{j,\delta}^\zeta (1 - P_{j,\delta}^\zeta)(1/N_j^\zeta + 1/\hat{N}_j^\zeta)}} \tag{10}$$

where N_j^ζ and $\hat{N}_j^\zeta$ respectively represent the observed and expected number of occurrences of sequence ζ_j, and where $P_{j,\delta}^\zeta = (P_j^\zeta - \hat{P}_j^\zeta)/(N_j^\zeta - \hat{N}_j^\zeta)$. The null hypothesis is that the test statistic is normally distributed. If we can reject the null hypothesis for such sequences, we can build an argument in support of the hot hand effect.[15]

4 Results

4.1 Distribution of rounds played

Using Equation 3, we find that $\chi^2 = 16.0$ (p-value of <0.01) for $\phi_0 = 0.5$. Thus, we can comfortably reject the constant probability hypothesis for the naive $\phi_0 = 0.5$.

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

Table 3 shows the alternate values for the constant round win probability assumption, each 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.

Table 4 shows the expected proportions of series ending in r rounds, $\hat{\Phi}_2(r)$, when re-applying Equation 2 for the maximum likelihood estimate $\phi_2 = 0.575$.[16] We observe that $\hat{\Phi}_2(r)$ more closely matches $\Phi(r)$ than $\hat{\Phi}_0(r)$.

---

[14]We calculate μ and σ^2 for each number of rounds $r \in R$.

[15]As with the binomial tests for the notional and MS post-streak proportions, we adjust the p-values with the BY correction.

[16]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).

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

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 round win probability (ϕ) and their corresponding χ^2 values, applying the three methods suggested by Mosteller to CoD SnD. Additionally, the naive constant round win probability ϕ_0 and its χ^2.

Method	ϕ	χ^2 (p-value)
ϕ_0: Naive	0.5000	16.0 (≤ 0.01)
ϕ_1: Method of moments	0.5725	3.6 (0.6)
ϕ_2: Maximum likelihood	0.5750	3.5 (0.62)
ϕ_3: Minimum χ^2	0.5775	3.5 (0.62)

The observation that $\hat{\Phi}_2(r)$ reasonably matches $\Phi(r)$ supports the implication from the failure to reject the null hypothesis for ϕ_2 (shown in Table 3): the constant round win probability assumption can be valid for CoD SnD 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. Table 5 compares the observed round win proportion, $P^{+|k=3,r}$, with the notional proportion, $\hat{P}^{+|k=3,r}$, following streaks of $k = 3$ round wins. The MS proportion, $\hat{P}_{MS}^{+|k=3,r}$, is also shown as a reference.

All binomial test p-values for both $\hat{P}_0^{+|k=3,r}$ and $\hat{P}_{MS}^{+|k=3,r}$ are found to be insignificant at the $\alpha = 0.05$ threshold, implying that we cannot reject the null hypothesis that the notional and MS post-streak round win proportions are different than the observed proportion for streaks of three.[17] When performing the same

---

[17]p-values are not shown for readability purposes.

Table 4: The observed proportion of CoD SnD series ending in r rounds, $\Phi(r)$, compared to the expected proportion, $\hat{\Phi}_2(r)$, under the maximum likelihood estimate, $\phi_2 = 0.575$. Additionally, the expected proportion, $\hat{\Phi}_0(r)$, under the naive assumption $\phi_0 = 0.5$ is repeated from Table 2 for comparison.

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: For streaks of $k = 3$ round wins in CoD SnD series ending in r rounds, the observed count of rounds wins, $r^{+|k=3,r}$, and proportion of round wins, $P^{+|k=3,r}$, among $N^{k=3,r}$ chances ($P^{+|k=3,r} = r^{+|k=3,r}/N^{k=3,r}$). 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%

Table 6: For streaks of $k = 3$ round wins in CoD SnD series ending in r rounds, the observed count of rounds wins, $r^{+|k=3,r,i}$, and proportion of round wins, $P^{+|k=3,r,i}$, among $N^{k=3,r,i}$ chances, given the index of the round, i, immediately following the streak. Additionally, the notional proportion, $\hat{P}_0^{+|k=3,r,i}$. Table restricted to $r \in [10, 11]$, $i < r$, and $N^{k=3,r,i} \geq 10$.

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%

tests for streaks of two, four, and five, there is no case in which we can reject the null hypothesis for $\hat{P}_0^{+|k,r}$. The null hypothesis can only be rejected in some extreme cases for $\hat{P}_{MS}^{+|k,r}$.[18]

Table 6 shows the observed and notional proportions of round wins immediately following streaks of three round wins when explicitly accounting for the round index, i.[19]

4.2.2 Wald-Wolfowitz runs test

In Table 7, the observed and expected count of sequences respectively, are shown for sequence for which we can reject the Wald-Wolfowitz null hypothesis[20] at a confidence level of $\alpha = 0.05$, for $r \in [6, 7, 8, 9, 10]$[21]. In addition to the 14 sequences shown, there are 12 additional sequences for $r = 11$.

The expected proportion of occurrences of each sequence ζ_j is 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

---

[18]These cases are: (1) $k = 4$, $r = 7$ ($N^{k,r} = 108$); (2) $k = 4$, $r = 10$ ($N^{k,r} = 45$); (3) $k = 5, r = 7$ ($N^{k,r} = 37$); (4) $k = 5, r = 8$ ($N^{k,r} = 4$). In each case, the team won the series.

[19]We omit the MS proportions $\hat{P}_{MS}^{+|k=3,r,i}$ since sample sizes are small and results may or may not be valid due to the questionable i.i.d assumption.

[20]Interestingly, we identify the same non-random sequences when using either $\phi = 0.575$ or $\phi = 0.5$.

[21]Since the normal method is invalid for sequences with only successes or failures, we use the exact method to evaluate (and reject) the null hypothesis for the Wald Wolfowitz runs test for $r = 6$.

Table 7: Sequences (ζ_j) for which we can reject Wald-Wolfowitz null hypothesis for $r \in [6, 7, 8, 9, 10]$, where round wins and losses are denoted with $+$ and $-$ respectively. The observed count, N_j^ζ, and proportion, P_j^ζ, of all possible sequences, as well as the expected count, $\hat{N}_j^\zeta$, and proportion, $\hat{P}_j^\zeta$. $\hat{N}_j^\zeta$ is scaled to the observed number of series played, hence its non-integer value.

r	ζ_j	N_j^ζ	P_j^ζ	$\hat{N}_j^\zeta$	$\hat{P}_j^\zeta$
6	$+\ +\ +\ +\ +\ +$	40	4.7%	32.9	3.86%
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%

of all sequences, P_j^ζ and $\hat{P}_j^\zeta$ respectively, indicates that the null hypothesis—that the two proportions are equal—cannot be rejected for any of the non-random 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".

However, our results do not provide evidence for momentum on several fronts: (1) the evidence supporting the constant round probability hypothesis; (2) the failure to reject the binomial null hypothesis for post-streak win probabilities; and (3) the failure to reject the equal proportions null hypothesis for non-random 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 go on streaks of consecutive round wins in SnD, compared to the counterfactual setting where the skill gap is "large" and better teams consistently defeat worse teams.

For the sake of brevity, we did not account for offensive or defensive role when considering win proportions in the rounds immediately following streaks. It would be interesting to look at whether the round win percentage is higher after streaks of round wins when the team starts the streak as the offensive team. Considering streaks of three, this would mean that the streaking team would be playing defense in the round immediately following the last streak win. Knowing that teams are more likely to win on defense—recall $\bar{\tau} = 47.8$—we should expect that teams on such streaks are more likely to win the round than teams that go on streaks of three starting with a defensive round win. On the other hand, streaks of three starting with an offensive round win are less likely to occur than such streaks starting with a defensive win—$\bar{\tau}^2(1 - \bar{\tau}) \approx 11.9\%$ in the former case and $\bar{\tau}(1 - \bar{\tau})^2 \approx 13.0\%$ in the latter case—so we would need to properly account for sample size differences.

Another idea would be to account for the quality of the teams, both in general and on specific maps. Even if we still cannot prove that the hot hand effect exists when doing so, a null finding would be insightful in and of itself.

Further, we have contacted the Twitter user "R11stats", who privately tracks in-round player engagements. Data regarding individual eliminations during each SnD round would facilitate research into player-specific momentum. While it seems there is no team-level hot hand effect in CoD SnD, perhaps there is with 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.

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