ALWAYS-ON EXPERIMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative AI has dramatically accelerated the rate at which new treatments—from novel pharmaceuticals to online marketing campaigns—can be conceived and deployed. As a result, modern experimentation platforms often run continuously, with treatments added as they are ready and removed when they underperform. We formalize this "Always-On" experimental setting, in which treatments can be dynamically generated, added to, and removed from a running experiment, and study the statistical problem of deciding whether to accept or reject each treatment while controlling for the false discovery rate. We develop sequential tests that achieve time-uniform Type-I error control under arbitrary stopping times and "predictable" treatment schedules. Our approach builds on the testing-by-betting framework: we construct test supermartingales for testing the average treatment effect of each treatment, and show that the construction of these test supermartingales is growth-rate optimal in an almost-sure sense.

1 INTRODUCTION

Generative AI has the promise to accelerate the rate of scientific discovery. It has reduced the barrier to the conception of new treatments, making it easier for practitioners to generate new hypotheses "on the fly." Existing experimentation frameworks, however, have not been designed for an era in which treatments can be easily and dynamically generated. Rather, to guarantee statistical validity existing methods typically require practitioners to know all treatments prior to starting the experiment and do not afford practitioners the freedom to continuously monitor the experiment in order to decide when to remove a treatment. In this work we introduce *Always-On Experimentation*, a statistical design framework suited for the era in which the conception of treatments is no longer the bottleneck in the experimentation pipeline.

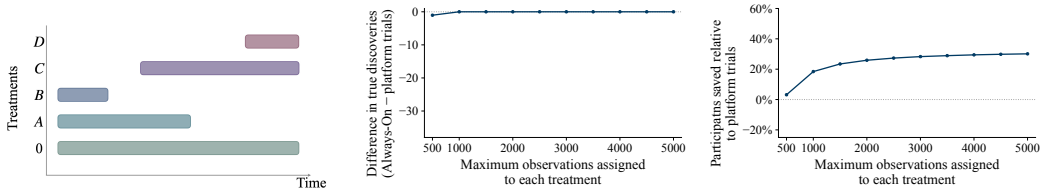

Figure 1: *(Left)* Depiction of an instance of Always-On experimentation in which four treatments were added and removed at different points in the experiment. *(Middle & Right)* Always-On experimentation achieves the same level of power as platform trials while reducing the number of units needed to perform such discoveries. For information about simulation setup and further simulations, see Section A.1.

Our Always-On design is based on simple desiderata: practitioners should have the flexibility to add and remove treatments at any point in time from a continuously running experiment. Such flexibility should be coupled with the ability to *continuously monitor* the experiment and decide to introduce or remove treatments in a data-driven manner, all while maintaining statistical validity. In our case, statistical validity consists of two components: controlling the type-I error for each treatment that has been introduced to the experiment, for the entire duration of the experiment; and safeguarding against multiple comparisons, since potentially multiple treatments will be tested within the same experiment. Our paper formalizes the Always-On design and describes how it satisfies these desiderata while maintaining statistical validity.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The rest of the paper is organized as follows. Section 2 connects the Always-On design to the existing experimental designs literature. Section 3 presents our Always-On design. Section 3.1 formalizes the design and instantiates it within the testing-by-betting framework. Section 3.2 describes how we tackle the multiple testing problem within our Always-On framework. Section 4 instantiates the Always-On framework in a setting where experimental units are only observed once for the entire duration of the experiment. We describe the target estimands and our estimators within this setting, and we present a strategy to construct "powerful" test statistics that enjoy an almost sure and asymptotic notion of growth-rate optimality. In Section 5 we present the performance of Always-On experimentation in a semi-synthetic experiment.

2 Related Work

Platform trials. Platform trials from biomedical research (Berry et al., 2015; Woodcock & LaVange, 2017; Santacatterina et al., 2025) focus on evaluating a range of interventions for a single disease or condition. Like Always-On experimentation, platform trials facilitate the addition and removal of interventions from a continuously running experiment. Platform trials, however, typically operate under a master protocol that prespecifies the timing or rules governing interim analyses and adaptations, including when treatment arms may be stopped or added (see, for example, Adaptive Platform Trials Coalition, 2019). By contrast, Always-On experimentation permits continuous monitoring and allows treatments to be added or removed at arbitrary times, provided these decisions depend only on past data.

Online multiple testing. Our Always-On design is a particular instantiation of the general online multiple testing framework introduced by Foster & Stine (2008). Most work on online multiple testing (e.g., (Javanmard & Montanari, 2015; 2018; Ramdas et al., 2018; Tian & Ramdas, 2019; Xu & Ramdas, 2024)) has considered a setting in which hypotheses are tested sequentially; that is, a hypothesis must either be accepted or rejected before a new hypothesis is tested. Zrnic et al. (2021) and Fischer et al. (2024) relaxed the synchronous testing requirement and considered settings in which hypotheses need not be accepted or rejected before others are introduced, and they provided algorithms to control the online false discovery rate (or other related error metrics). Our framework's multiple testing setting is most similar to that of the latter two works as treatments can be added to the experiment at any point to the experiment and their futility need not be determined before other treatments are introduced. Rather, treatment arms can remain active throughout the experiment until enough evidence is gathered to declare the treatment a discovery or a decision is made to remove the treatment from the experiment. Robertson et al. (2023) and Zehetmayer et al. (2022) study how existing algorithms that control the online false discovery rate can be used within platform trials and evaluate their performance empirically.

Online experimentation and multi-armed bandits. Modern experimentation platforms at technology companies run experimentation continuously (Kohavi et al., 2020), motivating our Always-On framework. A common approach is to run a single experiment that consists of multiple arms, typically referred to as A/B/n tests. Bartroff & Song (2020) focus on a setting in which data arrives sequentially and are interested in testing a predetermined set of hypotheses while controlling for the false discovery rate. They extend the procedure due to Benjamini & Hochberg (1995) to a setting in which data arrives sequentially. Jamieson & Jain (2018) consider a setup in which they get to dynamically sample from the available treatment arms, with the goal of minimizing the number of samples needed to exceed threshold on the proportion of true discoveries while controlling for the false discovery rate. Multi-armed bandits are another common approach for testing multiple treatments simultaneously in online platforms. The focus of these tests is to allocate units dynamically to the best performing arm in order to minimize (or maximize) some metric such as regret. Yang et al. (2017) propose a framework in which best-arm multi-armed bandits are sequentially tested with the aim of identifying the best arm within each instance while controlling the online false discovery rate. Most recently, Sandoval et al. (2026); Imbens et al. (2026) and Bharti et al. (2026) study a sequential hypothesis testing setting in which data can be sampled from multiple arms, and it is of interest to determine quickly whether any of the arms is non-null.

3 ALWAYS-ON EXPERIMENTATION

We study a design that affords the flexibility to add and remove treatments from a running experiment. We introduce the construct of *sub-experiments*, which we define to be those points in time in which treatments are added and removed from the running experiment, and note that the total number of sub-experiments $t \in \mathbb{N}$ need not be known a priori. Nonetheless, throughout this paper we assume the total number of sub-experiments is finite, meaning that $t < \infty$. In each sub-experiment $s \in [t] \coloneqq \{1, \dots, t\}$, the statistician tests the treatments belonging to the *set of active treatments* $\mathcal{A}_s$, and we assume that the control—which we denote as 0—is always present in each sub-experiment.[1] In other words, in each sub-experiment $s \in [t]$ the statistician has access to the *finite* set $\mathcal{A}_s \cup \{0\}$, where $|\mathcal{A}_s \cup \{0\}| < \infty$. We denote the set of treatments that were at some point active in the experiment (that consisted of t sub-experiments) as $\mathcal{A} \coloneqq \bigcup_{s=1}^{t} \mathcal{A}_s$. Since not all treatments are active in the same set of sub-experiments, we let $\mathcal{T}_t(a) \coloneqq \{s \in [t] \mid a \in \mathcal{A}_s\}$ be the set of sub-experiments in which treatment $a \in \mathcal{A}$ was active.

We assume units enter the experiment sequentially, and denote as $\mathcal{I}_s$ the set of units that participated in the experiment during sub-experiment $s \in [t]$ whose size we represent as $n_s \coloneqq |\mathcal{I}_s|$. We remark that the sets $\mathcal{I}_s$ could consist of different or the same units—settings that we consider in Section 4 and Section B, respectively. We use the notation $\mathcal{I}_s(j)$ to denote the set of the first $j \in [n_s]$ units whom have entered sub-experiment s. Furthermore, we denote the set of all units that have been part of the experiment starting from sub-experiment one up to sub-experiment s as $\mathcal{I}_{1:s} \coloneqq \bigcup_{r=1}^{s} \mathcal{I}_r$, and if $s = t$ (i.e., the last sub-experiment that has been run) we will use the shorthand $\mathcal{I} \equiv \mathcal{I}_{1:t}$. Lastly, we denote as $\mathcal{I}_{1:s}(j) \coloneqq \bigcup_{r=1}^{s-1} \mathcal{I}_r \cup \mathcal{I}_s(j)$ the set of all units that have appeared in sub-experiments one up to $t-1$ and those j units whom have entered sub-experiment s.

We adopt the potential outcomes framework (Neyman, 1923; Rubin, 1974) and assume each unit $i \in \mathcal{I}$ has potential outcomes $Y_i(\mathbf{z}) \in \mathbb{R}$ for each $\mathbf{z} \in \mathcal{Z}$, where $\mathcal{Z}$ denotes the set of all possible treatment paths. We will instantiate the set $\mathcal{Z}$ in the sections to come. Additionally, for the rest of the paper we make the following assumption on the potential outcomes.

Assumption 3.1. *We assume the potential outcomes for each unit are independently and identically distributed:* $(Y_i(\mathbf{z}))_{\mathbf{z} \in \mathcal{Z}} \sim \mathsf{P}$ *for each* $i \in \mathcal{I}$.

We emphasize that Assumption 3.1 is only with respect to how the units themselves are sampled and not on any putative dependence between their potential outcomes. Indeed, the potential outcomes $(Y_i(\mathbf{z}))_{\mathbf{z} \in \mathcal{Z}}$ for each unit can be highly dependent.

Because the set of active treatments differs across sub-experiments and to avoid temporal drift that might result from running the experiment over time, we perform a new series of treatment assignments in every sub-experiment. That is, in each sub-experiment $s \in [t]$ we randomly assign the treatments from the set $\mathcal{A}_s \cup \{0\}$ to those units in $\mathcal{I}_s$. Throughout, we denote the treatment assignment unit $i \in \mathcal{I}_s$ received in sub-experiment s as $A_{i,s} \in \mathcal{A}_s \cup \{0\}$. This treatment assignment induces the observed outcome for unit $i \in \mathcal{I}_s$, which we denote as $Y_{i,s}$, and defer its formal definitions to Sections B and 4. The filtration that arises in Always-On experimentation is with respect to the outcomes and treatment assignments that we have observed thus far in the experiment. More formally, we denote the filtration for sub-experiment $s \in [t]$ based on all the units that have been part of the experiment, including the j units that have entered sub-experiment s, as $\mathcal{F}_{j,s} \coloneqq \sigma\left(\left((A_{i,r}, Y_{i,r})_{i \in \mathcal{I}_r}, \mathcal{A}_r\right)_{r=1}^{s-1}, (A_{i,s}, Y_{i,s})_{i \in \mathcal{I}_s(j)}\right)$. We slightly overload notation, denoting the filtration with respect to observed outcomes and assignments from the previous $s-1$ sub-experiments as $\mathcal{F}_{s-1} \coloneqq \sigma\left(\left((A_i, Y_i)_{i \in \mathcal{I}_r}, \mathcal{A}_r\right)_{r=1}^{s-1}\right)$, with $\mathcal{F}_0 = \{\varnothing, \Omega\}$ as the trivial σ-algebra.

In addition to Assumption 3.1 we make the following assumptions on the treatment assignments.

Assumption 3.2 (Assumptions on the treatment assignments)**.** *Let* $A_{i,s} \in \mathcal{A}_s \cup \{0\}$ *denote the treatment assignment unit* $i \in \mathcal{I}$ *received in sub-experiment* $s \in [t]$*, and let* $\mathcal{Z}_{s:t}$ *denote the set of treatment paths in sub-experiment* s *up to* t*—with their respective potential outcomes being* $Y_i(\mathbf{z})$

[1] For simplicity of exposition we assume that all treatments within the experiment are compared against the same control, but our framework can be easily modified so that each sub-experiment can have a potentially different control.

*for each $\mathbf{z} \in \mathcal{Z}_{s:t}$ and $i \in \mathcal{I}$. We assume the following conditions hold for each sub-experiment $s \in [t]$ and unit $i \in \mathcal{I}$: (i) $A_{i,s} \perp\!\!\!\perp (Y_i(\mathbf{z}))_{\mathbf{z} \in \mathcal{Z}_{s:t}} \mid \mathcal{F}_{i-1,s}$ (*conditional ignorability*) ; (ii) for some deterministic constant $p_s \in (0,1)$, $\mathbb{P}_{\mathrm{RCT}_s}(A_{i,s} = a \mid \mathcal{F}_{s-1}) \geq p_s > 0$ for each $a \in \mathcal{A}_s \cup \{0\}$ (*positivity*), where $\mathbb{P}_{\mathrm{RCT}_s}(A_{i,s} = a \mid \mathcal{F}_{s-1})$ denotes the conditional probability of being assigned treatment $a \in \mathcal{A}_s$ when the treatment assignment is sampled from the randomization distribution represented by RCT_s.*

The assumptions from Assumption 3.2 are common in the causal inference literature, and are satisfied by design given that we have control over the treatment assignments in the setting we study.

We let $\boldsymbol{\pi}_s \coloneqq (\pi_s(a))_{a \in \mathcal{A}_s \cup \{0\}}$ be the *propensity score* over the set of active treatments and control in sub-experiment s, where we define $\pi_s(a) \coloneqq \mathbb{P}_{\mathrm{RCT}_s}(A_{i,s} = a \mid \mathcal{F}_{s-1})$ for each $a \in \mathcal{A}_s \cup \{0\}$. In addition to Assumption 3.2, which tells us that $\pi_s(a) \in (0,1)$ for each $a \in \mathcal{A}_s \cup \{0\}$, we assume that $\sum_{a \in \mathcal{A}_s \cup \{0\}} \pi_s(a) = 1$ and that the propensity score $\boldsymbol{\pi}_s$ is determined (i.e., fixed) at the beginning of each sub-experiment s and that it is $\mathcal{F}_{s-1}$-measurable. This means that the propensity score can be a function of the data seen in previous sub-experiments, but once the current sub-experiment starts it can no longer be modified (until the start of the next sub-experiment). Having defined the propensity score, we model the treatment assignment as follows: in every sub-experiment s we commit to drawing the treatment assignments from a multinomial distribution over the set of active treatments. Letting $\mathrm{RCT}_s \coloneqq \mathrm{Multinomial}_{\mathcal{A}_s \cup \{0\}}(\boldsymbol{\pi}_s)$, the treatment assignment for unit $i \in \mathcal{I}_s$ in sub-experiment s is drawn i.i.d as $A_{i,s} \sim \mathrm{RCT}_s$. To simplify the notation used throughout the rest paper, we write $\mathsf{P}_{\mathrm{RCT}} \coloneqq \mathsf{P} \otimes \mathrm{RCT}_1 \otimes \cdots \otimes \mathrm{RCT}_t$ to denote the joint law of the potential outcomes and treatment assignments under the induced experimental design.

In addition to our assumptions on the treatment assignments and propensity scores, we make the following assumptions for the treatment schedule.

Assumption 3.3 (Assumptions on the treatment schedule)**.** *We make the following assumptions on how the treatments are added and removed from the experiment: (i) Let $a \in \mathcal{A}_s$ be a treatment active in sub-experiment $s \in [t]$. Then, the decision to keep treatment a in sub-experiment $s+1$ is $\mathcal{F}_s$-measurable; (ii) Let $\mathcal{A}'_s$ be the set of new treatments to be added in sub-experiment s. We assume that the decision to add these treatments to the experiment is $\mathcal{F}_{s-1}$-measurable.*

Assumption 3.3 has to do with the information available to us when deciding to add and remove treatments, and it tells us that such decisions cannot depend on future information. Otherwise, these decisions can be arbitrary. For example, a decision to remove a treatment from the experiment could be based on the belief, formed from the evidence gathered, that such treatment will not yield a higher expected outcomes than any other active treatment in the experiment.

We formally state the data collection protocol of Always-On experimentation in Algorithm 2, which can be found in Section A. In the following section we introduce the type of hypotheses tests we consider, and describe how to construct sequential anytime valid test statistics.

3.1 Testing by betting in Always-On experimentation

Rather than pursuing accurate estimates of the average treatment effects, our goal is to create powerful test statistics that allow us to determine as quickly as possible whether each of the treatments added to the experiment is effective. In this section we provide an informal introduction to the concepts that are needed to obtain such statistics; for formal definitions see Sections B and 4.

For each of the active arms in sub-experiment $s \in [t]$, $a \in \mathcal{A}_s$, we are interested in testing whether their treatment effect $\psi_{\mathsf{P}}(a) \in \mathbb{R}$ is greater than some threshold $\delta \in \mathbb{R}$. We therefore run a hypothesis test for each of the active arms $a \in \mathcal{A}_s$, where under the null $\mathcal{P}(a)$ the treatment effect is less than or equal to the threshold δ while under the alternative $\mathcal{Q}(a)$ the treatment effect is greater than δ. That is, for each treatment $a \in \mathcal{A}$ the null and alternative hypotheses take the following form

$$\mathcal{P}(a) \coloneqq \{\mathsf{P} \mid \psi_{\mathsf{P}}(a) \leq \delta\} \quad \text{versus} \quad \mathcal{Q}(a) \coloneqq \{\mathsf{P} \mid \psi_{\mathsf{P}}(a) > \delta\}\ , \tag{1}$$

and we additionally define the sets $\mathcal{P} \coloneqq \bigcup_{a \in \mathcal{A}} \mathcal{P}(a)$ and $\mathcal{Q} \coloneqq \bigcup_{a \in \mathcal{A}} \mathcal{Q}(a)$.

Given that we are interested in testing each of the active arms, we construct a test supermartingale: a supermartingale that is almost surely nonnegative and whose expectation is upper bounded by one. We first introduce some additional notation. For sub-experiment $s \in \mathcal{T}_t(a)$ we let $\check{N}_a(s) \coloneqq$

$\sum_{i\in\mathcal{I}_s} \mathbb{1}\{A_{i,s} = a \text{ or } A_{i,s} = 0\}$ be the number of units in sub-experiment s which were assigned to the treatment $a \in \mathcal{A}_s$ or the control. Furthermore, we let $N_a(t) \coloneqq \sum_{s\in\mathcal{T}_t(a)} \check{N}_a(s)$ be the total number of units which have been assigned to either treatment $a \in \mathcal{A}$ or the control over those sub-experiments in which treatment a has been active. Let $\widehat{\psi}_{i,s}(a)$ be the estimator of the treatment effect of treatment $a \in \mathcal{A}_s$, $((\lambda_{i,s})_{i\in\mathcal{I}_s})_{s\in\mathcal{T}_t(a)}$ be a predictable process (which we also refer to as a *portfolio*), and for each sub-experiment s let $g_s : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a $\mathcal{F}_{s-1}$-measurable nondecreasing function that maps a real-valued variable into the positive reals. Having introduced the necessary concepts, we now present the form of the test statistics—to test the hypotheses given in (1)—that we consider for each treatment $a \in \mathcal{A}$ throughout the rest of the paper:

$$W_{N_a(t)} \coloneqq \prod_{s\in\mathcal{T}_t(a)} W_{\check{N}_a(s)} \quad \text{where} \quad W_{\check{N}_a(s)} \coloneqq \prod_{i\in\mathcal{I}_s} \Big(1 - \lambda_{i,s} + \lambda_{i,s} g_s(\widehat{\psi}_{i,s}(a))/g_s(\delta)\Big)^{\xi_{i,s}(a)}, \tag{2}$$

where we have used the shorthand $\xi_{i,s}(a) \coloneqq \mathbb{1}\{A_{i,s} = a \text{ or } A_{i,s} = 0\}$.

The test statistic from (2) pools together the information from those units which received the treatment $a \in \mathcal{A}$ and those which received the control across the different sub-experiments. The following lemma states that this test statistic is a test supermartingale; see Section D.1 for the proof.

Lemma 3.4. *Fix an arm $a \in \mathcal{A}$ and a distribution $\mathsf{P} \in \mathcal{P}(a)$. Suppose that $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}[g_s(\widehat{\psi}_{i,s}(a)) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1] = g_s(\psi_{\mathsf{P}}(a))$ for each $s \in [t]$ and that Assumptions 3.1 to 3.3 hold. Then, the process $W_{N_a(t)}$ forms a test $\mathcal{P}(a)$-supermartingale.*

This statistic (2) is a particular instantiation of a test supermartingale commonly used for testing the mean of bounded random variables (see for example (Waudby-Smith & Ramdas, 2024; Hendriks, 2021; Orabona & Jun, 2023; Ryu & Bhatt, 2024; Shekhar & Ramdas, 2023; Chugg et al., 2023; Podkopaev & Ramdas, 2023; Podkopaev et al., 2023; Ryu et al., 2025)) and which is in fact the only admissible test supermartingale for such problems (Clerico, 2025).

Lemma 3.4 together with Ville's inequality allow us to conclude that we are able to achieve time-uniform type-I error rate control when it comes to testing each individual treatment; see Equation (15) in Section C for a formal definition. However, in Always-On experimentation we are interested in testing multipe treatments $a \in \mathcal{A}$, which means that we are in a multiple testing setting and must also aim to control a metric that takes this into account. In the following section we describe how we address multiple testing in Always-On experimentation.

3.2 Multiple testing in Always-On experimentation

Recall that we are interested in testing whether the treatment effect of each of the arms that have been added to the experiment is greater than some pre-determined threshold. This means that the experiment could consist of multiple hypothesis tests, one for each of the treatments that was added. Given that we do not know a priori the number of treatments that will be tested, we must control an error metric that is compatible with the flexibility sought after in Always-On experimentation. The online false discovery rate (Foster & Stine, 2008) affords such flexibility and, as such, we use it as our error metric in Always-On experimentation. In the rest of this section we adapt the notion of online false discovery rate to Always-On experimentation and provide an algorithm that controls it.

For a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, we let $\mathcal{H}_0 \subseteq \mathcal{A}$ be the set of treatments that are null under P based on the hypotheses from (1), or those from (11) in Section B. We let $\mathcal{D}_s$ be the set of discoveries that have been made up to sub-experiment $s \in [t]$ (i.e., the set of treatments $a \in \bigcup_{r=1}^{s} \mathcal{A}_r$ that have been declared as non-null), and we note how $\mathcal{D}_1 \subseteq \cdots \subseteq \mathcal{D}_t$. The *false discovery rate* (Benjamini & Hochberg, 1995) for the set $\mathcal{D}$ is defined as

$$\mathrm{FDR}(\mathcal{D}) \coloneqq \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\frac{|\mathcal{D} \cap \mathcal{H}_0|}{|\mathcal{D}| \vee 1}\right].$$

That is, the false discovery rate is the expected proportion of treatments that have been incorrectly declared as discoveries. Our aim is, thus, to produce discovery sets $\mathcal{D}_s$ so that the false discovery rate of each sub-experiment, and thus the overall experiment, is controlled at some level $\alpha \in (0, 1)$; more formally, our goal is to obtain the following guarantee:

$$\mathrm{FDR}(\mathcal{D}_s) \leq \alpha \quad \text{for all } s \in [t]. \tag{3}$$

To control the variant of the online false discovery rate given in (3), we adapt the asynchronous e-LOND algorithm due to Xu & Ramdas (2024) to Always-On experimentation. Our algorithm to control the online false discovery rate is given in Algorithm 1.

Algorithm 1 Online false discovery rate control in Always-On experimentation

Input: Confidence level $\alpha \in (0, 1)$.

1. Choose a sequence of nonnegative reals, $(\gamma_r)_{r=1}^{\infty}$, such that $\sum_{r=1}^{\infty} \gamma_r \leq 1$.
2. Initialize the hypothesis counter: $c = 0$.

for *each sub-experiment* $s = 1, \ldots, t$ **do**

3. Determine the set of new treatments to be tested: $\mathcal{A}'_s \coloneqq \mathcal{A}_s \setminus \bigcup_{r=1}^{s-1} \mathcal{A}_r$

  for *each* $a \in \mathcal{A}'_s$ **do**

4. Update the hypothesis counter: $c = c + 1$ and set $\gamma_a \coloneqq \gamma_c$.
5. Compute the confidence level: $\alpha_a \coloneqq \alpha\gamma_a \left(|\mathcal{D}_{s-1}| + 1\right)$

  end

end

Notice how Algorithm 1 produces confidence levels for each of the treatments added to the experiment. We, in turn, use these confidence levels for each treatment arm $a \in \mathcal{A}$ to construct their level-α_a sequential tests for each sub-experiment $s \in [t]$,

$$\phi_s^{(\alpha)}(a) \coloneqq \mathbb{1}\left\{W_{N_a(s)} \geq 1/\alpha_a\right\} . \tag{4}$$

As soon as $\phi_s^{(\alpha)}(a) = 1$ we declare treatment a a discovery and remove it from the experiment. When we combine this procedure with Algorithm 1, we can guarantee that the online false discovery rate from (3) is correctly controlled.

Proposition 3.5. *Fix a confidence level $\alpha \in (0, 1)$ and a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$. Suppose Assumptions 3.1 to 3.3 hold, and for each treatment $a \in \mathcal{A}$ and any sub-experiment $s \in [t]$, let $W_{N_a(s)}$ be its test supermartingale as defined in* (2)*. Then if we use Algorithm 1 to select the confidence levels for each treatment and use the sequential level-α test as defined in* (4)*, we have that for all $s \in [t]$, $\mathrm{FDR}(\mathcal{D}_s) \leq \alpha$.*

The proof of Proposition 3.5, given in Section D.2, explicitly handles scenarios in which treatments can be added to the experiment based on past observations, as well as instances in which the evidence used to declare hypotheses as discoveries (i.e., the test supermartingales) evolves over time. From Proposition 3.5 we can conclude that as long as we select the confidence levels for each treatment using Algorithm 1 and use test supermartingales within our sequential tests, then we can control the false discovery rate in Always-On experimentation. Hence, the section to follow is solely focused on constructing log-optimal test supermartingales. To build our understanding, in the main text we consider a setting in which units are only observed once for the entirety of the experiment, and in Section B we consider a setting in which units are repeatedly observed throughout the experiment.

4 Constructing Powerful Tests for Always-On Experimentation

We now consider a setting in which units are only observed once across the overall experiment, and describe how to construct test supermartingales that are growth-rate optimal. We first make explicit the assumptions that we impose throughout this section.

Assumption 4.1. *We make the following assumptions on the units that are observed in each sub-experiment $s \in [t]$: (i) If a unit participated in sub-experiment s, then this means that they have not and will not participate in any other sub-experiment $s' \in [t]$. (ii) The potential outcomes of the units take their values from the unit interval: $Y_i(\mathbf{z}) \in [0, 1]$, for each $i \in \mathcal{I}_s$ and $\mathbf{z} \in \mathcal{Z}$.*

A consequence of Assumption 4.1 is that the units' outcomes do not depend on past treatments—as they can only receive one treatment per experiment. Therefore, in this setting the set of all possible

treatment paths takes the form $\mathcal{Z} = \mathcal{A} \cup \{0\}$, and the observed outcome for unit $i \in \mathcal{I}_s$ is given by $Y_{i,s} \equiv Y_{i,s}(A_{i,s}) \coloneqq \sum_{a \in \mathcal{A} \cup \{0\}} \mathbb{1}\{A_{i,s} = a\} Y_{i,s}(a)$.

In the present setting, we are interested in testing whether the average treatment effect under distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, defined for any $a \in \mathcal{A}$ as $\psi_{\mathsf{P}}(a) \coloneqq \mathbb{E}_{\mathsf{P}}[Y_1(a) - Y_1(0)]$, is less than the threshold $\delta \in [-1, 1]$. The average treatment effect is unobservable as it depends on both the potential outcome under treatment a and under control for each unit. Nonetheless, under Assumptions 3.1, 3.2 and 4.1 it is an identifiable quantity, meaning that under each sub-experiment $s \in [t]$ and letting $i \in \mathcal{I}_s$: $\psi_{\mathsf{P}}(a) = \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}[Y_{i,s} \mid A_{i,s} = a] - \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}[Y_{i,s} \mid A_{i,s} = 0]$.

We use a slightly modified Horvitz-Thompson estimator (Horvitz & Thompson, 1952) (also known as inverse propensity weighting) to estimate the average treatment effect for each treatment in sub-experiment s using the data of unit $i \in \mathcal{I}_s$. For each treatment $a \in \mathcal{A}_s$ it takes the form

$$\widehat{\psi}_{i,s}(a) \coloneqq Y_{i,s} \left(\frac{\mathbb{1}\{A_{i,s} = a\}}{\widehat{\pi}_s(a)} - \frac{\mathbb{1}\{A_{i,s} = 0\}}{\widehat{\pi}_s(0)} \right), \tag{5}$$

where the modified propensity scores for testing the average effect of treatment a take the form

$$\widehat{\pi}_s(a) \coloneqq \frac{\pi_s(a)}{\pi_s(a) + \pi_s(0)} \quad \text{and} \quad \widehat{\pi}_s(0) \coloneqq \frac{\pi_s(0)}{\pi_s(a) + \pi_s(0)}. \tag{6}$$

We have slightly abused the notation and made implicit within the definition of the propensity score for the control the fact that it depends on the treatment being tested. An important property of the Horvitz-Thompson estimator from (5) is that it is conditionally unbiased.

Lemma 4.2. *Fix a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$ and a sub-experiment $s \in [t]$. Under Assumptions 3.1 to 3.3 and 4.1 we have that for any treatment $a \in \mathcal{A}_s$ and unit $i \in \mathcal{I}_s$, the Horvitz-Thompson estimator* (5) *is a conditionally unbiased estimator: $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\widehat{\psi}_{i,s}(a) \,\middle|\, \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] = \psi_{\mathsf{P}}(a)$.*

Under our assumption that the potential outcomes are bounded within the unit interval, the Horvitz-Thompson estimator is almost surely bounded as $-1/\widehat{\pi}_s(0) \leq \widehat{\psi}_{i,s}(a) \leq 1/\widehat{\pi}_s(a)$. In addition, to use the test supermartingale from (2), we must ensure that the random variables (and also threshold) are almost surely nonnegative. For this reason, for each sub-experiment $s \in \mathcal{T}_t(a)$ we define the transform $g_s : [-1/\widehat{\pi}_s(0), 1/\widehat{\pi}_s(a)] \to \mathbb{R}_{\geq 0}$ as follows:

$$g_s(x) \coloneqq 1 + \widehat{\pi}_s(0) x. \tag{7}$$

We use the map from (7) and the Horvitz-Thompson estimator from (5) in the test statistics defined in (2). Doing so leads to the following proposition, whose proof is given in Section D.4, which states that the test statistics given in (2) are test supermartingales.

Proposition 4.3. *Fix an arm $a \in \mathcal{A}$ and a distribution $\mathsf{P} \in \mathcal{P}(a)$. Let $W_{N_a(t)}$ be the test statistic from* (2) *constructed using the Horvitz-Thompson estimator given in* (5) *and transform defined in* (7)*. Then, under Assumptions 3.1 to 3.3 and 4.1, $W_{N_a(t)}$ forms a test $\mathcal{P}(a)$-supermartingale.*

Proposition 4.3 allows us to conclude that we can achieve time-uniform type-I error rate control in the present setting. Nevertheless, we are also interested in constructing test supermartingales that exhibit a large growth rate whenever we are under the alternative, no matter the the treatment schedule and units whom were part of the experiment.

To evaluate the growth-rate optimality of test supermartingales in Always-On experimentation, we hold the schedule of the overall experiment fixed. More specifically, we assume that the number of sub-experiments $t \in \mathbb{N}$ as well as the sets of active treatments $\mathcal{A}_s$ in each sub-experiment $s \in [t]$ are fixed and that the only growing quantity is the number of units that participate in each sub-experiment. To denote a growing number of units in each sub-experiment, we define $n_{\underline{t}} \coloneqq \min_{s \in [t]} n_s$ to be the minimum number of units in any of the sub-experiments. Thus, $n_{\underline{t}} \to \infty$ implies that $n_s \to \infty$ for all $s \in [t]$. Under the present setting, we assess the growth-rate optimamllity of test supermartingales that use the estimator from (5) and transform from (7) with respect to the following comparator class:

$$\mathcal{W}(a) \coloneqq \{W : W \text{ is a test } \mathcal{P}(a)\text{-supermartingale of the form (2)}\}, \tag{8}$$

for each treatment $a \in \mathcal{A}$. Having defined the necessary concepts, we now introduce the notion of log-optimality in dynamic experiments.

Definition 4.4 (Almost-sure log-optimality in Always-On experimentation)**.** *Fix an arm $a \in \mathcal{A}$ and a distribution $\mathsf{Q} \in \mathcal{Q}(a)$. Let $\mathcal{W}(a)$ be the class defined in* (8)*. We say that the test supermartingale $W_{N_a(t)}$ is* $\mathsf{Q}_{\mathrm{RCT}}$-almost surely log-optimal *with respect to the class $\mathcal{W}(a)$, if for any $\widetilde{W}_{N_a(t)} \in \mathcal{W}(a)$,*

$$\liminf_{n_{\underline{t}} \to \infty} \frac{1}{N_a(t)} \left(\log \left(W_{N_a(t)} \right) - \log \left(\widetilde{W}_{N_a(t)} \right) \right) \geq 0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-}\textit{almost surely}\,.$$

If $N_a(t) = 0$, then we say that the growth rates of $W_{N_a(t)}$ and $\widetilde{W}_{N_a(t)}$ are equivalent.

To construct growth-rate optimal test supermartingales, in the sense of Definition 4.4, we run independent copies of the Universal Portfolio algorithm (Cover, 1991) (as given in (17)) for each arm $a \in \mathcal{A}$ and restart the algorithms corresponding to the treatments in the active set of arms in every sub-experiment $s \in [t]$. More explicitly, in every sub-experiment $s \in [t]$ and for each unit $i \in \mathcal{I}_s$ we compute the portfolio corresponding to treatment $a \in \mathcal{A}_s$ as

$$\lambda_{i,s}(a) \equiv \mathrm{UP}\left(W^{(\lambda)}_{\check{N}_a(s;i-1)} \right) \quad \text{where} \quad W^{(\lambda)}_{\check{N}_a(s;i-1)} \coloneqq \prod_{j \in \mathcal{I}_s(i-1)} \left[(1-\lambda) + \lambda g_s(\widehat{\psi}_{j,s}(a)) / g_s(\delta) \right]^{\xi_{j,s}(a)}, \tag{9}$$

where we denote the portfolios as a function of a in order to make it explicit that the portfolios are computed independently across treatment arms. In the following theorem we establish the almost sure log-optimality of our test supermartingales when we select the portfolios using the strategy established in (9).

Theorem 4.5. *Fix the treatment schedule for the experiment and an arm $a \in \mathcal{A}$. Let $W_{N_a(t)}$ be a test $\mathcal{P}(a)$-supemartingale of the form* (2) *constructed using the transformed Horvitz-Thompson estimators and thresholds given in* (5)*, and whose portfolios are selected using the strategy from* (9)*. Suppose that Assumptions 3.1 to 3.3 and 4.1 hold. Then, for any $\mathsf{Q} \in \mathcal{Q}(a)$, $W_{N_a(t)}$ is $\mathsf{Q}_{\mathrm{RCT}}$-almost sure log-optimal in the sense of Definition 4.4.*

The proof of Theorem 4.5 can be found in Section D.5. It consists of deriving a nonasymptotic inequality on the absolute difference between $W_{N_a(t)}$ and the process that has oracle access to the numeraire portfolios (see Section C.1 for more details on these objects). The idea behind the proof resembles those found in Waudby-Smith et al. (2025) and Sandoval et al. (2026), but departs from these works as the number of data points used to construct the test supermartingale $W_{N_a(t)}$ is random under the present setting. Importantly, Theorem 4.5 establishes that no matter what the treatment schedule is, the way in which we are constructing the test supermartingales (i.e., selecting the portfolios) ensures that there is no other test supermartingale within the comparator class $\mathcal{W}(a)$ that could achieve a limiting growth-rate larger than ours. This includes a test supermartingale that has oracle access to and uses the numeraire portfolios for each treatment arm in each sub-experiment, as can be seen in Figure 2.

5 Semi-Synthetic Experiments

To test the efficacy of the proposed methods in realistic settings we constructed a set of semi-synthetic experiments using the Upworthy Research Archive (Matias et al., 2021), which is a corpus of randomized headline experiments that were run on an online platform between 2013 and 2015. Treatments consist of headline image combinations that were shown to users with at least 1,000 impressions with click through rate being the outcome of interest. On average each month contains 634 treatments per month. We construct a semi-synthetic Always-On experiment by assembling all tests that occurred within a given calendar month (18 total months), and using the observed timestamp for each treatment as the starting time within the overall Always-On experiment. See Section A.2 for details.

We compare the proposed method to an oracle that has access to the true treatment effects, to an ablation of our method that ignores the discovery term from the confidence schedule (whiwh we denote as "Ours Sans Reward"), to a version of our method without false discovery rate correction ("uncorrected"); and to early peeking in which results are monitored without correction ("peeking"). Figures 2 and 6 show the results. Consistent with our theoretical results, the proposed method

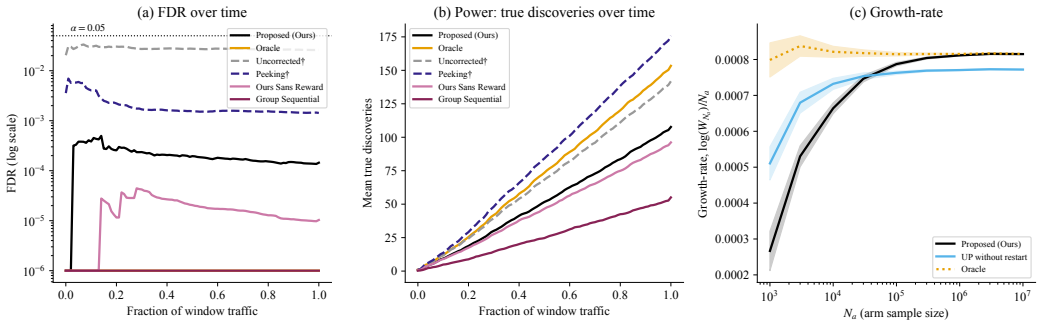

Figure 2: (a) FDR over window traffic, with 10 concurrent treatments. (b) Mean true discoveries over window traffic, with 10 concurrent treatments. (c) Empirical growth-rates for our proposed process, one that selects the portfolios via the Universal Portfolio algorithm but without restarting the algorithm in each sub-experiment, and a process that has oracle access to the numeraire portfolios.

controls the false discovery rate (FDR) below the nominal $\alpha = 0.05$ level across the traffic levels. By contrast, the uncorrected variants have an order-of-magnitude worse FDR control. We also see that our method discovers roughly 1,900 true effects per replication (see Figure 6), 12% more than the ablated version without reward and within 30% of the oracle. Finally, the group sequential baseline has a lower false discovery rate at the cost of a substantially more conservative discovery rate.

6 Conclusion

In this work we introduced a statistical framework to run randomized experiments in a way that allows treatments to by dynamically added and removed. The goal in our framework is to test the effectiveness of treatments while simultaneously controlling the online false discovery rate. As such, we based our framework on the sequential hypothesis testing-by-betting literature, and showed how to construct test supermartingales that test whether the treatment effect is greater than some threshold. We considered two settings: one in which each unit is observed at most once within the same experiment and another in which the experiment is run over the same set of units. We provided strategies to select the portfolios under these two settings that result in growth-rate optimal test supermartingales, no matter the treatment schedule. Furthermore, we provided and analyzed an algorithm that controls the online false discovery rate in the Always-On setting. Finally, experiments conducted using semi-synthetic data based on the Upworthy repository of online experiments provide additional evidence that the proposed methodology closely follows theory in empirical settings.

We envision that our Always-On framework will be useful in settings that encompass drug discovery, experimentation in tech companies, and testing for (emerging) capabilities of LLMs, amongst many others. To further improve the power of our framework, we outline some interesting directions for future work. One direction for future work is how to adaptively select the propensity scores to maximize the number of discoveries made in the experiment. A second direction for future work is designing the treatment schedule; that is, coming up with strategies to remove treatments from the experiment while controlling for the type-II error.

AI use statement

In this work, we used generative AI tools for writing the code used for the simulations and semi-synthetic experiments. We also used generative AI tools to help find related work and check the correctness of the proofs of the mathematical statements. We have not used generative AI tools to help with paper writing, generating synthetic datasets. As such, these are not applicable to this work. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Adaptive Platform Trials Coalition. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nature Reviews Drug Discovery*, 18(10):797–807, 2019.

Jay Bartroff and Jinlin Song. Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates. *Sequential analysis*, 39(1):65–91, 2020.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

Scott M Berry, Jason T Connor, and Roger J Lewis. The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA*, 313(16):1619–1620, 2015.

Beepul Bharti, Ambar Pal, and Jeremias Sulam. Global sequential testing for multi-stream auditing. *arXiv preprint arXiv:2602.21479*, 2026.

Iavor Bojinov, Ashesh Rambachan, and Neil Shephard. Panel experiments and dynamic causal effects: A finite population perspective. *Quantitative Economics*, 12(4):1171–1214, 2021.

Audrey Boruvka, Daniel Almirall, Katie Witkiewitz, and Susan A Murphy. Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, 113 (523):1112–1121, 2018.

L Breiman. Optimal gambling systems for favorable games. In *Fourth Berkeley Symposium: University of California*, 1961.

Ben Chugg, Santiago Cortes-Gomez, Bryan Wilder, and Aaditya Ramdas. Auditing fairness by betting. *Advances in Neural Information Processing Systems*, 36:6070–6091, 2023.

Eugenio Clerico. On the optimality of coin-betting for mean estimation. *International Journal of Approximate Reasoning*, pp. 109550, 2025.

Thomas M Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Thomas M Cover and Erik Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348–363, 1996.

Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 1999.

Donald A. Darling and Herbert E. Robbins. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences*, 58 1:66–8, 1967.

Conor P Farrington and Godfrey Manning. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in medicine*, 9(12):1447–1454, 1990.

Lasse Fischer, Ziyu Xu, and Aaditya Ramdas. An online generalization of the (e-) benjamini-hochberg procedure. *arXiv preprint arXiv:2407.20683*, 2024.

Dean P Foster and Robert A Stine. α-investing: a procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(2): 429–444, 2008.

Peter Grünwald, Rianne de Heide, and Wouter Koolen. Safe testing. *Journal of the Royal Statistical Society B (with discussion)*, 86(5):1091–1128, 2024.

Harrie Hendriks. Test martingales for bounded random variables. *arXiv preprint arXiv:2109.08923*, 2021.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021.

Guido Imbens, Lorenzo Masoero, Alexander Rakhlin, Thomas S Richardson, and Suhas Vijaykumar. Demonstration experiments. *arXiv preprint arXiv:2603.06941*, 2026.

Kevin Jamieson and Lalit Jain. A bandit approach to sequential experimental design with false discovery control. *Advances in Neural Information Processing Systems*, 31, 2018.

Adel Javanmard and Andrea Montanari. On online control of false discovery rate. *arXiv preprint arXiv:1502.06197*, 2015.

Adel Javanmard and Andrea Montanari. Online rules for control of false discovery rate and false discovery exceedance. *The Annals of statistics*, 46(2):526–554, 2018.

Ioannis Karatzas and Constantinos Kardaras. The numéraire portfolio in semimartingale financial models. *Finance and Stochastics*, 11:447–493, 2007.

John L Kelly. A new interpretation of information rate. *The Bell System Technical Journal*, 35(4): 917–926, 1956.

Ron Kohavi, Diane Tang, and Ya Xu. *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press, 2020.

Martin Larsson, Aaditya Ramdas, and Johannes Ruf. The numeraire e-variable and reverse information projection. *The Annals of Statistics*, 53(3):1015–1043, 2025.

Peng Liao, Predrag Klasnja, Ambuj Tewari, and Susan A Murphy. Sample size calculations for micro-randomized trials in mhealth. *Statistics in medicine*, 35(12):1944–1971, 2016.

John B Long Jr. The numeraire portfolio. *Journal of Financial economics*, 26(1):29–69, 1990.

J Nathan Matias, Kevin Munger, Marianne Aubin Le Quere, and Charles Ebersole. The upworthy research archive, a time series of 32,487 experiments in us media. *Scientific Data*, 8(1):195, 2021.

Jerzy Neyman. Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10(1):1–51, 1923.

Francesco Orabona and Kwang-Sung Jun. Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*, 70(1):436–455, 2023.

Aleksandr Podkopaev and Aaditya Ramdas. Sequential predictive two-sample and independence testing. *Advances in Neural Information Processing Systems*, 36:53275–53307, 2023.

Aleksandr Podkopaev, Patrick Blöbaum, Shiva Kasiviswanathan, and Aaditya Ramdas. Sequential kernelized independence testing. In *International Conference on Machine Learning*, pp. 27957–27993. PMLR, 2023.

Aaditya Ramdas and Ruodu Wang. Hypothesis testing with e-values. *Foundations and Trends® in Statistics*, 1(1-2):1–390, 2025.

Aaditya Ramdas, Tijana Zrnic, Martin Wainwright, and Michael Jordan. Saffron: an adaptive algorithm for online control of the false discovery rate. In *International conference on machine learning*, pp. 4286–4294. PMLR, 2018.

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv preprint arXiv:2009.03167*, 2020.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601, 2023.

Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409, 1970.

Herbert Robbins and David Siegmund. Iterated logarithm inequalities and related statistical procedures. *Mathematics of the Decision Sciences*, 2:267–279, 1968.

Herbert Robbins and David Siegmund. The expected sample size of some tests of power one. *The Annals of Statistics*, 2(3):415–436, 1974.

David S Robertson, James MS Wason, Franz König, Martin Posch, and Thomas Jaki. Online error rate control for platform trials. *Statistics in Medicine*, 42(14):2475–2495, 2023.

James M Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods*, 23(8):2379–2412, 1994.

James M Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics: analysis of correlated data*, pp. 189–326. Springer, 2004.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

J Jon Ryu and Alankrita Bhatt. On confidence sequences for bounded random processes via universal gambling strategies. *IEEE Transactions on Information Theory*, 2024.

J Jon Ryu, Jeongyeol Kwon, Benjamin Koppe, and Kwang-Sung Jun. Improved offline contextual bandits with second-order bounds: Betting and freezing. *arXiv preprint arXiv:2502.10826*, 2025.

Ricardo J Sandoval, Ian Waudby-Smith, and Michael I Jordan. Multi-armed sequential hypothesis testing by betting. *arXiv preprint arXiv:2603.17925*, 2026.

Michele Santacatterina, Federico Macchiavelli Giron, Xinyi Zhang, and Iván Díaz. Identification and estimation of causal effects using non-concurrent controls in platform trials. *Statistics in Medicine*, 44(6):e70017, 2025.

Stephen Senn. *Cross-over Trials in Clinical Research*. Wiley, 2nd edition, 2002.

Glenn Shafer and Vladimir Vovk. *Probability and finance: it's only a game!*, volume 491. John Wiley & Sons, 2005.

Glenn Shafer and Vladimir Vovk. *Game-theoretic foundations for probability and finance*. John Wiley & Sons, 2019.

Shubhanshu Shekhar and Aaditya Ramdas. Nonparametric two-sample testing by betting. *IEEE Transactions on Information Theory*, 70(2):1178–1203, 2023.

Jinjin Tian and Aaditya Ramdas. Addis: an adaptive discarding algorithm for online fdr control with conservative nulls. *Advances in neural information processing systems*, 32, 2019.

Jean Ville. Étude critique de la notion de collectif. *Bulletin of the American Mathematical Society*, 45(11):824, 1939.

Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16 (2):117–186, 1945.

Abraham Wald. *Sequential Analysis*. Wiley, New York, 1947.

Qiuqi Wang, Ruodu Wang, and Johanna Ziegel. E-backtesting. *Management Science*, 2025.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B*, 86(1):1–27, 2024.

Ian Waudby-Smith, Ricardo Sandoval, and Michael I Jordan. Universal log-optimality for general classes of e-processes and sequential hypothesis tests. *arXiv preprint arXiv:2504.02818*, 2025.

Janet Woodcock and Lisa M LaVange. Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine*, 377(1):62–70, 2017.

Ziyu Xu and Aaditya Ramdas. Online multiple testing with e-values. In *International Conference on Artificial Intelligence and Statistics*, pp. 3997–4005. PMLR, 2024.

Fanny Yang, Aaditya Ramdas, Kevin G Jamieson, and Martin J Wainwright. A framework for multi-a (rmed)/b (andit) testing with online fdr control. *Advances in Neural Information Processing Systems*, 30, 2017.

Sonja Zehetmayer, Martin Posch, and Franz Koenig. Online control of the false discovery rate in group-sequential platform trials. *Statistical Methods in Medical Research*, 31(12):2470–2485, 2022.

Tijana Zrnic, Aaditya Ramdas, and Michael I Jordan. Asynchronous online testing of multiple hypotheses. *Journal of Machine Learning Research*, 22(33):1–39, 2021.

A Further Information on Always-On Experimentation

In this section we provide further details about our Always-On experimentation framework. Recall that one of our goals in proposing this framework is to give practitioners the flexibility to add and remove treatments from a running experiment. We formalize the data collection protocol of the Always-On experimentation framework in Algorithm 2.

Algorithm 2 Data collection protocol

for *each sub-experiment* $s = 1, \dots, t$ **do**

1. Determine the set of active treatments $\mathcal{A}_s$.
2. Choose the propensity score $\boldsymbol{\pi}_s \coloneqq (\pi_s(a))_{a \in \mathcal{A}_s \cup \{0\}}$.
3. Nature samples the potential outcomes $(Y_i(\mathbf{z}))_{\mathbf{z} \in \mathcal{Z}} \sim \mathsf{P}$ for units $i \in \mathcal{I}_s$.

for *each unit* $i \in \mathcal{I}_s$ **do**

4. Randomly draw the treatment assignment $A_{i,s} \sim \mathrm{RCT}_s \equiv \mathrm{Multinomial}_{\mathcal{A}_s \cup \{0\}}(\boldsymbol{\pi}_s)$.
5. Observe the outcome $Y_{i,s}$.

end

end

As we described in Section 3.1, the data collected according to Algorithm 2 is used to construct test supermartingales for each of the active treatments. An important concept—which we expand on in Section C—in testing by betting is that of portfolio regret, as it provides a sufficient condition for achieving asymptotic almost sure log-optimality. Nonetheless, in Always-On experimentation we must consider a modified definition of portfolio regret. More specifically, we must consider a definition of portfolio regret for a particular sub-experiment $s \in [t]$ with respect to a particular arm $a \in \mathcal{A}_s$.

Definition A.1 (Per-sub-experiment portfolio regret)**.** *Fix a sub-experiment* $s \in [t]$*, a treatment* $a \in \mathcal{A}_s$*, and let* $(A_{i,s}, Y_{i,s})_{i \in \mathcal{I}_s}$ *be the data collected in sub-experiment* s*. Let* $W_{\check{N}_a(s)}$ *form a process of the form* (2)*. We define the* per-sub-experiment portfolio regret *for treatment* a *in sub-experiment* s *to be*

$$\mathcal{R}_{\check{N}_a(s)} \coloneqq \max_{\lambda \in [0,1]} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a) \log\left(1 - \lambda + \lambda g_s(\widehat{\psi}_{i,s}(a))/g_s(\delta)\right) - \log\left(W_{\check{N}_a(s)}\right),$$

where we recall that $\xi_{i,s}(a) \coloneqq \mathbb{1}\{A_{i,s} = a \text{ or } A_{i,s} = 0\}$.

The reason we must now consider a per-sub-experiment and per-arm definition of portfolio regret is because each sub-experiment has a particular joint distribution over the potential outcomes and treatment assignments. These different distributions, in turn, induce their own numeraire portfolios (see Section C.1), which means that for each sub-experiment we must construct processes that are "competitive" against those that have oracle access to these numeraire portfolios. Waudby-Smith et al. (2025) showed that to construct processes that are competitive against an oracle that has access to the numeraire portfolios is to use algorithms that select the portfolios in such a way that the test supermartingales they induce enjoy *sub-linear* portfolio regret; hence, explaining our portfolio selection strategy from Section 4. In the following subsection we present experiments whose aim is to provide some empirical comparison between Always-On experimentation and platform trials.

A.1 SIMULATIONS TO COMPARE ALWAYS-ON EXPERIMENTATION VERSUS PLATFORM TRIALS

Conceptually, Always-On experimentation is a way to run platform trials while allowing for continuous monitoring and arbitrary treatment schedules. In this section we explore what is the price we have to pay in Always-On experimentation relative to platform trials.

To understand the price to pay, we run various simulations comparing our Always-On experimentation framework to an adapted version of the platform trials design from Zehetmayer et al. (2022). More specifically, we use the gs-LOND algorithm proposed by Zehetmayer et al. (2022) in conjunction with the O'Brien–Fleming boundary to allow for interim analyses in a multiple testing setting. Furthermore, we use the Farrington-Manning test (Farrington & Manning, 1990) as our statistic to assess the effect of each treatment in the platform trial, as we consider Bernoulli outcomes within our simulations.

In the simulations we compare the difference in true discoveries made by our Always-On experimentation framework against the true discoveries made by the specification of the platform trial described in the previous paragraph. In each of the simulations we introduce a total of 60 treatments to the experiment, where 30 of them are under the null while the rest are under the alternative. In each of the simulations, we endow all the treatments under the alternative with the same effect (and similarly for those under the null) and vary the effect of the treatments under the alternative across simulations.

In the simulations we vary the maximum number of units allowed to be enrolled into a treatment (i.e., the cap on the number of units that can be assigned to each treatment) and consider different "peeking schedules" (i.e., number of times at which the test statistic can be computed) for always-on experimentation. Treatments are removed as soon as they are declared a discovery or whenever they hit the cap on the allowed number of units. Below we present the results of simulations for different treatment effects when testing that the effect is greater than the threshold $\delta = 0.05$.

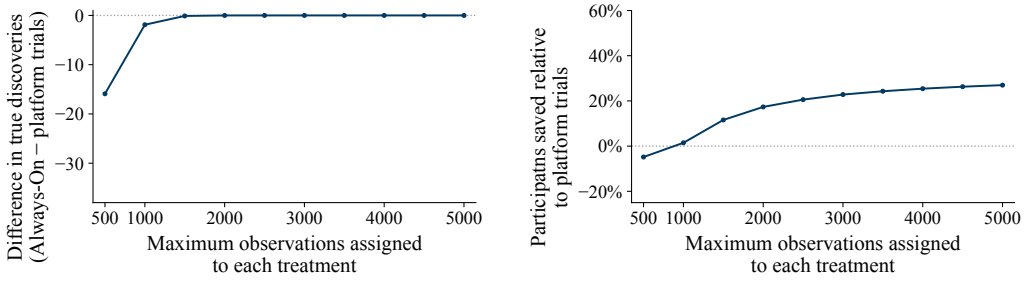

Figure 3: The true treatment effect of all 30 treatments under the alternative is equal to 0.25. The interim analysis in the platform trials only take place when 50% of the maximum number of units allowed for each treatment have been observed.

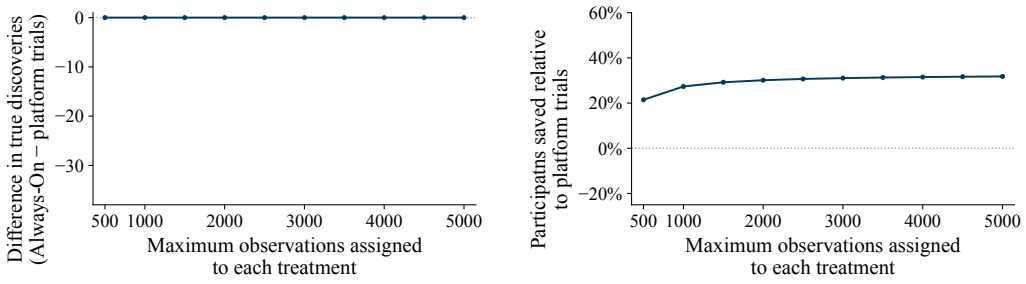

Figure 4: The true treatment effect of all 30 treatments under the alternative is equal to 0.55. The interim analysis in the platform trials only take place when 50% of the maximum number of units allowed for each treatment have been observed.

We are also interested in understanding the effect of being able to perform more interim analyses in platform trials has on the difference in true discoveries and the units needed to make such discoveries. Figure 5 depicts the effect that being able to perform more interim analyses in platform trials has on the difference in true discoveries with Always-On experimentation and the relative number of participants needed to make such discoveries.

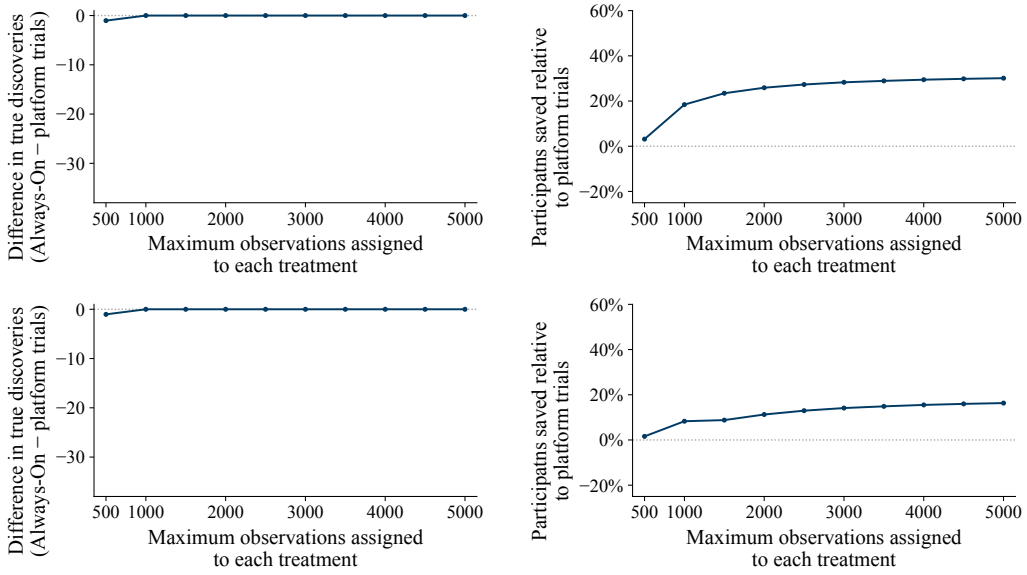

Figure 5: The true treatment effect of all 30 treatments under the alternative is equal to 0.35. In the top row we only perform one interim analysis when 50% of the maximum number of units allowed for each treatment have been observed. In the bottom row we perform interim analyses when 25%, 50%, and 75% of the maximum number of units allowed for each treatment have been observed.

A.2 Further information about the semi-synthetic experiment.

We fix a maximum number of concurrent treatments at ten, and create a queue in which a treatment is allowed to enter only after another treatment has been removed. Within each month, the control mean is simulated by taking the median click-through-rate (CTR) of all of the controls tested throughout the month, i.e., $\mathrm{CTR}_0 = \mathrm{median}(\mathrm{CTR}_{0,j})$, where j indexes the individual tests run that month. The outcomes are then simulated using the historical CTRs for each of the treatments, using the relative lift, $\rho_a = \mathrm{CTR}_a/\mathrm{CTR}_{0,j}$, where a indexes the treatment and j is the corresponding test for treatment a. We then use the same proportion with respect to the constructed always-on control, $\mathrm{CTR}_a^{\mathrm{sim}} = \rho_a \cdot \mathrm{CTR}_0$. A treatment is removed from an experiment either when it has been declared statistically significant with respect to the control or once it exceeds the total observation budget, i.e., the number of impressions it received in the dataset, $N_a \geq N_a^{\max}$. Every month was run with 200 replications each.

B Always-On Experimentation for Testing Treatment Histories

In this section we focus on a setting in which it is of interest to test whether there exists a treatment history for which a treatment induces a positive treatment effect. For this reason, the experiment is run over the same set of units—meaning that units receive multiple treatments throughout their experiment. This, in turn, implies that the observed outcomes are now functions of the history of treatments (or treatment path) a unit has received. We define the *treatment history* of unit i starting in sub-experiment $r \in [t]$ and going up to sub-experiment $s \in [t]$, such that $r \leq s$, as $\mathbf{Z}_{i,r:s} \coloneqq (A_{i,r}, \ldots, A_{i,s})$. We denote the set of all observable treatment histories starting in sub-experiment r and going up to sub-experiment s as

$$\mathcal{Z}_{r:s} \coloneqq \left\{ (a_r, \ldots, a_s) \in \mathcal{A}^{s-r+1} \;\middle|\; \forall j \in [s:r],\ a_j \in \mathcal{A}_j \right\} .$$

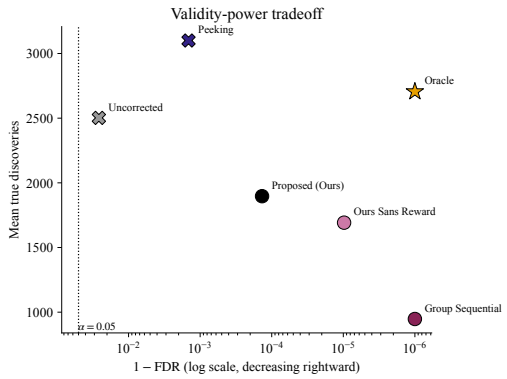

Figure 6: Validity and power at the primary traffic level (10 concurrent treatments): mean false discovery rate and mean true discoveries per replication.

Since units can receive more than one treatment in the experiment, we must now worry about carryover effects, and denote as $\ell \in \mathbb{N}_0$ the duration (in terms of sub-experiments) of the carryover effects—see (Senn, 2002; Bojinov et al., 2021) for examples of other settings in which carryover effects are of concern. Having carryover effects that last for ℓ sub-experiments implies that the *effective* treatment history of unit $i \in \mathcal{I}$ in sub-experiment $s \in [t]$ is given by $\mathbf{Z}_{i,s} \equiv \mathbf{Z}_{i,s-\ell:s} \coloneqq (A_{i,s-\ell}, \ldots, A_{i,s})$, and the effective set of observable treatment histories is $\mathcal{Z}_{s-\ell:s}$—where we remark that if $\ell > s$, then we let $s - \ell = 0$.

We assume each unit $i \in \mathcal{I}$ has potential outcomes $Y_i(\mathbf{z})$ for each $\mathbf{z} \in \bigcup_{s=1}^{t} \mathcal{Z}_{s-\ell:s}$. Given that the carryover effects last for ℓ sub-experiments, the observed outcome for unit i in sub-experiment s now takes the form $Y_{i,s} \coloneqq Y_i(\mathbf{Z}_{i,s}) = \sum_{\mathbf{z} \in \mathcal{Z}_{s-\ell,s}} \mathbb{1}\{\mathbf{Z}_{i,s} = \mathbf{z}\} Y_i(\mathbf{z})$. In the present setting we assume that the potential outcomes can take only a finite number of outcome levels.

Assumption B.1 (Finite number of outcome levels)**.** *We assume all potential outcomes take a finite number of outcome levels, and that they take their values from the set $\{0, \ldots, m\}$ for some $m \in \mathbb{N}$ such that $m < \infty$.*

For the rest of the paper, we define the shorthand $\mathbf{Y}_{i,r:s} \coloneqq (Y_{i,r}, \ldots, Y_{i,s})$ to denote the outcomes that have been observed for unit i starting from sub-experiment r and going up to sub-experiment s. We define the history of observed outcomes and treatments for unit $i \in \mathcal{I}$ as $\mathbf{H}_{i,r:s} \coloneqq (\mathbf{Y}_{i,r:s}, \mathbf{Z}_{i,r:s})$. Furthermore, we denote the set of *observed* histories starting in sub-experiment $r \in [t]$ and going up to sub-experiment $s \in [t]$ as $\mathcal{H}_{r:s} \coloneqq \{\mathbf{H}_{i,r:s}, \forall i \in [n]\}$, and we use the shorthand $\mathcal{H}_s \equiv \mathcal{H}_{s-\ell:s}$.

In addition to assuming that the outcomes are functions of only the past ℓ treatments, we impose the following Markovian assumption on the distribution from which the outcomes are sampled.

Assumption B.2 (Markov distribution over outcomes)**.** *Fix a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, let $\ell \in \mathbb{N}_0$ be the duration of the carryover effects, and define the realized history for unit i from sub-experiment r up to sub-experiment $s-1$ as $\mathbf{h}_{i,r:s-1} \coloneqq (\mathbf{y}_{i,r:s-1}, \mathbf{z}_{i,r:s-1})$. We say that the distribution over the outcomes is ℓ-*Markov *if for each sub-experiment $s \in [t]$, treatment $a \in \mathcal{A}_s$, and outcome value $y \in \{0, \ldots, m\}$ it holds that*

$$\mathbb{P}_{\mathsf{P}}\left(Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) = y \mid \mathbf{H}_{i,1:s-1} = \mathbf{h}_{i,1:s-1}\right) = \mathbb{P}_{\mathsf{P}}\left(Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) = y \mid \mathbf{H}_{i,s-\ell:s-1} = \mathbf{h}_{i,s-\ell:s-1}\right) .$$

A consequence of Assumption B.2 is that the expected outcomes are also ℓ-Markov—i.e., only depend on the treatments and observed outcomes from the past ℓ sub-experiments. We leverage this assumption and define our target estimand with respect to the histories from the set $\mathcal{H}_{s-1}$. More formally, in sub-experiment $s \in [t]$, we define the *per-sub-experiment treatment effect* for each $a \in \mathcal{A}_s$ for those units whose history $\mathbf{H}_{i,s-1} \equiv \mathbf{H}_{i,s-\ell:s-1}$ is equal to $\mathbf{h} \in \mathcal{H}_{s-1}$ as

$$\psi_{\mathsf{P}}(a, \mathbf{h}) \coloneqq \mathbb{E}_{\mathsf{P}}\left[Y_1(\mathbf{Z}_{1,s-\ell:s-1}, a) - Y_1(\mathbf{Z}_{1,s-\ell:s-1}, 0) \mid \mathbf{H}_{1,s-1} = \mathbf{h}\right] . \tag{10}$$

Our estimand from (10) captures the effect of a treatment $a \in \mathcal{A}_s$ when compared against the control and when we condition on the history of realized outcomes and treatments that were assigned throughout the past ℓ sub-experiments. For the first sub-experiment we can see how the per

sub-experiment treatment effect simplifies to $\psi_{\mathsf{P}}(a) \coloneqq \mathbb{E}_{\mathsf{P}}[Y_1(a) - Y_1(0)]$, which is the standard average treatment effect for treatment $a \in \mathcal{A}_1$. We note how the estimand in (10) resembles the blip function (Robins, 1994; 2004); albeit, it only contrasts the outcome under treatment a against the control in sub-experiment s and not two different treatment regimes. Moreover, our estimand is analogous to the proximal effect from micro-randomized trials (Liao et al., 2016; Boruvka et al., 2018), but differs from the proximal effect because these estimands typically condition on the availability of units and a summary of the pre-treatment history.

In the present Always-On setting, we slightly modify the hypotheses tests from (1) and test for each treatment $a \in \mathcal{A}$ the following null versus alternative hypotheses

$$\mathcal{P}(a) \coloneqq \{\mathsf{P} \mid \forall \mathbf{h} \in \overline{\mathcal{H}}(a)\,,\ \psi_{\mathsf{P}}(a, \mathbf{h}) \leq \delta\} \quad \text{versus} \quad \mathcal{Q}(a) \coloneqq \{\mathsf{P} \mid \exists \mathbf{h} \in \overline{\mathcal{H}}(a)\,,\ \psi_{\mathsf{P}}(a, \mathbf{h}) > \delta\}\,, \tag{11}$$

where $\overline{\mathcal{H}}(a) \coloneqq \bigcup_{s \in \mathcal{T}_t(a)} \mathcal{H}_s$. We remark that $\overline{\mathcal{H}}(a)$ is partially observed throughout the experiment and that it is only fully observed until the last sub-experiment. This is due to the fact that $\overline{\mathcal{H}}(a)$ consists of all possible histories that were observed in the sub-experiments in which treatment a was active, and we can only determine all these histories until after the experiment as a whole comes to an end.

The hypotheses from (11) depend on the per-sub-experiment treatment effect, a quantity that we must estimate. For each sub-experiment $s \in [t]$ we use the Horvitz-Thompson estimator to estimate the per-sub-experiment treatment effect for treatment $a \in \mathcal{A}_s$ for those units whose treatment histories were $\mathbf{h} \in \mathcal{H}_{s-1}$ and this takes the form

$$\widehat{\psi}_{i,s}(a, \mathbf{h}) \coloneqq Y_{i,s} \left(\frac{\mathbb{1}\{A_{i,s} = a\}}{\widehat{\pi}_s(a)} - \frac{\mathbb{1}\{A_{i,s} = 0\}}{\widehat{\pi}_s(0)} \right)\,, \tag{12}$$

where the modified propensity scores for testing the average treatment effect of treatment a are defined as in (6). We remark that the observed outcome $Y_{i,s}$ is a function of both the treatment history $\mathbf{Z}_{i,s-1}$ and the treatment assignment in sub-experiment s; that is $Y_{i,s} \equiv Y_i(\mathbf{Z}_{i,s-\ell:s-1}, A_{i,s})$. Furthermore, $Y_{i,s}$ can be dependent on the observed outcomes $\mathbf{Y}_{i,s-1} = \mathbf{y}_{s-1}$ encoded in the history $\mathbf{h}$. Under the present construction, the estimator from (12) is a conditionally unbiased estimator for the per-sub-experiment treatment effect—where the conditioning is on the filtration $\mathcal{F}_{i-1,s}$ and inclusion indicator $\xi_{i,s}(a, \mathbf{h}) \coloneqq \mathbb{1}\{(A_{i,s} = a \text{ or } A_{i,s} = 0) \text{ and } \mathbf{H}_{i,s-1} = \mathbf{h}\}$ being one.

Lemma B.3. *Fix a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, a sub-experiment $s \in [t]$, and an arm $a \in \mathcal{A}_s$. Under Assumptions B.1, B.2 and 3.1 to 3.3, the Horvitz-Thompson estimator from* (12) *for unit $i \in \mathcal{I}_s$ belonging to strata $\mathbf{h} \in \mathcal{H}_s$ is a conditionally unbiased estimator, meaning that* $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\widehat{\psi}_{i,s}(a, \mathbf{h}) \,\middle|\, \mathcal{F}_{i-1,s}, \xi_{i,s}(a, \mathbf{h}) = 1\right] = \psi_{\mathsf{P}}(a, \mathbf{h})$.

The proof of Lemma B.3 can be found in Section D.6. The Horvitz-Thompson estimator in the present setting takes it values from the interval $\widehat{\psi}_{i,s}(a) \in [-m/\widehat{\pi}_s(0), m/\widehat{\pi}_s(a)]$, which means that we must again apply a transformation so that its values are almost surely nonnegative. In this setting, the transformation function for each sub-experiment $s \in \mathcal{T}_t(a)$, $g_s : [-m/\widehat{\pi}_s(0), m/\widehat{\pi}_s(a)] \to \mathbb{R}_{\geq 0}$, is given by

$$g_s(x) \coloneqq m + \widehat{\pi}_s(0)x\,. \tag{13}$$

Using the map from (13) alongside the estimator from (12) in the test statistic given in (2) results in a test $\mathcal{P}(a)$-supermartingale.

Proposition B.4. *Fix an arm $a \in \mathcal{A}$ and a distribution $\mathsf{P} \in \mathcal{P}(a)$. Let $W_{N_a(t)}$ be the test statistic from* (2) *constructed using the Horvitz-Thompson estimator from* (12) *and the transform defined in* (13)*. Then, under Assumptions B.1, B.2 and 3.1 to 3.3, $W_{N_a(t)}$ forms a test $\mathcal{P}(a)$-supermartingale.*

Proof. Proposition B.4 follows immediately from Lemmas B.3 and 3.4, and the fact that $g_s(\cdot)$ is a non-decreasing linear map for each $s \in [t]$. □

As in the main text, one of our goals is to construct powerful test statistics. Nonetheless, the comparator class we consider in this section is one whose test supermartingales are constructed using the estimator from (12) and transform from (13). For each treatment $a \in \mathcal{A}$, we denote it as

$$\overline{\mathcal{W}}(a) \coloneqq \{W : W \text{ is a test } \mathcal{P}(a)\text{-supermartingale of the form (2)}\}\,.$$

In the present setting, we can no longer follow the same strategy to select the portfolios as that from the main text. This is because in the present setting we get to run the experiment over the same set of units; thus, we get to repeatedly observe the same unit and must take into account the temporal dependency this induces—since the potential outcomes for each unit can be dependent with each other. Assumptions B.1 and B.2 alongside our carryover effect assumption limit the extent of the temporal dependency. More specifically, the temporal dependency in sub-experiment s is encoded in the history $\mathbf{h} \in \mathcal{H}_{s-1}$. As a consequence, each history $\mathbf{h} \in \mathcal{H}_{s-1}$ and treatment $a \in \mathcal{A}_s$ pair induces (possibly) different numeraire portfolios. For this reason, in the present setting we must control a version of per-sub-experiment portfolio regret that takes into account each history $\mathbf{h} \in \mathcal{H}_{s-1}$.

Definition B.5 (Per-sub-experiment and history portfolio regret)**.** *Fix a sub-experiment $s \in [t]$, a treatment $a \in \mathcal{A}_s$, and let $(A_{i,s}, Y_{i,s})_{i \in \mathcal{I}_s}$ be the data collected in sub-experiment s. Let $W_{\check{N}_{a,\mathbf{h}}}$ form a process of the form* (2) *but just over those units whose history is equal to $\mathbf{h} \in \mathcal{H}_{s-1}$. We define the* per-sub-experiment and history portfolio regret *for treatment a and history $\mathbf{h} \in \mathcal{H}_{s-1}$ in sub-experiment s to be*

$$\mathcal{R}_{\check{N}_{a,\mathbf{h}}(s)} \coloneqq \max_{\lambda \in [0,1]} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a, \mathbf{h}) \log \left(1 - \lambda + \lambda g_s(\widehat{\psi}_{i,s}(a, \mathbf{h}))/g_s(\delta)\right) - \log \left(W_{\check{N}_{a,\mathbf{h}}(s)}\right) ,$$

where we have defined $\xi_{i,s}(a, \mathbf{h}) \coloneqq \mathbb{1}\left\{(A_{i,s} = a \text{ or } A_{i,s} = 0) \text{ and } \mathbf{H}_{i,s} = \mathbf{h}\right\}$.

Hence, we must now select the portfolios in such a way that controls the per-sub-experiment and history portfolio regret. To do so, we now use independent copies of the universal portfolio algorithm for each history in each sub-experiment. Hence, the portfolio for arm $a \in \mathcal{A}_s$ in sub-experiment $s \in [t]$ for unit $i \in \mathcal{I}_s$ whose history is $\mathbf{h} \in \mathcal{H}_{s-1}$ is given by

$$\lambda_{i,s}(a, \mathbf{h}) \equiv \text{UP}\left(W^{(\lambda)}_{\check{N}_{a,\mathbf{h}}(s;i-1)}\right) \tag{14}$$
$$\text{where} \quad W^{(\lambda)}_{\check{N}_{a,\mathbf{h}}(s;i-1)} \coloneqq \prod_{j \in \mathcal{I}_s(i-1)} \left[(1-\lambda) + \lambda g_s(\widehat{\psi}_{j,s}(a, \mathbf{h}))/g_s(\delta)\right]^{\xi_{j,s}(a,\mathbf{h})} ,$$

and where we recall the definition $\xi_{i,s}(a, \mathbf{h}) \coloneqq \mathbb{1}\left\{(A_{i,s} = a \text{ or } A_{i,s} = 0) \text{ and } \mathbf{H}_{i,s} = \mathbf{h}\right\}$. This strategy results in test supermartingales that are asymptotically almost sure log-optimal.

Theorem B.6. *Fix the treatment schedule for the experiment and an arm $a \in \mathcal{A}$. Let $W_{N_a(t)}$ be a test $\mathcal{P}(a)$-supemartingale of the form* (2) *constructed using the transformed Horvitz-Thompson estimators and thresholds given in* (12) *and* (13)*, respectively, and whose portfolios are selected using the strategy from* (14)*. Suppose that Assumptions B.1, B.2 and 3.1 to 3.3 hold. Then, for any $\mathsf{Q} \in \mathcal{Q}(a)$, $W_{N_a(t)}$ is Q_{RCT}-almost sure log-optimal—in the sense of Definition 4.4—within the class $\overline{\mathcal{W}}(a)$.*

We provide the proof of Theorem B.6 in Section D.7. It follows similar steps to the proof of Theorem 4.5, but controls the additional complexity that arises from running the experiment over the same set of units. Hence, Theorem B.6 states that regardless of the treatment schedule, we can construct powerful test statistics in Always-On experimentation even in settings where we get to repeatedly test over the same set of units.

C SEQUENTIAL HYPOTHESIS TESTING BY BETTING

We now review some important concepts of sequential hypothesis testing by betting—a research program whose inception can be traced back to the works of Abraham Wald (Wald, 1945; 1947) and Herbert Robbins and colleagues (Darling & Robbins, 1967; Robbins & Siegmund, 1968; Robbins, 1970; Robbins & Siegmund, 1974). For in-depth reviews, see the books of Shafer and Vovk (Shafer & Vovk, 2005; 2019), Ramdas and Wang (Ramdas & Wang, 2025), and the review paper of Ramdas et al. (2023).

Let $(Y_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables on a filtered space $(\Omega, \mathcal{F})$, where $\mathcal{F} \equiv (\mathcal{F}_n)_{n \in \mathbb{N}_0}$ is the filtration generated by $(Y_n)_{n \in \mathbb{N}}$ and $\mathcal{F}_0 = \{\Omega, \varnothing\}$ is the trivial sigma-algebra. We are particularly interested in testing some null hypothesis—which is represented by the collection of

distributions $\mathcal{P}$—against an alternative—from a collection of distributions $\mathcal{Q}$—such that $\mathcal{P} \cap \mathcal{Q} = \varnothing$. In testing by betting, one fixes a desired type-I error rate $\alpha \in (0, 1)$ and aims to construct a binary-valued $\mathcal{F}$-adapted map $(\phi_n^{(\alpha)})_{n\in\mathbb{N}}$, where $\phi_n^{(\alpha)} = 1$ should be interpreted as "reject the null" and $\phi_n^{(\alpha)} = 0$ as "do not reject." The map $\phi_n^{(\alpha)}$ is commonly referred to as the "level-α sequential test." Importantly, it is desired that the level-α sequential test provides the following guarantee:

$$\sup_{\mathsf{P}\in\mathcal{P}} \mathbb{P}_{\mathsf{P}}\left(\exists n \in \mathbb{N} : \phi_n^{(\alpha)} = 1\right) \leq \alpha \quad \text{or equivalently} \quad \sup_{\mathsf{P}\in\mathcal{P}} \mathbb{P}_{\mathsf{P}}\left(\phi_\tau^{(\alpha)} = 1\right) \leq \alpha\,, \tag{15}$$

where τ is an arbitrary $\mathcal{F}$-stopping time, meaning that $\{\tau = n\} \in \mathcal{F}_n$ for any $n \in \mathbb{N}$. One way of constructing such tests is via *test supermartingales*.

Definition C.1 (Test supermartingales)**.** *Let $(W_n)_{n\in\mathbb{N}}$ be an $\mathcal{F}$-adapted process, meaning that W_n is $\mathcal{F}_n$-measurable for each $n \in \mathbb{N}$. For a given distribution $\mathsf{P} \in \mathcal{P}$, we say that $(W_n)_{n\in\mathbb{N}}$ is a test P-supermartingale if:*

(i) $(W_n)_{n\in\mathbb{N}_0}$ *is a* P*-supermartingale, meaning that* $\mathbb{E}_{\mathsf{P}}[W_n \mid \mathcal{F}_{n-1}] \leq W_{n-1}$ P*-almost surely,*

(ii) $W_n \geq 0$ *for every* $n \in \mathbb{N}$ P*-almost surely, and*

(iii) $\mathbb{E}_{\mathsf{P}}[W_1] \leq 1$.

We say that $(W_n)_{n\in\mathbb{N}_0}$ is a test $\mathcal{P}$-supermartingale if it is a test P-supermartingale for every $\mathsf{P} \in \mathcal{P}$.

In this work we focus on test supermartingales that take the following form

$$W_n \coloneqq \prod_{i=1}^{n} \left[(1 - \lambda_i) + \lambda_i E_i\right]\,, \tag{16}$$

where $(\lambda_n)_{n\in\mathbb{N}}$ is a $[0, 1]$-valued predictable process (in the sense that λ_n is $\mathcal{F}_{n-1}$-measurable for every $n \in \mathbb{N}$), and $(E_n)_{n\in\mathbb{N}}$ are sequential e-variables under the null hypothesis $\mathcal{P}$—meaning that E_n is nonnegative with P-probability one and $\mathbb{E}_{\mathsf{P}}[E_n \mid \mathcal{F}_{n-1}] \leq 1$ for every $\mathsf{P} \in \mathcal{P}$.

Level-α sequential tests can be constructed by thresholding test P-supermartingales at $1/\alpha$, and Ramdas et al. (2020) show that this is the only admissible way to construct level-α sequential tests. More specifically, letting $\phi_n^{(\alpha)} \coloneqq \mathbb{1}\{W_n \geq 1/\alpha\}$ we can satisfy the type-I error control from (15) by appealing to Ville's inequality for nonnegative supermartingales (Ville, 1939):

$$\sup_{\mathsf{P}\in\mathcal{P}} \mathbb{P}_{\mathsf{P}}\left(\exists n \in \mathbb{N} : \phi_n^{(\alpha)} = 1\right) = \sup_{\mathsf{P}\in\mathcal{P}} \mathbb{P}_{\mathsf{P}}\left(\exists n \in \mathbb{N} : W_n \geq 1/\alpha\right) \leq \alpha\mathbb{E}_{\mathsf{P}}[W_1] \leq \alpha\,.$$

Indeed, many of the sequential tests that have been proposed in the literature have been constructed in this way, and we refer the reader to (Robbins, 1970; Robbins & Siegmund, 1974; Howard et al., 2021; Waudby-Smith & Ramdas, 2024; Orabona & Jun, 2023; Waudby-Smith et al., 2025; Sandoval et al., 2026) for a non-exhaustive list of examples.

The level-α sequential test $(\phi_n^{(\alpha)})$ surfaces the following interpretation of the test supermartingale W_n: it is our evidence against the null $\mathcal{P}$. For this reason, it is of interest to obtain test supermartingales that diverge to infinity as quickly as possible under the alternative. Test supermartingales with large growth rates are, therefore, considered powerful—a designation that can be traced back to the works of Kelly (1956), Breiman (1961), and Long Jr (1990). A more recent literature has made connections between large growth rates and hypothesis testing (Grünwald et al., 2024; Waudby-Smith & Ramdas, 2024; Larsson et al., 2025; Waudby-Smith et al., 2025; Orabona & Jun, 2023; Wang et al., 2025), and most recently Waudby-Smith et al. (2025) introduce the following definition of *almost-sure (growth-rate) log-optimality* of test supermartingales.

Definition C.2 (Almost sure log-optimality (Waudby-Smith et al., 2025))**.** *Fix a null $\mathcal{P}$ and an alternative $\mathcal{Q}$. For $\mathsf{Q} \in \mathcal{Q}$, a $\mathcal{P}$-test supermartingale $W \equiv (W_n)_{n\in\mathbb{N}}$ is said to be Q-log-optimal in a class of test supermartingales $\mathcal{W}$ if for any $W' \in \mathcal{W}'$,*

$$\liminf_{n\to\infty}\left(\frac{1}{n}\log(W_n) - \frac{1}{n}\log(W'_n)\right) \geq 0 \quad \mathsf{Q}\text{-almost surely}\,.$$

We say that W is $\mathcal{Q}$-log-optimal if the same process is Q-log-optimal for every $\mathsf{Q} \in \mathcal{Q}$.

Instantiating the class of test supermartingales $\mathcal{W}$ referenced in Definition C.2 as

$$\mathcal{W}' \coloneqq \{W : W \text{ is a test } \mathcal{P}\text{-supermartingale of the form given in (16)}\} \ ,$$

Waudby-Smith et al. (2025) showed[2] that a sufficient condition to achieve log-optimality with respect to $\mathcal{W}$ is to select the *portfolios* $(\lambda_n)_{n\in\mathbb{N}}$ using algorithms that enjoy sub-linear portfolio regret.

Definition C.3 (Portfolio regret)**.** *Let $(e_n)_{n\in\mathbb{N}}$ be a deterministic sequence lying in $[0,\infty)$. We define the* portfolio regret *of the real sequence $(W_n)_{n\in\mathbb{N}}$ as*

$$\mathcal{R}_n \equiv \mathcal{R}_n(W_n) \coloneqq \max_{\lambda\in[0,1]} \sum_{i=1}^{n} \log\left((1-\lambda)+\lambda e_i\right) - \log\left(W_n\right) \ .$$

Importantly, the Universal Portfolio algorithm from Cover (1991) achieves sub-linear portfolio regret (Cover & Ordentlich, 1996), meaning that $\mathcal{R}_n = o(n)$, and it computes the portfolios as

$$\mathrm{UP}(W_{n-1}^{(\lambda)}) \coloneqq \frac{\int_{\lambda\in[0,1]} \lambda W_{n-1}^{(\lambda)} \mathrm{d}F(\lambda)}{\int_{\lambda\in[0,1]} W_{n-1}^{(\lambda)} \mathrm{d}F(\lambda)} \, , \tag{17}$$

where $F(\cdot)$ is set to be the Beta$(1/2,1/2)$ distribution, and $W_{n-1}^{(\lambda)} \coloneqq \prod_{i=1}^{n-1}[(1-\lambda)+\lambda E_i]$. Hence, to achieve the notion of optimality defined in Definition C.2 it is sufficient to select the portfolios $(\lambda_n)_{n\in\mathbb{N}}$ using Cover's Universal Portfolio algorithm from (17).

C.1 Numeraire Portfolios

In this section we briefly review numeraire portfolios (Long Jr, 1990; Karatzas & Kardaras, 2007; Larsson et al., 2025), objects that are fundamental to many of our proofs. To ease the notation in the discussion and proofs to follow, we employ the following shorthand throughout the appendix:

$$\boldsymbol{\lambda}_{i,t} \coloneqq (1-\lambda_{i,t} \ \ \lambda_{i,t})^\top \quad \text{and} \quad \mathbf{E}_{i,t}(a) \coloneqq \left(1 \ \ g_t(\widehat{\psi}_{i,t}(a))/g_t(\delta)\right)^\top \, ,$$

for all treatment arms $a \in \mathcal{A}$, with analogous notation for the setting in which units can be repeatedly observed in the experiment (see Section B). The above notation allows us to write the process from (2) in the following compact form

$$W_{N_a(t)} = \prod_{s\in\mathcal{T}_t(a)} \prod_{i\in\mathcal{I}_s} \left(\boldsymbol{\lambda}_{i,s}^\top \mathbf{E}_{i,s}(a)\right)^{\xi_{i,s}(a)} \, .$$

Numeraire portfolios are those portfolios which maximize the expected logarithmic increment under the distribution generating the data—which can also be thought of as those portfolios which maximize the growth-rate of the test supermartingale. In our Always-On setting it is the case that the random variables we work with are generated by different distributions, as the distributions over the active set of treatments differ across sub-experiments. For this reason, every sub-experiment $s \in [t]$ has its own set of numeraire portfolios corresponding to each of the active arms $a \in \mathcal{A}_s$. Hence, the numeraire portfolios for the setting considered in Section 4 are defined as

$$\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a) \coloneqq \operatorname*{arg\,max}_{\boldsymbol{\lambda}\in\triangle_1} \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\log\left(\boldsymbol{\lambda}^\top \mathbf{E}_{1,s}(a)\right) \,\middle|\, \mathcal{F}_{s-1}, \xi_{1,s}(a)=1\right] \, , \tag{18}$$

for each $s \in [t]$ and $a \in \mathcal{A}_s$. The numeriare portfolios for the setting considered in Section B are defined in a similar way, but under this setting we endow a numeraire portfolios to each arm $a \in \mathcal{A}_s$ and history $\mathbf{h} \in \mathcal{H}_{s-1}$ in sub-experiment $s \in [t]$,

$$\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a,\mathbf{h}) \coloneqq \operatorname*{arg\,max}_{\boldsymbol{\lambda}\in\triangle_1} \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\log\left(\boldsymbol{\lambda}^\top \mathbf{E}_{1,s}(a,\mathbf{h})\right) \,\middle|\, \mathcal{F}_{s-1}, \xi_{1,s}(a,\mathbf{h})=1\right] \, . \tag{19}$$

An important property of numeraire portfolios is that for any other $\boldsymbol{\lambda} \in \triangle_d$, the expected sub-optimality ratio is upper bounded by one.

[2] Waudby-Smith et al. (2025) provide their results for a more general class of test supermartingales than that in (16), but we focus our discussion on test supermartingales of the form (16) to avoid introducing any extra notation.

Lemma C.4. *Fix a sub-experiment $s \in [t]$ and treatment $a \in \mathcal{A}_s$. Suppose that Assumptions 3.1 to 3.3 hold. Then, for any unit $i \in \mathcal{I}_s$ and $\boldsymbol{\lambda} \in \triangle_1$ we have that*

$$\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\frac{\boldsymbol{\lambda}^\top \mathbf{E}_{i,s}(a)}{\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a)^\top \mathbf{E}_{i,s}(a)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] \leq 1\,.$$

Furthermore, suppose that Assumption B.2 also holds. We then have that for any history $\mathbf{h} \in \mathcal{H}_{s-1}$ the following inequality holds for any $\boldsymbol{\lambda} \in \triangle_1$,

$$\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\frac{\boldsymbol{\lambda}^\top \mathbf{E}_{i,s}(a,\mathbf{h})}{\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a,\mathbf{h})^\top \mathbf{E}_{i,s}(a,\mathbf{h})} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] \leq 1\,.$$

Proof. The proofs of both statements follow from the definitions of their numeraire portfolios—given in (18) and (19)—and from (Cover & Thomas, 1999, Theorem 15.2.2). □

In addition to "inducing e-values" under the alternative distribution, numeraire portfolios help characterize certain optimality quantities of test supermartingales, but we differ a detailed treatment of this fact to the work of Waudby-Smith et al. (2025).

D Proofs of the Main Results

In this section we present the proofs of our main results. Some of the proofs to follow make use of processes that have oracle access to either the numeraire portfolios or the best in hindsight portfolios. We define these processes below. The proofs of the results from Section 4 make use of the process that has access to the numeraire portfolios in each sub-experiment

$$W^{(\mathsf{Q})}_{N_a(t)} \coloneqq \prod_{s \in \mathcal{T}_t(a)} \prod_{i \in \mathcal{I}_s} \left(\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a)^\top \mathbf{E}_{i,s}(a)\right)^{\xi_{i,s}(a)}\,,$$

where $\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a)$ is defined as in Equation (18) for each $s \in [t]$ and $a \in \mathcal{A}_s$. The second process is one that uses the best in hindsight portfolio for each of the active treatment arms in every sub-experiment:

$$W^{\star}_{N_a(t)} \coloneqq \prod_{s \in \mathcal{T}_t(a)} \prod_{i \in \mathcal{I}_s} \left(\boldsymbol{\lambda}_s^{\star}(a)^\top \mathbf{E}_{i,s}(a)\right)^{\xi_{i,s}(a)} \quad \text{where} \quad \boldsymbol{\lambda}_s^{\star}(a) \coloneqq \operatorname*{arg\,max}_{\boldsymbol{\lambda} \in \triangle_1} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a) \log\left(\boldsymbol{\lambda}^\top \mathbf{E}_{i,s}(a)\right)\,.$$

We slightly abuse the notation and define the two analogous oracle processes used in the proofs of the results from Section B as follows—where the processes that has oracle access to the numeraire portfolios in each sub-experiment and for each strata in that sub-experiment is given by

$$W^{(\mathsf{Q})}_{N_a(t)} \coloneqq \prod_{s \in \mathcal{T}_t(a)} \prod_{i \in \mathcal{I}_s} \prod_{\mathbf{h} \in \mathcal{H}_{s-1}} \left(\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a,\mathbf{h})^\top \mathbf{E}_{i,s}(a,\mathbf{h})\right)^{\xi_{i,s}(a,\mathbf{h})}\,, \tag{20}$$

where $\boldsymbol{\lambda}_s^{(\mathsf{Q})}(a,\mathbf{h})$ is defined as in Equation (19) for each $s \in [t]$, $a \in \mathcal{A}_s$, and $\mathbf{h} \in \mathcal{H}_{s-1}$. While the second process is the one that has oracle access to the best in hindsight portfolio for each of the active treatment arms in every sub-experiment and for each stratum $\mathbf{h} \in \mathcal{H}_{s-1}$ takes the form

$$W^{\star}_{N_a(t)} \coloneqq \prod_{s \in \mathcal{T}_t(a)} \prod_{i \in \mathcal{I}_s} \prod_{\mathbf{h} \in \mathcal{H}_{s-1}} \left(\boldsymbol{\lambda}_s^{\star}(a,\mathbf{h})^\top \mathbf{E}_{i,s}(a,\mathbf{h})\right)^{\xi_{i,s}(a,\mathbf{h})}$$

$$\text{where for each } s \in [t] \text{ and } \mathbf{h} \in \mathcal{H}_{s-1}\,,\ \boldsymbol{\lambda}_s^{\star}(a,\mathbf{h}) \coloneqq \operatorname*{arg\,max}_{\boldsymbol{\lambda} \in \triangle_1} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a,\mathbf{h}) \log\left(\boldsymbol{\lambda}^\top \mathbf{E}_{i,s}(a,\mathbf{h})\right)\,.$$

It will be clear from context which of these processes we are invoking in each of the proofs.

D.1 PROOF OF LEMMA 3.4

Proof of Lemma 3.4. To show that $W_{N_a(t)}$ (given in (2)) forms a test supermartingale we must argue that $W_{N_a(t)}$ is nonnegative almost surely, that it forms a supermartingale, and that its mean is upper bounded by one. Indeed, nonnegativity follows by construction so all it remains to show is that $W_{N_a(t)}$ is a P-supermartingale and that its mean is upper bounded by one.

We start by showing that $W_{N_a(t)}$ is a P-supermartingale. To do so, we first notice how $W_{N_a(t)}$ can be equivalently characterized as

$$W_{N_a(t)} = \prod_{s=1}^{t} \left(W_{\check{N}_a(s)}\right)^{\mathbb{1}\{a \in \mathcal{A}_s\}} .$$

Let $n^\star$ be the index of the last unit to enter the experiment—which corresponds to the index of the last unit to enter sub-experiment t—and $n^\star - 1$ correspond to the index of the second to last unit to enter the experiment. The conditional expectation of $W_{N_a(t)}$ simplifies to

$$\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[W_{N_a(t)} \mid \mathcal{F}_{n^\star-1,t}\right] = \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\left(W_{\check{N}_a(t)}\right) \,\middle|\, \mathcal{F}_{n^\star-1,t}\right] \prod_{s=1}^{t-1} \left(W_{\check{N}_a(s)}\right)^{\mathbb{1}\{a \in \mathcal{A}_s\}} .$$

From Assumption 3.2 we have that the decision to keep treatment a in sub-experiment t is $\mathcal{F}_{t-1}$-measurable, which means that if $a \notin \mathcal{A}_t$ then $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[W_{N_a(t)} \mid \mathcal{F}_{n^\star-1,t}\right] = W_{N_a(t-1)}$ $\mathsf{P}_{\mathrm{RCT}}$-almost surely. On the other hand, if $a \in \mathcal{A}_t$ we have that the conditional expectation term is equal to

$$\begin{aligned}
&\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\prod_{i\in\mathcal{I}_t} \left(\boldsymbol{\lambda}_{i,t}^\top \mathbf{E}_{i,t}(a)\right)^{\xi_{i,t}(a)} \,\middle|\, \mathcal{F}_{n^\star-1,t}\right] \\
&\quad= \underbrace{\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\left(\boldsymbol{\lambda}_{n^\star,t}^\top \mathbf{E}_{n^\star,t}(a)\right)^{\xi_{n^\star,t}} \,\middle|\, \mathcal{F}_{n^\star-1,t}\right]}_{(21).(1)} \prod_{i\in\mathcal{I}_t(n^\star-1)} \left(\boldsymbol{\lambda}_{i,t}^\top \mathbf{E}_{i,t}(a)\right)^{\xi_{i,t}(a)} . \qquad (21)
\end{aligned}$$

We now focus on $(21).(1)$ and apply to law of total expectation to see how it can be upper bounded as

$$\begin{aligned}
(21).(1) &= \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right) \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\boldsymbol{\lambda}_{n^\star,t}^\top \mathbf{E}_{n^\star,t}(a) \mid \mathcal{F}_{n^\star-1,t}, \xi_{n^\star,t}(a) = 1\right] \\
&\quad + \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 0 \mid \mathcal{F}_{n^\star-1,t}\right) \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[1 \mid \mathcal{F}_{n^\star-1,t}, \xi_{n^\star,t} = 0\right] \\
&= \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right) \Big((1 - \lambda_{n^\star,t}) \\
&\quad + \lambda_{n^\star,t} \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[g_t(\widehat{\psi}_{n^\star,t}(a)) \,\middle|\, \mathcal{F}_{n^\star-1,t}, \xi_{n^\star,t}(a) = 1\right] / g_t(\delta)\Big) \qquad (22) \\
&\quad + \left(1 - \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\right) \\
&= \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right) \left((1 - \lambda_{n^\star,t}) + \lambda_{n^\star,t} g_t(\psi_{\mathsf{P}}(a))/g_t(\delta)\right) \qquad (23) \\
&\quad + \left(1 - \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\right) \\
&\leq \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right) \left((1 - \lambda_{n^\star,t}) + \lambda_{n^\star,t}\right) + \left(1 - \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t} = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\right) \qquad (24) \\
&= 1 ,
\end{aligned}$$

where (22) follows from the predictability of the portfolios and the transformation $g_t(\cdot)$ on the threshold δ; (23) follows from the assumption that $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[g_t(\widehat{\psi}_{n^\star,t}(a)) \,\middle|\, \mathcal{F}_{n^\star-1,t}, \xi_{n^\star,t}(a) = 1\right] = g_t(\psi_{\mathsf{P}}(a))$; and (24) follows because under the null, $\mathcal{P}(a)$, $\psi_{\mathsf{P}}(a) \leq \delta$.

Putting the previous steps together we can conclude that

$$\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[W_{N_a(t)} \mid \mathcal{F}_{n^\star-1}, t\right] \leq W_{N_a(t-1)} \left[\prod_{i\in\mathcal{I}_t(n^\star-1)} \left(\boldsymbol{\lambda}_{i,t}^\top \mathbf{E}_{i,t}(a)\right)^{\xi_{i,t}(a)}\right]^{a\in\mathcal{A}_t} ,$$

which shows that $W_{N_a(t)}$ is a supermartingale. Instantiating the above for the first sub-experiment with only one unit, and through similar arguments we can conclude that $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\left(\boldsymbol{\lambda}_{1,1}^\top \mathbf{E}_{1,1}(a)\right)^{\xi_{1,1}(a)}\right] \leq 1$, which completes the proof. □

D.2 Proof of Proposition 3.5

Proof of Proposition 3.5. Fix a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, a sub-experiment $s \in [t]$, and define $\overline{\mathcal{A}}_s \coloneqq \bigcup_{r=1}^{s} \mathcal{A}_r$. We start by noting how the number of false discoveries can be written as

$$|\mathcal{D}_s \cap \mathcal{H}_0| = \sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \mathbb{1}\left\{W_{N_a(s)} \geq 1/\alpha_a\right\},$$

where we recall that for any treatment $a \in \mathcal{A}_s$, the treatment is removed from the experiment as soon as its test supermartingale surpasses the threshold. This is equivalent to "freezing" the test supermartingale for treatment a; that is, for all subsequent sub-experiments the value of the test supermartingale is not updated.

By the definition of sub-experiments, we know that $\mathcal{D}_s \supseteq \mathcal{D}_{s-1}$ for all $s \in [t]$ where we let $\mathcal{D}_0 = \varnothing$. This implies that $|\mathcal{D}_s| \geq |\mathcal{D}_{s-1}|$ for all $s \in [t]$, where $|\varnothing| = 0$. Let $s_0(a) \coloneqq \min\{r \in [s] \mid a \in \mathcal{A}_r\}$ be the sub-experiment in which treatment a was introduced. We therefore have the inequalities $|\mathcal{D}_s| \geq \left|\mathcal{D}_{s_0(a)}\right| \geq \left|\mathcal{D}_{s_0(a)-1}\right|$. We use these inequalities to upper bound the proportion of false discoveries as

$$\begin{aligned}
\frac{|\mathcal{D}_s \cap \mathcal{H}_0|}{|\mathcal{D}_s| \vee 1} &= \sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \frac{\mathbb{1}\left\{W_{N_a(s)} \geq 1/\alpha_a\right\}}{|\mathcal{D}_s| \vee 1} \\
&\leq \sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \frac{\mathbb{1}\left\{W_{N_a(s)} \geq 1/\alpha_a\right\}}{\left|\mathcal{D}_{s_0(a)-1}\right| + 1} \\
&\leq \sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \frac{\alpha_a W_{N_a(s)}}{\left|\mathcal{D}_{s_0(a)-1}\right| + 1}. \qquad (25)
\end{aligned}$$

Recall the definition of $\alpha_a \coloneqq \alpha\gamma_a\left(\left|\mathcal{D}_{s_0(a)-1}\right| + 1\right)$. Substituting the definition of α_a into (25) gives us

$$\begin{aligned}
(25) &= \sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \frac{\alpha\gamma_a\left(\left|\mathcal{D}_{s_0(a)-1}\right| + 1\right) W_{N_a(s)}}{\left|\mathcal{D}_{s_0(a)-1}\right| + 1} \\
&= \sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \alpha\gamma_a W_{N_a(s)}.
\end{aligned}$$

The above upper bounds are deterministic, which allow us to upper bound the false discover rate as

$$\mathrm{FDR}(\mathcal{D}_s) \leq \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \alpha\gamma_a W_{N_a(s)}\right].$$

We now focus on the upper bound on the false discovery rate and note how due to the law of total expectation it can be equivalently written as

$$\begin{aligned}
&\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \alpha\gamma_a W_{N_a(s)}\right] \\
&= \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \mathcal{A}'_s \cap \mathcal{H}_0} \alpha\gamma_a W_{N_a(s)} + \sum_{a \in \overline{\mathcal{A}}_{s-1} \cap \mathcal{H}_0} \alpha\gamma_a W_{N_a(s)} \,\middle|\, \mathcal{F}_{s-1}\right]\right] \\
&= \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \mathcal{A}'_s \cap \mathcal{H}_0} \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\alpha\gamma_a W_{N_a(s)} \mid \mathcal{F}_{s-1}\right]\right] + \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \overline{\mathcal{A}}_{s-1} \cap \mathcal{H}_0} \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\alpha\gamma_a W_{N_a(s)} \mid \mathcal{F}_{s-1}\right]\right], \qquad (26)
\end{aligned}$$

where the last equality follows from the $\mathcal{F}_{s-1}$-measurability of the treatments in $\overline{\mathcal{A}}_s$ and the linearity of expectations. Since for each $a \in \mathcal{A}'_s$ we have that $W_{N_a(s)}$ is a test supermartingale we have the

following upper bound on (26),

$$(26) \leq \alpha \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \mathcal{A}'_s \cap \mathcal{H}_0} \gamma_a\right] + \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \overline{\mathcal{A}}_{s-1} \cap \mathcal{H}_0} \alpha \gamma_a W_{N_a(s-1)}\right].$$

By continuously applying the above arguments to the resulting upper bounds, we end up with the upper bound

$$\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{a \in \overline{\mathcal{A}}_s \cap \mathcal{H}_0} \alpha \gamma_a W_{N_a(s)}\right] \leq \alpha \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{r=1}^{s} \sum_{a \in \mathcal{A}'_r \cap \mathcal{H}_0} \gamma_a\right]. \tag{27}$$

Since the sequence $(\gamma_r)_{r=1}^{\infty}$ is chosen to satisfy the inequality $\sum_{r=1}^{\infty} \gamma_r \leq 1$, we have that the upper bound in (27) can be further upper bounded as

$$\alpha \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{r=1}^{s} \sum_{a \in \mathcal{A}'_r \cap \mathcal{H}_0} \gamma_a\right] \leq \alpha \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\sum_{r=1}^{\infty} \gamma_r\right] \leq \alpha\,.$$

Hence, we can conclude that

$$\mathrm{FDR}(\mathcal{D}_s) \leq \alpha\,.$$

Since the choice of $s \in [t]$ was arbitrary, we have that

$$\mathrm{FDR}(\mathcal{D}_s) \leq \alpha \quad \text{for all } s \in [t]\,,$$

completing the proof. □

D.3 Proof of Lemma 4.2

Proof of Lemma 4.2. Fix a $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, a sub-experiment $s \in [t]$, and any treatment $a \in \mathcal{A}_s$. Our goal is to show that the Horivtz-Thompson estimator from (5) is conditionally unbiased when we condition on the filtration and the event that the inclusion indicator is equal to one. Hence, we can write the conditional expectation as

$$\begin{aligned}\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\widehat{\psi}_{i,s}(a) \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] &= \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(a)\frac{\mathbb{1}\{A_{i,s} = a\}}{\widehat{\pi}_s(a)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] \\ &\quad - \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(0)\frac{\mathbb{1}\{A_{i,s} = 0\}}{\widehat{\pi}_s(0)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right].\end{aligned}$$

We now focus on the conditional expectation under treatment a. Due to the $\mathcal{F}_{s-1}$-measurability of $\pi_s(a)$ and $\pi_s(0)$—and hence of $\widehat{\pi}_s(a)$—alongside the conditional conditional ignorability assumption (Assumption 3.2.(i)) we can further rewrite the conditional expectation as

$$\begin{aligned}&\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(a)\frac{\mathbb{1}\{A_{i,s} = a\}}{\widehat{\pi}_s(a)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] \\ &\quad = \frac{1}{\widehat{\pi}_s(a)}\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(a) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right]\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\{A_{i,s} = a\} \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right]. \quad (28)\end{aligned}$$

Since units are independent and identically distributed together with the conditional ignorability assumption (Assumption 3.2.(i)) we have the identity $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(a) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] = \mathbb{E}_{\mathsf{P}}\left[Y_1(a)\right]$.

We now turn our attention to the conditional expectation of the indicator. The conditional expectation of the indicator follows the identity $\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\{A_{i,s} = a\} \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] = \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(A_{i,s} = a \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right)$. Since we are conditioning on the event $\{\xi_{i,s}(a) = 1\}$ we know that $A_{i,s} = a$ or $A_{i,s} = 0$. This, in turn, implies that

$$\mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(A_{i,s} = a \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right) = \frac{\pi_s(a)}{\pi_s(a) + \pi_s(0)} \equiv \widehat{\pi}_s(a)\,.$$

The above arguments imply that (28) is equal to

$$\frac{1}{\widehat{\pi}_s(a)}\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(a) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right]\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\{A_{i,s} = a\} \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a) = 1\right] = \mathbb{E}_{\mathsf{P}}\left[Y_1(a)\right].$$

Through the same set of steps we can conclude that

$$\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_{i,s}(0)\frac{\mathbb{1}\left\{A_{i,s}=0\right\}}{\widehat{\pi}_s(0)}\;\middle|\;\mathcal{F}_{i-1,s},\xi_{i,s}(a)=1\right]=\mathbb{E}_{\mathsf{P}}\left[Y_1(0)\right].$$

Hence, all of the above steps imply that

$$\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\widehat{\psi}_{i,s}(a)\;\middle|\;\mathcal{F}_{i-1,s},\xi_{i,s}(a)=1\right]=\psi_{\mathsf{P}}(a)\,,$$

completing the proof of Lemma 4.2. □

D.4 Proof of Proposition 4.3

Proof of Proposition 4.3. Proposition 4.3 is an immediate consequence of Lemmas 3.4 and 4.2, and the fact that $g_s(\cdot)$ is a non-decreasing linear map for each $s \in [t]$. □

D.5 Proof of Theorem 4.5

In this subsection we present the proof of Theorem 4.5. Nonetheless, before providing the proof we present the following lemma (whose proof is given in Section E.1) that is central to the proof of Theorem 4.5.

Lemma D.1. *Fix an arm $a \in \mathcal{A}$ and a probability $\mathsf{Q} \in \mathcal{Q}(a)$. Let $t \in \mathbb{N}$ denote the number of sub-experiments that have transpired in the experiment, and let $R_{\check{N}_a(s)} \coloneqq \log\left(\check{N}_a(s)+1\right)/2+\log(2)$ be the upper bound on the portfolio regret of the Universal Portfolio algorithm (Cover & Ordentlich, 1996) used to select the portfolios under treatment a in each sub-experiment $s \in \mathcal{T}_t(a)$. Then, under Assumptions 3.1 to 3.3 and 4.1 the following inequality holds for any $\varepsilon > 0$,*

$$\begin{aligned}&\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\left|\log\left(W_{N_a(t)}\right)-\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right|\geq\varepsilon\right)\\&\quad\leq\sum_{n_{\underline{t}}=m}^{\infty}\exp\left\{-2(1-\exp\{-\varepsilon\})\sum_{s\in\mathcal{T}_t(a)}n_s p_s\right\}+\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\sum_{s\in\mathcal{T}_t(a)}R_{\check{N}_a(s)}\geq\varepsilon\right).\end{aligned}$$

With Lemma D.1 in hand, we now proceed to prove Theorem 4.5. The proof follows two steps. In the first step, we show that the normalized $\log\left(W_{N_a(t)}\right)$ and $\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)$ are almost surely equivalent as the total number of units who enter the experiment goes to infinity. In the second step, we use the aforementioned result to show that in the limit no other process $\widetilde{W}_{N_a(t)}$ can exhibit a larger growth rate than $W_{N_a(t)}$.

Proof of the almost sure equivalence with the oracle process. We begin by upper bounding the following probability for any $\varepsilon > 0$,

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\left|\log\left(W_{N_a(t)}\right)-\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right|\geq\varepsilon\right). \tag{29}$$

From Lemma D.1 we have that

$$\begin{aligned}(29)&\leq\sum_{n_{\underline{t}}=m}^{\infty}\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)}\left|\log\left(W_{N_a(t)}\right)-\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right|\geq\varepsilon\right)\\&\leq\underbrace{\sum_{n_{\underline{t}}=m}\exp\left\{-2(1-\exp\{-\varepsilon\})\sum_{s\in\mathcal{T}_t(a)}n_s p_s\right\}}_{(30).(1)}+\underbrace{\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\sum_{s\in\mathcal{T}_t(a)}R_{\check{N}_a(s)}\geq\varepsilon\right)}_{(30).(2)}.\end{aligned} \tag{30}$$

We now recall the fact that for any random variables $(X_n)_{n\in\mathbb{N}}$ drawn from a distribution P,

$$\mathbb{P}_{\mathsf{P}}\left(\lim_{n\to\infty}|X_n|=0\right)=1 \quad \text{if and only if} \quad \forall\varepsilon>0\,, \quad \lim_{m\to\infty}\mathbb{P}_{\mathsf{P}}\left(\sup_{n\geq m}|X_n|>\varepsilon\right)=0\,.$$

Hence, as we take $m\to\infty$ we can see that (30).$(1)\to 0$ almost surely. Similarly, we have that (30).$(2)\to 0$ as $m\to\infty$, which follows from Lemma E.1. Thus, we can conclude that

$$\lim_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\left(\log\left(W_{N_a(t)}\right)-\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right)=0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-almost surely}\,. \tag{31}$$

□

Proof of the log-optimality of $W_{N_a(t)}$. We begin by lower bounding the difference as

$$\begin{aligned}
&\liminf_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\left(\log\left(W_{N_a(t)}\right)-\log\left(\widetilde{W}_{N_a(t)}\right)\right)\\
&\quad=\liminf_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\left(\log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)+\log\left(W^{(\mathsf{Q})}_{N_a(t)}/\widetilde{W}_{N_a(t)}\right)\right)\\
&\quad\geq\underbrace{\liminf_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)}_{(32).(1)}+\underbrace{\liminf_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\log\left(W^{(\mathsf{Q})}_{N_a(t)}/\widetilde{W}_{N_a(t)}\right)}_{(32).(2)}
\end{aligned} \tag{32}$$

Notice how (32).$(1)=0$ which follows from the proof of the almost sure equivalence between $W_{N_a(t)}$ and the oracle process—the result which is given in (31). Hence, all it remains to show is that (32).$(2)\geq 0$.

From the arm-wise numeraire property (Sandoval et al., 2026, Lemma B.1) we have that for any $\varepsilon_{n_{\underline{t}}}>0$,

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\geq\varepsilon_{n_{\underline{t}}}\right)\leq\frac{1}{\varepsilon_{n_{\underline{t}}}}\,,$$

from which we can also conclude that

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)}\log\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)\geq\frac{1}{N_a(t)}\log\left(\varepsilon_{n_{\underline{t}}}\right)\right)\leq\frac{1}{\varepsilon_{n_{\underline{t}}}}\,.$$

Letting $\varepsilon_{n_{\underline{t}}}=n_{\underline{t}}^2$ and summing over $n_{\underline{t}}$ we can see how

$$\sum_{n_{\underline{t}}=1}^{\infty}\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)}\log\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)\geq\frac{2\log\left(n_{\underline{t}}\right)}{N_a(t)}\right)\leq\sum_{n_{\underline{t}}=1}^{\infty}\frac{1}{n_{\underline{t}}^2}=\frac{\pi^2}{6}\,.$$

Hence, by the Borel-Cantelli lemma we have that

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\limsup_{n_{\underline{t}}\to\infty}\left(\frac{1}{N_a(t)}\log\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)-\frac{2\log\left(n_{\underline{t}}\right)}{N_a(t)}\right)>0\right)=0\,.$$

By Lemma E.5 and the definition of $n_{\underline{t}}$ it is the case that $2\log\left(n_{\underline{t}}\right)/N_a(t)\to 0$ $\mathsf{Q}_{\mathrm{RCT}}$-almost surely, which together with the above arguments we can conclude that

$$\liminf_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\log\left(W^{(\mathsf{Q})}_{N_a(t)}/\widetilde{W}_{N_a(t)}\right)\geq 0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-almost surely}\,.$$

Putting all of the previous steps together we can conclude that

$$\liminf_{n_{\underline{t}}\to\infty}\frac{1}{N_a(t)}\left(\log\left(W_{N_a(t)}\right)-\log\left(\widetilde{W}_{N_a(t)}\right)\right)\geq 0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-almost surely}\,,$$

completing the proof of the log-optimality of $W_{N_a(t)}$. □

D.6 Proof of Lemma B.3

Proof of Lemma B.3. Fix a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$, a sub-experiment $s \in [t]$, a treatment arm $a \in \mathcal{A}_s$, and a history $\mathbf{h} \in \mathcal{H}_{s-1}$. Taking the conditional expectation of the Horvitz-Thompson estimator for treatment a for units whose history is equal to $\mathbf{h}$ in sub-experiment s, we have that:

$$\begin{aligned}
&\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\widehat{\psi}_{i,s}(a,\mathbf{h}) \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] \\
&= \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_{i,s}\left(\frac{\mathbb{1}\left\{A_{i,s}=a\right\}}{\widehat{\pi}_s(a)} - \frac{\mathbb{1}\left\{A_{i,s}=0\right\}}{\widehat{\pi}_s(0)}\right) \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] \\
&= \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, A_{i,s})\left(\frac{\mathbb{1}\left\{A_{i,s}=a\right\}}{\widehat{\pi}_s(a)} - \frac{\mathbb{1}\left\{A_{i,s}=0\right\}}{\widehat{\pi}_s(0)}\right) \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right],
\end{aligned} \tag{33}$$

where the second equality follows from the definition of the observed outcome in sub-experiment s. Due to the linearity of expectation we can see how (33) can be rewritten as:

$$\begin{aligned}
(33) = \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}&\left[\frac{Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a)\mathbb{1}\left\{A_{i,s}=a\right\}}{\widehat{\pi}_t(a)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] \\
&- \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\frac{Y_i(\mathbf{Z}_{i,s-\ell:s-1}, 0)\mathbb{1}\left\{A_{i,s}=0\right\}}{\widehat{\pi}_s(0)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right].
\end{aligned}$$

Focusing on the first summand we can see how it can be further simplified to:

$$\begin{aligned}
&\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\frac{Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a)\mathbb{1}\left\{A_{i,s}=a\right\}}{\widehat{\pi}_s(a)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] \\
&= \frac{\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right]\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\left\{A_{i,s}=a\right\} \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right]}{\widehat{\pi}_s(a)} \\
&= \frac{\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right]\widehat{\pi}_s(a)}{\widehat{\pi}_s(a)} \\
&= \mathbb{E}_{\mathsf{P}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right],
\end{aligned}$$

where the first equality follows from the independence of the treatment assignment (Assumption 3.2) and the $\mathcal{F}_{s-1}$-measurability of the propensity score. The second equality follows because the treatment assignments in each sub-experiment are drawn from the same distribution no matter which history $\mathbf{h}$ the unit experiences, and since conditioning on the event $\{\xi_{i,s}(a,\mathbf{h}) = 1\}$ implies that $A_{i,s} = a$ or $A_{i,s} = 0$. Hence, we have that

$$\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\left\{A_{i,s}=a\right\} \mid \mathcal{F}_{s-1}, \xi_{i,s}(a) = 1\right] = \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(A_{i,s}=a \mid \mathcal{F}_{s-1}, \xi_{i,s}(a) = 1\right) = \frac{\pi_s(a)}{\pi_s(a)+\pi_s(0)} \equiv \widehat{\pi}_s(a).$$

Through the same set of steps one can show that

$$\mathbb{E}_{\mathsf{P}}\left[\frac{Y_i(\mathbf{Z}_{i,s-\ell:s-1}, 0)\mathbb{1}\left\{A_{i,s}=0\right\}}{\widehat{\pi}_s(0)} \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] = \mathbb{E}_{\mathsf{P}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, 0) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right].$$

Hence, putting the previous steps together we have that

$$\mathbb{E}_{\mathsf{P}}\left[\widehat{\psi}_{i,s}(a,\mathbf{h}) \;\middle|\; \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right] = \mathbb{E}_{\mathsf{P}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) - Y_i(\mathbf{Z}_{i,s-\ell:s-1}, 0) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h}) = 1\right].$$

The result follows by noting how the potential outcomes for unit i are independent from those of all other units $j \in \mathcal{I}$ (such that $j \neq i$) as well as their treatment assignments. Furthermore, from Assumption 3.2.(i) and Assumption B.2 we have that the expected outcomes in the current sub-experiment s only depend on the observed outcomes and treatment assignments from the previous ℓ sub-experiments. We therefore have that

$$\begin{aligned}
&\mathbb{E}_{\mathsf{P}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) - Y_i(\mathbf{Z}_{i,s-\ell:s-1}, 0) \mid \mathcal{F}_{i-1,s}, \xi_{i,s}(a,\mathbf{h})\right] \\
&\quad = \mathbb{E}_{\mathsf{P}}\left[Y_i(\mathbf{Z}_{i,s-\ell:s-1}, a) - Y_i(\mathbf{Z}_{i,s-\ell:s-1}, 0) \mid \mathbf{H}_{i,s} = \mathbf{h}\right] \equiv \psi_{\mathsf{P}}(a,\mathbf{h}),
\end{aligned}$$

completing the proof. □

D.7 Proof of Theorem B.6

In this subsection we present the proof of Theorem B.6. It follows the same steps as the proof of Theorem 4.5, but we provide them here for the sake of completeness. Nonetheless, before providing the proof we state the following lemma (whose proof is given in Section E.2) that plays a central role in the proof of Theorem B.6.

Lemma D.2. *Fix an arm $a \in \mathcal{A}$ and a probability $\mathsf{Q} \in \mathcal{Q}(a)$. Let $t \in \mathbb{N}$ denote the number of sub-experiments that were run, and let $R_{\check{N}_{a,\mathbf{h}}(s)} \coloneqq \log\left(\check{N}_{a,\mathbf{h}}(s)+1\right)/2 + \log(2)$ be the upper bound on the portfolio regret of the Universal Portfolio algorithm used to select the portfolios under treatment a, for history $\mathbf{h}$, in each sub-experiment $s \in \mathcal{T}_t(a)$. Then, under Assumptions B.1, B.2 and 3.1 to 3.3 the following inequality holds for any $\varepsilon > 0$,*

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \left|\log\left(W_{N_a(t)}\right) - \log\left(W_{N_a(t)}\right)^{(\mathsf{Q})}\right| \geq \varepsilon\right)$$
$$\leq \sum_{n_{\underline{t}}=m}^{\infty} \exp\left\{-2(1-\exp\{-\varepsilon\}) \sum_{s\in\mathcal{T}_t(a)} n_s p_s\right\} + \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \sum_{s\in\mathcal{T}_t(a)} \sum_{\mathbf{h}\in\mathcal{H}_{s-1}} R_{\check{N}_{a,\mathbf{h}}(s)} \geq \varepsilon\right).$$

Having stated Lemma D.2, we now prove Theorem B.6. The proof proceeds in two steps. In the first step, we show that $W_{N_a(t)}$ is almost surely equivalent to a process that has oracle access to all the numeraire portfolios, $W^{(\mathsf{Q})}_{N_a(t)}$. We then use this result in the second step to argue about the log-optimality of our process.

Proof of the almost sure equivalence with the oracle process. We begin by upper bounding the following probability for any $\varepsilon > 0$,

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \left|\log\left(W_{N_a(t)}\right) - \log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right| \geq \varepsilon\right). \tag{34}$$

From Lemma D.2 we have that

$$\begin{aligned}(34) &\leq \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)} \left|\log\left(W_{N_a(t)}\right) - \log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right| \geq \varepsilon\right) \\ &\leq \underbrace{\sum_{n_{\underline{t}}=m} \exp\left\{-2(1-\exp\{-\varepsilon\}) \sum_{s\in\mathcal{T}_t(a)} n_s p_s\right\}}_{(35).(1)} + \underbrace{\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \sum_{s\in\mathcal{T}_t(a)} \sum_{\mathbf{h}\in\mathcal{H}_{s-1}} R_{\check{N}_{a,\mathbf{h}}(s)} \geq \varepsilon\right)}_{(35).(2)}.\end{aligned} \tag{35}$$

We now recall the fact that for any random variables $(X_n)_{n\in\mathbb{N}}$ drawn from a distribution P,

$$\mathbb{P}_{\mathsf{P}}\left(\lim_{n\to\infty} |X_n| = 0\right) = 1 \quad \text{if and only if} \quad \forall \varepsilon > 0\,, \quad \lim_{m\to\infty} \mathbb{P}_{\mathsf{P}}\left(\sup_{n\geq m} |X_n| > \varepsilon\right) = 0\,.$$

Hence, as we take $m \to \infty$ we can see that $(35).(1) \to 0$ almost surely. Similarly, we have that $(35).(2) \to 0$ as $m \to \infty$, which follows from Lemma E.2. Thus, we can conclude that

$$\lim_{n_{\underline{t}}\to\infty} \frac{1}{N_a(t)} \left(\log\left(W_{N_a(t)}\right) - \log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right) = 0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-almost surely}\,, \tag{36}$$

completing the proof of the almost sure equivalence between our process and the oracle process. □

Proof of the log-optimality of $W_{N_a(t)}$. We begin by lower bounding the difference as

$$\begin{aligned}
&\liminf_{n_{\underline{t}}\to\infty} \frac{1}{N_a(t)} \left(\log\left(W_{N_a(t)}\right) - \log\left(\widetilde{W}_{N_a(t)}\right)\right) \\
&\quad = \liminf_{n_{\underline{t}}\to\infty} \frac{1}{N_a(t)} \left(\log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) + \log\left(W^{(\mathsf{Q})}_{N_a(t)}/\widetilde{W}_{N_a(t)}\right)\right) \\
&\quad \geq \underbrace{\liminf_{n_{\underline{t}}\to\infty} \frac{1}{N_a(t)} \log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)}_{(37).(1)} + \underbrace{\liminf_{n_{\underline{t}}\to\infty} \frac{1}{N_a(t)} \log\left(W^{(\mathsf{Q})}_{N_a(t)}/\widetilde{W}_{N_a(t)}\right)}_{(37).(2)}
\end{aligned} \tag{37}$$

Notice how $(37).(1) = 0$ which follows from the proof of the almost sure equivalence between $W_{N_a(t)}$ and the oracle process (36). Hence, all it remains to show is that $(37).(2) \geq 0$.

From the arm-wise numeraire property (Sandoval et al., 2026, Lemma B.1) we have that for any $\varepsilon_{n_{\underline{t}}} > 0$,

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)} \geq \varepsilon_{n_{\underline{t}}}\right) \leq \frac{1}{\varepsilon_{n_{\underline{t}}}},$$

from which we can also conclude that

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)} \log\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq \frac{1}{N_a(t)} \log\left(\varepsilon_{n_{\underline{t}}}\right)\right) \leq \frac{1}{\varepsilon_{n_{\underline{t}}}}.$$

Letting $\varepsilon_{n_{\underline{t}}} = n_{\underline{t}}^2$ and summing over $n_{\underline{t}}$ we can see how

$$\sum_{n_{\underline{t}}=1}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)} \log\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq \frac{2\log\left(n_{\underline{t}}\right)}{N_a(t)}\right) \leq \sum_{n_{\underline{t}}=1}^{\infty} \frac{1}{n_{\underline{t}}^2} = \frac{\pi^2}{6}.$$

Hence, by the Borel-Cantelli lemma we have that

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\limsup_{n_{\underline{t}}\to\infty}\left(\frac{1}{N_a(t)} \log\left(\widetilde{W}_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) - \frac{2\log\left(n_{\underline{t}}\right)}{N_a(t)}\right) > 0\right) = 0.$$

By Lemma E.5 and the definition of $n_{\underline{t}}$ it is the case that $2\log\left(n_{\underline{t}}\right)/N_a(t) \to 0$ RCT-almost surely, which together with the above arguments we can conclude that

$$\liminf_{n\to\infty} \frac{1}{N_a(t)} \log\left(W^{(\mathsf{Q})}_{N_a(t)}/\widetilde{W}_{N_a(t)}\right) \geq 0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-almost surely}.$$

Putting the steps together we can conclude that

$$\liminf_{n_{\underline{t}}\to\infty} \frac{1}{N_a(t)} \left(\log\left(W_{N_a(t)}\right) - \log\left(\widetilde{W}_{N_a(t)}\right)\right) \geq 0 \quad \mathsf{Q}_{\mathrm{RCT}}\text{-almost surely},$$

completing the proof of the log-optimality of $W_{N_a(t)}$. □

E Auxiliary Results

E.1 Proof of Lemma D.1

Proof of Lemma D.1. Our aim is to control the following probability for any $\varepsilon > 0$,

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m} \frac{1}{N_a(t)} \left|\log\left(W_{N_a(t)}\right) - \log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right| \geq \varepsilon\right). \tag{38}$$

We start by upper bounding it by the probability of the upper and lower deviations

$$(38) \leq \underbrace{\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m} \frac{1}{N_a(t)} \log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq \varepsilon\right)}_{(39).(1)} + \underbrace{\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m} \frac{1}{N_a(t)} \log\left(W^{(\mathsf{Q})}_{N_a(t)}/W_{N_a(t)}\right) \geq \varepsilon\right)}_{(39).(2)}, \tag{39}$$

and we proceed by analyzing the terms in the upper bound independently.

We first analyze (39).(1), and proceed as follows:

$$\begin{aligned}
(39).(1) &\leq \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)}\log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq \varepsilon\right) \\
&= \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)} \geq \exp\left\{N_a(t)\varepsilon\right\}\right) \\
&= \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\exp\left\{-N_a(t)\varepsilon\right\}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq 1\right) \\
&\leq \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\left\{-N_a(t)\varepsilon\right\}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)\right] \\
&\leq \sum_{n_{\underline{t}}=m}^{\infty} \exp\left\{-2(1-\exp\left\{-\varepsilon\right\})\sum_{s\in\mathcal{T}_t(a)} n_s p_s\right\}
\end{aligned}$$

where the first inequality is follows from a union bound, the second inequality is due to Markov's inequality, and the third inequality is due to Lemma E.3.

We now turn our attention to (39).(2), and recall that by definition $\log\left(W^{\star}_{N_a(t)}\right) \geq \log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)$. Hence, we have the following series of upper bounds

$$\begin{aligned}
(39).(2) &\leq \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\left(\log\left(W^{\star}_{N_a(t)}\right)-\log\left(W_{N_a(t)}\right)\right) \geq \varepsilon\right) \\
&= \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\left(\sum_{s\in\mathcal{T}_t(a)}\left[\sum_{i\in\mathcal{I}_s}\xi_{i,s}(a)\log\left(\boldsymbol{\lambda}^{\star}_s(a)^{\top}\mathbf{E}_{i,s}(a)\right)-\sum_{i\in\mathcal{I}_s}\xi_{i,s}(a)\log\left(\boldsymbol{\lambda}_{i,s}(a)^{\top}\mathbf{E}_{i,s}(a)\right)\right]\right) \geq \varepsilon\right) \\
&\leq \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\sum_{s\in\mathcal{T}_t(a)} R_{\check{N}_a(s)} \geq \varepsilon\right),
\end{aligned}$$

where the second inequality follows because for each arm and in each sub-experiment we run a new instance of the Universal Portfolio algorithm (Cover, 1991) whose upper bound on the portfolio regret is denoted by $R_{\check{N}_a(s)}$ for each sub-experiment $s \in \mathcal{T}_t(a)$.

Putting the above steps together we have that

$$\begin{aligned}
&\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\left|\log\left(W_{N_a(t)}\right)-\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right| \geq \varepsilon\right) \\
&\quad\leq \sum_{n_{\underline{t}}=m}^{\infty}\exp\left\{-2(1-\exp\left\{-\varepsilon\right\})\sum_{s\in\mathcal{T}_t(a)} n_s p_s\right\} + \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\sum_{s\in\mathcal{T}_t(a)} R_{\check{N}_a(s)} \geq \varepsilon\right),
\end{aligned}$$

completing the proof. □

E.2 Proof of Lemma D.2

Proof of Lemma D.2. Our aim is to control the following probability for any $\varepsilon > 0$,

$$\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\frac{1}{N_a(t)}\left|\log\left(W_{N_a(t)}\right)-\log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right| \geq \varepsilon\right). \tag{40}$$

We start by upper bounding it by the probability of the upper and lower deviations

$$(40) \leq \underbrace{\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq \varepsilon\right)}_{(41).(1)} + \underbrace{\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \log\left(W^{(\mathsf{Q})}_{N_a(t)}/W_{N_a(t)}\right) \geq \varepsilon\right)}_{(41).(2)}, \tag{41}$$

and we proceed by analyzing the terms in the upper bound independently.

We first analyze (41).(1), and proceed as follows:

$$\begin{aligned}
(41).(1) &\leq \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\frac{1}{N_a(t)} \log\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq \varepsilon\right) \\
&= \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)} \geq \exp\left\{N_a(t)\varepsilon\right\}\right) \\
&= \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\exp\left\{-N_a(t)\varepsilon\right\}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right) \geq 1\right) \\
&\leq \sum_{n_{\underline{t}}=m}^{\infty} \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\left\{-N_a(t)\varepsilon\right\}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)\right] \\
&\leq \sum_{n_{\underline{t}}=m}^{\infty} \exp\left\{-2(1-\exp\left\{-\varepsilon\right\}) \sum_{s \in \mathcal{T}_t(a)} n_s p_s\right\}
\end{aligned}$$

where the first inequality is due to a union bound, the second inequality is due to Markov's inequality, and the third inequality is due to Lemma E.3.

We now turn our attention to (41).(2). Notice how by definition of $\boldsymbol{\lambda}^{\star}_s(a,\mathbf{h})$ and $\boldsymbol{\lambda}^{(\mathsf{Q})}_s(a,\mathbf{h})$ we have the following upper bound for each $s \in \mathcal{T}_t(a)$,

$$\sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a,\mathbf{h}) \log\left(\boldsymbol{\lambda}^{\star}_s(a,\mathbf{h})^{\top}\mathbf{E}_{i,s}\right) \geq \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a,\mathbf{h}) \log\left(\boldsymbol{\lambda}^{(\mathsf{Q})}_s(a,\mathbf{h})^{\top}\mathbf{E}_{i,s}\right).$$

From the portfolio regret of the Universal Portfolio algorithm (Cover & Ordentlich, 1996) we have that for each $s \in \mathcal{T}_t(a)$,

$$\sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a,\mathbf{h}) \log\left(\boldsymbol{\lambda}^{\star}_s(a,\mathbf{h})^{\top}\mathbf{E}_{i,s}\right) - \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \sum_{i \in \mathcal{I}_s} \xi_{i,s}(a,\mathbf{h}) \log\left(\boldsymbol{\lambda}_{i,s}(a,\mathbf{h})^{\top}\mathbf{E}_{i,s}\right) \leq \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} R_{\check{N}_{a,\mathbf{h}}(s)}.$$

Putting the inequalities from the above two displays together allows us to upper bound (41).(2) as

$$(41).(2) \leq \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \sum_{s \in \mathcal{T}_t(a)} \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} R_{\check{N}_{a,\mathbf{h}}(a)} \geq \varepsilon\right).$$

Hence, all of the above steps allow us to conclude that

$$\begin{aligned}
&\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \left|\log\left(W_{N_a(t)}\right) - \log\left(W^{(\mathsf{Q})}_{N_a(t)}\right)\right| \geq \varepsilon\right) \\
&\quad \leq \sum_{n_{\underline{t}}=m}^{\infty} \exp\left\{-2(1-\exp\left\{-\varepsilon\right\}) \sum_{s \in \mathcal{T}_t(a)} n_s p_s\right\} + \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \frac{1}{N_a(t)} \sum_{s \in \mathcal{T}_t(a)} \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} R_{\check{N}_{a,\mathbf{h}}(s)} \geq \varepsilon\right),
\end{aligned}$$

completing the proof. □

E.3 Path-wise control of portfolio regret

Lemma E.1. *Fix a distribution $\mathsf{P} \in \mathcal{P} \cup \mathcal{Q}$ and for any $n \in \mathbb{N}$ let $R_n \coloneqq \log(n+1)/2 + \log(2)$ denote the portfolio regret bound from the Universal Portfolio algorithm (Cover & Ordentlich, 1996).*

We then have that for any $\varepsilon > 0$,

$$\lim_{m\to\infty} \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{N_a(t)}\geq\varepsilon\right)=0\,.$$

Proof. Fix a distribution $\mathsf{P}\in\mathcal{P}\cup\mathcal{Q}$. We start by upper bounding the normalized experiment-wide regret as

$$\sum_{s\in\mathcal{T}_t(a)}\frac{1}{N_a(t)}R_{\check{N}_a(s)}=\sum_{s\in\mathcal{T}_t(a)}\frac{\check{N}_a(s)}{N_a(t)}\frac{R_{\check{N}_a(s)}}{\check{N}_a(s)}\leq\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{\check{N}_a(s)}\,,$$

where the inequality follows from the definition of $N_a(t)=\sum_{s\in\mathcal{T}_t(a)}\check{N}_a(s)$ which implies that for any $s\in\mathcal{T}_t(a)$, $\check{N}_a(s)/N_a(t)\leq 1$.

From assumption we have that for every $s\in[t]$, $n_s\to\infty$ as $n_{\underline{t}}\to\infty$. When combined with Lemma E.4, which tells us that $\check{N}_a(s)\to\infty$ $\mathsf{P}_{\mathrm{RCT}}$-almost surely as $n_s\to\infty$, we can conclude that

$$\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{N_a(t)}\leq\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{\check{N}_a(s)}\to 0\quad\mathsf{P}_{\mathrm{RCT}}\text{-almost surely}$$

because $R_n=\log(n+1)/2+\log(2)$ for every $n\in\mathbb{N}$. Since the convergence holds pathwise we can conclude that as $m\to\infty$,

$$\sup_{n_{\underline{t}}\geq m}\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{N_a(t)}\to 0\quad\mathsf{P}_{\mathrm{RCT}}\text{-almost surely}\,,$$

which in turn implies that for any $\varepsilon>0$ as $m\to\infty$,

$$\mathbb{1}\left\{\sup_{n_{\underline{t}}\geq m}\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{N_a(t)}\geq\varepsilon\right\}\to 0\quad\mathsf{P}_{\mathrm{RCT}}\text{-almost surely}\,.$$

Putting all of the above steps together we have that as $m\to\infty$,

$$\mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{N_a(t)}\geq\varepsilon\right)=\mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\left\{\sup_{n_{\underline{t}}\geq m}\sum_{s\in\mathcal{T}_t(a)}\frac{R_{\check{N}_a(s)}}{N_a(t)}\geq\varepsilon\right\}\right]\to 0\,,$$

completing the proof. □

Lemma E.2. *Fix a distribution $\mathsf{P}\in\mathcal{P}\cup\mathcal{Q}$ and for any $n\in\mathbb{N}$ let $R_n\coloneqq\log(n+1)/2+\log(2)$ denote the portfolio regret bound from the Universal Portfolio algorithm (Cover & Ordentlich, 1996). We then have that for any $\varepsilon>0$,*

$$\lim_{m\to\infty}\mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}}\geq m}\sum_{s\in\mathcal{T}_t(a)}\sum_{\mathbf{h}\in\mathcal{H}_{s-1}}\frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{N_a(t)}\geq\varepsilon\right)=0\,.$$

Proof. Fix a distribution $\mathsf{P}\in\mathcal{P}\cup\mathcal{Q}$. We start by upper bounding the normalized experiment-wide regret as

$$\sum_{s\in\mathcal{T}_t(a)}\sum_{\mathbf{h}\in\mathcal{H}_{s-1}}\frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{N_a(t)}=\sum_{s\in\mathcal{T}_t(a)}\sum_{\mathbf{h}\in\mathcal{H}_{s-1}}\frac{\check{N}_a(s)}{N_a(t)}\frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{\check{N}_a(s)}\leq\sum_{s\in\mathcal{T}_t(a)}\sum_{\mathbf{h}\in\mathcal{H}_{s-1}}\frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{\check{N}_a(s)}\,,$$

where the inequality follows from the definition of $N_a(t)=\sum_{s\in\mathcal{T}_t(a)}\check{N}_a(s)$ which implies that for any $s\in\mathcal{T}_t(a)$, $\check{N}_a(s)/N_a(t)\leq 1$. Recall that $R_n=\log(n+1)/2+\log(2)$ for every $n\in\mathbb{N}$. Since R_n is nondecreasing we have that for any $\mathbf{h}\in\mathcal{H}_{s-1}$, $R_{\check{N}_{a,\mathbf{h}}(s)}\leq R_{\check{N}_a(s)}$. This in turn implies that

$$\sum_{s\in\mathcal{T}_t(a)}\sum_{\mathbf{h}\in\mathcal{H}_{s-1}}\frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{\check{N}_a(s)}\leq\sum_{s\in\mathcal{T}_t(a)}|\mathcal{H}_{s-1}|\,\frac{R_{\check{N}_a(s)}}{\check{N}_a(s)}\,.$$

We note how $|\mathcal{H}_{s-1}| < \infty$ (i.e., is finite) which is a consequence of Assumption B.1 and our assumption that the carryover effects last for ℓ sub-experiments. From assumption we have that for every $s \in [t]$, $n_s \to \infty$ as $n_{\underline{t}} \to \infty$. Hence, from the finiteness of $|\mathcal{H}_s|$, the sublinearity of R_n, and Lemma E.4, which tells us that $\check{N}_a(s) \to \infty$ $\mathsf{P}_{\mathrm{RCT}}$-almost surely as $n_s \to \infty$, we can conclude that as $n_{\underline{t}} \to \infty$,

$$\sum_{s \in \mathcal{T}_s(a)} |\mathcal{H}_{s-1}| \frac{R_{\check{N}_a(s)}}{\check{N}_a(s)} \to 0 \quad \mathsf{P}_{\mathrm{RCT}}\text{-almost surely}\,.$$

Since the convergence holds pathwise we can conclude that as $m \to \infty$,

$$\sup_{n_{\underline{t}} \geq m} \sum_{s \in \mathcal{T}_t(a)} \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{N_a(t)} \to 0 \quad \mathsf{P}_{\mathrm{RCT}}\text{-almost surely}\,,$$

which in turn implies that for every $\varepsilon > 0$ as $m \to \infty$,

$$\mathbb{1}\left\{ \sup_{n_{\underline{t}} \geq m} \sum_{s \in \mathcal{T}_t(a)} \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{N_a(t)} \geq \varepsilon \right\} \to 0 \quad \mathsf{P}_{\mathrm{RCT}}\text{-almost surely}\,.$$

Putting all of the above steps together we have that as $m \to \infty$,

$$\mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\sup_{n_{\underline{t}} \geq m} \sum_{s \in \mathcal{T}_t(a)} \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{N_a(t)} \geq \varepsilon \right) = \mathbb{E}_{\mathsf{P}_{\mathrm{RCT}}}\left[\mathbb{1}\left\{ \sup_{n_{\underline{t}} \geq m} \sum_{s \in \mathcal{T}_t(a)} \sum_{\mathbf{h} \in \mathcal{H}_{s-1}} \frac{R_{\check{N}_{a,\mathbf{h}}(s)}}{N_a(t)} \geq \varepsilon \right\} \right] \to 0\,,$$

completing the proof. □

E.4 Upper bound on the expected exponential decay as a function of the number of units in a treatment arm.

In this subsection we present an upper bound on the expected exponential decay as a function of the number of units assigned to a treatment arm when multiplied by the sub-optimality ratio of wealth processes. We prove the result for the setting in which units are classified into different strata in each sub-experiment, as the same result in which units are not stratified in each sub-experiment follows as a consequence.

Lemma E.3. *Fix the total number of sub-experiments $t \in \mathbb{N}$, an arm $a \in \mathcal{A}$ and a distribution $\mathsf{Q} \in \mathcal{Q}(a)$. Let $W^{(\mathsf{Q})}_{N_a(t)}$ be the process that has oracle access to the numeraire portfolios in every sub-experiment, as defined in* (20)*. We then have that*

$$\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\{-N_a(t)\varepsilon\} \left(W_{N_a(t)} / W^{(\mathsf{Q})}_{N_a(t)} \right) \right] \leq \exp\left\{ -2\left(1 - \exp\{-\varepsilon\}\right) \sum_{s \in \mathcal{T}_t(a)} n_s p_s \right\}.$$

Proof. Fix the total number of sub-experiments $t \in \mathbb{N}$, an arm $a \in \mathcal{A}$, and a distribution $\mathsf{Q} \in \mathcal{Q}(a)$. Let $n^\star$ be the index of the last unit to enter the experiment and $n^\star - 1$ be the index of the second to last unit to enter the experiment. The law of iterated expectation gives us the following identity on the expectation

$$\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\{-N_a(t)\varepsilon\} \left(W_{N_a(t)} / W^{(\mathsf{Q})}_{N_a(t)} \right) \right] = \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\{-N_a(t)\varepsilon\} \left(W_{N_a(t)} / W^{(\mathsf{Q})}_{N_a(t)} \right) \,\middle|\, \mathcal{F}_{n^\star - 1, t} \right] \right].$$

For now, we will focus on the conditional expectation, which simplifies to

$$\begin{aligned} &\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\{-N_a(t)\varepsilon\} \left(W_{N_a(t)} / W^{(\mathsf{Q})}_{N_a(t)} \right) \,\middle|\, \mathcal{F}_{n^\star - 1, t} \right] \\ &= \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\left[\exp\{-\varepsilon \check{N}_a(t)\} \left(W_{\check{N}_a(t)} / W^{(\mathsf{Q})}_{\check{N}_a(t)} \right) \right]^{\mathbb{1}\{a \in \mathcal{A}_t\}} \,\middle|\, \mathcal{F}_{n^\star - 1, t} \right] \\ &\quad \cdot \left(W_{N_a(t-1)} / W^{(\mathsf{Q})}_{N_a(t-1)} \right) \exp\{-N_a(t-1)\varepsilon\}\,. \end{aligned} \tag{42}$$

Further focusing on the conditional expectation from (42), we can see that if $a \notin \mathcal{A}_s$ then (42) $= 1$. Otherwise, if $a \in \mathcal{A}_s$, the conditional expectation from (42) decomposes into

$$(42) = \mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\prod_{\mathbf{h}\in\mathcal{H}_{t-1}} \exp\left\{-\varepsilon\xi_{n^\star,t}(a,\mathbf{h})\right\}\left(\frac{\boldsymbol{\lambda}_{n^\star,t}^\top \mathbf{E}_{n^\star,t}(a)}{\boldsymbol{\lambda}_t^{(\mathsf{Q})}(a,\mathbf{h})^\top \mathbf{E}_{n^\star,t}(a)}\right)^{\xi_{n^\star,t}(a,\mathbf{h})} \middle| \mathcal{F}_{n^\star-1,t}\right] \tag{43}$$

$$\cdot \prod_{i\in\mathcal{I}_t(n^\star-1)}\prod_{\mathbf{h}\in\mathcal{H}_{t-1}}\left(\frac{\boldsymbol{\lambda}_{i,t}^\top \mathbf{E}_{i,t}(a)}{(\boldsymbol{\lambda}_t^{(\mathsf{Q})}(a,\mathbf{h}))^\top \mathbf{E}_{i,t}(a)}\right)^{\xi_{i,t}(a,\mathbf{h})} \prod_{i\in\mathcal{I}_t(n^\star-1)}\prod_{\mathbf{h}\in\mathcal{H}_{t-1}} \exp\left\{-\varepsilon\xi_{i,t}(a,\mathbf{h})\right\} .$$

Focusing on the conditional expectation from (43) we have that conditioning on the filtration $\mathcal{F}_{n^\star-1}$ only the inclusion indicator, $\xi_{n^\star,t}(a,\mathbf{h})$, corresponding to $\mathbf{H}_{n^\star,t} = \mathbf{h}$ will not be equal to zero—as the filtration includes the history of unit $n^\star$, which determines the strata the unit belongs to. Nonetheless, the event $\{A_{n^\star,t} = a \text{ or } A_{n^\star,t} = 0\}$ is still random as the treatment assignments are randomly assigned in each sub-experiment. Hence, focusing on the inclusion indicator for which $\mathbf{H}_{n^\star,t} = \mathbf{h}$ and applying the law of total expectation we have that

$$\begin{aligned}
&\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\left\{-\varepsilon\xi_{n^\star,t}(a,\mathbf{h})\right\}\left(\frac{\boldsymbol{\lambda}_{n^\star,t}^\top \mathbf{E}_{n^\star,t}(a)}{\boldsymbol{\lambda}_t^{(\mathsf{Q})}(a,\mathbf{h})^\top \mathbf{E}_{n^\star,t}(a)}\right)^{\xi_{n^\star,t}(a,\mathbf{h})} \middle| \mathcal{F}_{n^\star-1,t}\right] \\
&= \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 0 \mid \mathcal{F}_{n^\star-1,t}\right) \\
&\quad + \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\exp\left\{-\varepsilon\right\}\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\frac{\boldsymbol{\lambda}_{n^\star,t}^\top \mathbf{E}_{n^\star,t}(a)}{\boldsymbol{\lambda}_t^{(\mathsf{Q})}(a,\mathbf{h})^\top \mathbf{E}_{n^\star,t}(a)} \middle| \mathcal{F}_{n^\star-1,t}, \xi_{n^\star,t}(a,\mathbf{h}) = 1\right] \\
&\leq \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 0 \mid \mathcal{F}_{n^\star-1,t}\right) + \mathbb{P}_{\mathsf{P}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 1\right)\exp\left\{-\varepsilon\right\} \\
&= 1 - \mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\left(1-\exp\left\{-\varepsilon\right\}\right) \\
&\leq \exp\left\{-\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\left(1-\exp\left\{-\varepsilon\right\}\right)\right\} ,
\end{aligned}$$

where the first inequality follows from the numeraire property of $\boldsymbol{\lambda}_t^{(\mathsf{Q})}(a,\mathbf{h})$ (Cover & Thomas, 1999, Theorem 15.2.2) (see also (Waudby-Smith et al., 2025; Sandoval et al., 2026)) and the second inequality follows from $1 + x \leq \exp\{x\}$ for all $x \in \mathbb{R}$.

Notice how $\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a,\mathbf{h}) = 1 \mid \mathcal{F}_{n^\star-1,t}\right) = \pi_t(a) + \pi_t(0)$, which follows from the fact that conditioning on $\mathcal{F}_{n^\star,t}$ the realization of $\mathbf{H}_{n^\star,t}$ is known and fixed. Thus, from Assumption 3.2 we have that $\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a) = 1 \mid \mathcal{F}_{n^\star-1,t}\right) \geq 2p_t$, which leads to the following upper bound

$$\exp\left\{-\mathbb{P}_{\mathsf{Q}_{\mathrm{RCT}}}\left(\xi_{n^\star,t}(a) = 1 \mid \mathcal{F}_{n^\star-1,t}\right)\left(1-\exp\left\{-\varepsilon\right\}\right)\right\} \leq \exp\left\{-2p_t(1-\exp\left\{-\varepsilon\right\})\right\} .$$

Recursively applying the above arguments we get the following inequality

$$\mathbb{E}_{\mathsf{Q}_{\mathrm{RCT}}}\left[\exp\left\{-N_a(t)\varepsilon\right\}\left(W_{N_a(t)}/W^{(\mathsf{Q})}_{N_a(t)}\right)\right] \leq \exp\left\{-2\left(1-\exp\left\{-\varepsilon\right\}\right)\sum_{s\in\mathcal{T}_t(a)} n_s p_s\right\} ,$$

proving Lemma E.3. We end the proof by noting that the same result follows for the setting in which units are not stratified in each sub-experiment. This is because in this setting the process no longer has the product over strata in each sub-experiment, which as we argued above has no effect on the conditional probability of the inclusion indicators and, thus, the effective sample size per treatment arm. □

E.5 Almost sure infinite per-arm allocation

Lemma E.4. *Fix a sub-experiment $s \in [t]$ and a treatment $a \in \mathcal{A}_s$. Recall the definitions $n_s \coloneqq |\mathcal{I}_s|$ and $\check{N}_a(s) \coloneqq \sum_{i\in\mathcal{I}_s} \mathbb{1}\{A_{i,s} = a \text{ or } A_{i,s} = 0\}$, where $A_{i,s} \sim \mathrm{Multinomial}_{\mathcal{A}_s\cup\{0\}}(\boldsymbol{\pi}_s)$. Under Assumption 3.2 we have that*

$$\check{N}_a(s) \to \infty \quad \text{as} \quad n_s \to \infty .$$

Proof. We define the shorthand $\mathrm{RCT}_s \coloneqq \mathrm{Multinomial}_{\mathcal{A}_s\cup\{0\}}(\boldsymbol{\pi}_s)$. By the strong law of large numbers we have that

$$\lim_{n_s\to\infty} \frac{\check{N}_a(s)}{n_s} = \mathbb{P}_{\mathrm{RCT}_s}\left(A_{i,s} = a \text{ or } A_{i,s} = 0\right) \quad \mathrm{RCT}_s\text{-almost surely} .$$

Indeed, notice how the probability can be written as:

$$\begin{aligned}\mathbb{P}_{\mathrm{RCT}_s}\left(\mathbb{1}\left\{A_{i,s}=a \text{ or } A_{i,s}=0\right\}=1\right) &= \mathbb{P}_{\mathrm{RCT}_s}\left(A_{i,s}=a \text{ or } A_{i,s}=0\right)\\ &= \mathbb{P}_{\mathrm{RCT}_s}\left(A_{i,s}=a\right)+\mathbb{P}_{\mathrm{RCT}_s}\left(A_{i,s}=0\right)>0\,,\end{aligned}$$

where the second equality follows from the fact that the events $\{A_{i,s}=a\}$ and $\{A_{i,s}=0\}$ are mutually exclusive, and the inequality is due to Assumption 3.2 which tells us that $\mathbb{P}_{\mathrm{RCT}_s}\left(A_{i,s}=a\right)>0$ and $\mathbb{P}_{\mathrm{RCT}_s}\left(A_{i,s}=0\right)>0$. Hence, we can conclude that $\check{N}_a(s)\to\infty$ almost surely as $n_s\to\infty$, completing the proof. □

Lemma E.5. *Fix a treatment $a\in\mathcal{A}$ and recall that $\underline{t}\coloneqq\arg\min_{s\in[t]}n_s$. Then, under Assumption 3.2 and the data collection protocol from Algorithm 2 we have that*

$$N_a(t)\to\infty \quad \text{as} \quad n_{\underline{t}}\to\infty\,.$$

Proof of Lemma E.5. Fix a treatment $a\in\mathcal{A}$. We start by recalling the definition $\underline{t}\coloneqq\arg\min_{s\in[t]}n_s$, from which we can see how

$$n_{\underline{t}}\to\infty \quad \text{implies that} \quad n_s\to\infty\,,\ \forall s\in[t]\,.$$

Hence, together with the definition $N_a(t)\coloneqq\sum_{s\in\mathcal{T}_t(a)}\check{N}_a(s)$ and Lemma E.4 it implies that

$$N_a(t)\to\infty \quad \text{as} \quad n_{\underline{t}}\to\infty\,,$$

completing the proof. □