The Minimax-Regret Decision Framework for Online A/B Tests*

Khai X. Chiong† Joonhwi Joo‡

January 2023

Abstract

In online A/B tests, where the sample flows in over time from a very large population and multiple policies can be experimented concurrently, an analyst has to make decisions on the efficient traffic allocation, when to stop the experiment, and which of the policies to adopt. We develop the minimax-regret decision framework for online A/B tests, providing integrated optimal solutions for the aforementioned decision problems. Our minimax-regret decision framework controls for the maximum regret, which is the maximum anticipated net payoff loss from making an error in a decision, instead of the maximum Type I error probabilities. Notably, our framework rationalizes and advocates a much less conservative cutoff decision rule in favor of new, innovative policies than the conventional hypothesis-testing cutoffs. We apply our framework to a large mobile game company's multi-arm experiment data, in which our minimax-regret decision framework arrives at drastically different decisions than the conventional hypothesis-testing framework. We then validate our framework by running a series of Monte Carlo simulations that mimic the data-generating process of our focal company's experiments. Our minimax-regret decision criteria attain better performance in designating the correct decisions as well as achieving lower net payoff loss than the hypothesis-testing counterpart. Without sacrificing the accuracy of the decisions, our minimax-regret efficient traffic allocation reduces the wait time and sample size by more than 30% relative to the ad-hoc traffic allocation.

Keywords: A/B tests, minimax regret, decision theory

*We thank Ron Berman, Giovanni Compiani, Max Joo, Kyeongbae Kim, TI Kim, Samir Mamadehussene, Ram Rao, Andrey Simonov, Kosuke Uetake, seminar participants at LSE, KAIST Business School, Tilburg University, and WashU Olin for helpful comments. The authors' names are in alphabetical order and each contributed equally.

†Naveen Jindal School of Management, The University of Texas at Dallas. `khai.chiong@utdallas.edu`

‡Naveen Jindal School of Management, The University of Texas at Dallas. `joonhwi.joo@utdallas.edu`

1 Introduction

Firms constantly run online randomized field experiments in digital marketing, referred to as online A/B tests, to find a new marketing policy that improves the payoff of interest, e.g., revenue, retention, and conversion. For instance, search engines test hundreds of different website layout features, such as sponsored ad location, font size, and background color. Video hosting platforms vary the users' ad-exposure frequency and duration by seconds. Mobile game studios examine new game features such as difficulty level, character design, and item prices, by making only very small changes to them. In these examples, the per-user effect size is presumably small because the new marketing policy experiments being run involve only a piecemeal change from the status quo, and the underlying population is very large compared to the test sample. The exploration-exploitation tradeoff for the payoffs in the A/B test sample can be ignored in such a setup.

The Neyman-Pearson hypothesis test has been the standard decision framework applied universally to online A/B tests regardless of the context and goal of running the A/B tests. What the hypothesis test controls for is only the maximum-tolerable Type I error probability under the desired significance level; it does not control for Type II error probability or the magnitude of foregone payoffs from committing an error. However, using the Type I error probability and the associated statistical significance as the policy-sorting and/or policy-adoption-decision criteria can be at odds with improving the payoff, because a policy with a larger p-value (less significant) may have a higher expected payoff. Another problem with hypothesis tests is the absence of a sensible adaptive sample-collection stopping criterion that does not peek at the effect-size estimate or its p-value midway through the sample collection. It has led to a widespread practice of a p-hacking scheme referred to as "peeking and optional stopping," whereby the analyst monitors the p-value during the sample collection and terminates the experiment if the p-value falls below the conventional significance level (e.g., 0.05 or 0.01). Such a p-hacking scheme inflates the actual Type I error probability more than the intended maximum-tolerable Type I error probability, invalidating the decisions from the hypothesis tests.[1][2]

In this paper, we propose an alternative decision-theoretic framework for online A/B tests, which provides an integrated solution for the aforementioned problems. The idea is to account for both the magnitude of the foregone payoff and the probability of committing either Type I or Type II errors in decision-making, rather than accounting for the Type I error probability alone. Specifically, we propose controlling for the maximum regret, which is the worst-case {error probability $\times$ net

[1] Suppose the analyst is allowed to peek at the experiment at least once during the collection of a size-n sample and stop the sample collection if the "naive" p-value is smaller than the pre-set significance level α. Then, the event of rejecting the null occurs only more often than no peeking is allowed.

[2] See, e.g., Armitage et al. (1969); McPherson and Armitage (1971); Deng et al. (2016); Miller and Hosanagar (2020); Johari et al. (2017, Forthcoming), and also (all accessed in December 2022):
https://blog.analytics-toolkit.com/2017/bayesian-ab-testing-not-immune-to-optional-stopping-issues/
https://blog.analytics-toolkit.com/2017/the-bane-of-ab-testing-reaching-statistical-significance/
https://blog.acolyer.org/2017/09/28/peeking-at-ab-tests-continuous-monitoring-without-pain/
https://www.evanmiller.org/sequential-ab-testing.html

foregone payoff} by committing either type of an error. Controlling for the maximum regret is payoff optimal by construction. It allows for a direct comparison across multiple experiment arms in choosing the best arm, because the unit of the maximum regret is identical to the payoff of interest (e.g., USD for revenue, no. of users for retention). The maximum regret is decreasing in the sample size, and therefore, it is a natural object to be optimized in deciding the overall sample size and the sample allocation ratio across multiple experiment arms. We devise a metric to be monitored during sample collection based on the maximum regret, which can be calculated without peeking at the current effect-size estimates and is therefore not subject to the p-hacking scheme mentioned above. Finally, we show in our empirical application and Monte Carlo validation that our proposed decision framework achieves a higher payoff and reduces the sample size substantially without sacrificing the accuracy of a decision, compared with the conventional hypothesis-testing framework.

We build our decision framework upon the prior work of Tetenov (2012), who developed the baseline minimax-regret decision rule for two-arm experiments when the experiment data are exogenously given to the analyst. We innovate upon Tetenov's baseline minimax-regret decision rule and make several novel methodological contributions by developing an integrated minimax-regret decision framework for (i) multi-arm experiments, (ii) the experiment-stopping rule, and (iii) the efficient dynamic traffic allocation, where (i) and (ii) pertain to the optimal experiment *evaluation* problem and (iii) to the optimal experiment *design* problem. In a (i) multi-arm experiment, the analyst needs to decide not only whether to reject the status quo policy (i.e., the control), but also which new alternative policy to adopt if multiple new policies (i.e., treatments) are preferred to the status quo policy. We propose filtering and sorting the alternative new policies by the maximum regret of rejecting the status quo policy. The filtering and sorting criteria are directly relevant to the payoffs and give a unique and consistent ordering across all the policies being examined. For the (ii) experiment-stopping rule, we propose monitoring and controlling for the worst-case minimax regret at a level that an analyst decides, which does not involve monitoring the effect-size estimate itself or the associated p-values. Turning to the optimal experiment *design*, (iii) the efficient dynamic traffic-allocation problem can be viewed as a combination of (i) and (ii). We provide the dynamic optimal traffic-allocation scheme that minimizes the overall sample size, which translates directly to the wait time before making a policy-adoption decision.

Notably, when the regret is symmetric, the optimal decision cutoff threshold in our minimax-regret framework turns out to be 0, which is much less conservative than 1.645, 1.960, or 2.326 (corresponding to the significance levels of 0.05, 0.025, or 0.01, respectively) in traditional z-tests. We also show our zero decision cutoff threshold derived from the mini*max*-regret criterion can also be derived from the problem of minimizing *average* regret, as long as the analyst does not have any prior information about the true distribution of the effect size, that is, when the prior is symmetric around zero.

We apply our minimax-regret decision framework to real-world field-experiment data. The data come from the multi-arm experiment conducted by a large mobile game studio that ran a series of randomized experiments to examine the effect of game-difficulty adjustment on customers' in-game currency spending. According to the conventional hypothesis-testing decision rule, the t-stats for all three new proposed policies do not exceed the threshold needed to reject the status quo policy. By contrast, all three new policies exceeded the optimal minimax-regret decision threshold. The analysis of the field-experiment data highlights that the payoff-optimal decision rules may diverge drastically from hypothesis tests in practice.

To validate our minimax-regret decision framework, we run a series of Monte Carlo experiments that closely replicate our mobile game company's experiments. In the experiment-*evaluation* Monte Carlo, we examine and compare the performance of the proposed minimax-regret decision framework against the hypothesis-testing decision rule. Because the true parameter values are known in the simulated data, we can calculate the percentage of the incorrect decisions and the associated realized payoff loss. We find our minimax-regret decision framework performs remarkably well in designating the correct decision, and our focal company could potentially gain several thousands of dollars per day by switching to the proposed minimax-regret decision rule from the conventional hypothesis-testing decision rule. In the experiment-*design* Monte Carlo, we examine the performances of the proposed efficient traffic allocation. Compared with the benchmark in which the company allocates the traffic at an ad-hoc ratio, our method can reduce the required sample size by more than 30% without sacrificing the accuracy of the decisions. In the context of the field experiment we analyzed, it translates into a reduction in the wait time for data collection from 21 days to 14 days to reach the same decision. Adopting our minimax decision framework would help the company make timely decisions with minimal wait time spent on sample collection.

Our work contributes to an emerging literature on developing new statistical decision frameworks based on a more relevant metric than Type I error probabilities in marketing and management science (e.g., Chick and Inoue, 2001; Chick and Gans, 2009; Chick and Frazier, 2012; Ascarza, 2018; Feit and Berman, 2019; Schwartz et al., 2017; Berman and Van den Bulte, Forthcoming). Our minimax-regret decision rule formalizes the notion of Ascarza (2018)'s "value lift" in the context of online A/B tests and provides a systematic guideline to control for the maximum foregone net value lift. The most closely related work to ours is Feit and Berman (2019). Based on the Bayesian perspective with a fully specified parametric shape of the priors, Feit and Berman study the optimal ex-ante profit-maximizing decision criteria and sample-size selection problem in two-arm A/B tests when the population is finite and relatively small, for which the costs of A/B tests and gains/losses from the test sample cannot be ignored. Even though the approach and decision context that Feit and Berman consider are different from ours, in that we approach the problem from a frequentist's perspective and we study the "pure-exploration bandits" problem whereby the rewards earned during the exploration period can be ignored due to a very large underlying population size and/or relatively small per-user effect size, the zero optimal cutoff they derive is the same as ours in the case of two-arm experiments. This paper therefore complements Feit and Berman and reconfirms that a less conservative decision rule than what is implied by hypothesis testing is payoff optimal across different experiment settings.

In addition, our minimax-regret decision framework is amenable to optimizing the site traffic and choosing the best policy in multi-arm experiments.

This paper contributes to the literature that develops the statistical decision rules based on the notion of *maximum regret* (e.g., Manski, 2000, 2004; Dehejia, 2005; Hirano and Porter, 2009; Tetenov, 2012; Hirano and Porter, 2020), and the asymmetric minimax-regret framework developed by Hirano and Porter (2009); Stoye (2009); Tetenov (2012). We add to this literature by innovating upon Tetenov (2012)'s baseline asymmetric minimax-regret framework to several dimensions, as explained above. Furthermore, we are the first to implement the asymmetric minimax-regret framework to real-world online A/B test data.

Our work contributes to the literature that studies the stopping rule in online A/B test sample collection. At least since Armitage et al. (1969); McPherson and Armitage (1971), testing the hypothesis of null effects repeatedly during the sample collection is known to make the usual p-value invalid in controlling for the Type I error probability. Recent attempts have sought to provide valid "optional stopping" criteria in terms of correct Type I error probabilities when the A/B tests are continuously monitored (e.g., Deng et al., 2016; Johari et al., 2017, Forthcoming). Our work provides alternative stopping criteria that allow for the real-time monitoring of the worst-case maximum regret, but not the effect-size estimator itself, and therefore are not subject to the aforementioned p-hacking schemes.

We also contribute to the literature on efficient sample allocation in multi-armed bandit experiments. Lai and Robbins (1985); Agrawal (1995) derived the average-regret-based efficient adaptive sample allocation rules. Other closely related research in this vein is Schwartz et al. (2017), who propose a revenue- or acquisition-maximizing dynamic sample allocation problem in multi-armed bandit experiments from the Bayesian perspective. We add to this literature by proposing the criteria that control for the maximum regret from a frequentist's perspective.

The remainder of the paper is organized as follows. Section 2 develops the experiment-evaluation part of the minimax-regret decision framework for online A/B tests, taking the experiment design as given. Section 3 develops the minimax-regret-based experiment-design part, where we propose the efficient traffic allocation algorithm. Section 4 presents an empirical application of our minimax-regret decision framework, and Section 5 presents the results from our Monte Carlo studies. Section 6 concludes.

2 The Minimax-Regret Decision Rule for Evaluating Online A/B Tests

We assume the goal of the company is to find a new policy that improves the payoff of interest from the status quo, and the gains/losses from the test sample can be ignored. The payoff of interest may include but is not limited to revenue, retention, conversion, or other performance indicators that may affect the company's profit over time, and we abstract away from its relationship with the short- or long-run profits.

In this section, we focus on the experiment-*evaluation* problem of online A/B tests, taking the experiment *design* as given. Subsection 2.1 primarily follows Tetenov (2012), setting up the minimax-regret decision problem for two-arm experiments. Our major contributions then follow – we derive the rules for the sample-collection stopping criteria and the policy-adoption decision rules for multi-arm experiments in Subsections 2.2 and 2.3.

2.1 The Minimax-Regret Decision Criteria for Two-Arm Experiments

We propose a cutoff decision rule based on the notion of *maximum regret*, which accounts not only for the maximum probability of committing an error but also for the maximum *net foregone payoff*. In the following, we explain how the optimal cutoff threshold can be determined according to the decision rule that minimizes (over the action) the maximum (over the true parameter value) regrets.

Let θ be the unknown true effect size of the new, innovative policy being proposed. θ can be either positive or negative. Without losing generality, we assume higher θ is more desirable for the firm, and the firm wants to test the null $H_0 : \theta \leq 0$ against the alternative $H_1 : \theta > 0$. We focus on the two-arm policy-adoption decision problem in the current subsection, by following the classical decision-theoretic setup in which the analyst has to decide whether to adopt the new innovative policy based on the evidence from a size-n sample.

A maintained assumption in this paper is that a consistent and (asymptotically) Normally distributed estimator $\hat{\theta}$ is available to the analyst, formalized as follows:

Assumption 1. *A consistent and (asymptotically) Normally distributed estimator $\hat{\theta}$ for θ is available from an experimental variation created by the analyst; that is, $\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2\right)$, where σ^2 is the variance of the estimator $\hat{\theta}$ that would generally depend on the sample size n.*

This is a mild assumption because it applies to virtually all effect-size estimators used in A/B testing. Examples include the usual mean-differences estimator, regression estimator, covariate and propensity-score matching, difference-in-differences estimator with covariates, and machine-learning-based estimators proposed in, for example, Athey and Wager (2021); Farrell et al. (2021).

Our focus is on the class of cutoff decision rules with the following form:

$$\begin{cases} \text{Reject the status-quo policy (accept the new policy)} & \text{if } \hat{\theta} > T^* \\ \text{Accept the status-quo policy (reject the new policy)} & \text{if } \hat{\theta} \leq T^* \end{cases} \tag{2.1}$$

for some cutoff threshold $T^* \in \mathbb{R}$.[3] Our goal is to find the optimal threshold T^* from the problem of selecting the policy that minimizes the *maximum regret* we define below.

[3]It can be shown that when the distribution of $\hat{\theta}$ exhibits the monotone likelihood ratio property and the associated loss function of the underlying decision problem is "reasonable," which is the case in our setup, the class of the cutoff decision rules is *essentially complete* in the sense that the decision rule is not dominated by any other set of decision

Following Hirano and Porter (2009); Tetenov (2012), we define the Type I and Type II *regret* evaluated at any candidate threshold $T \in \mathbb{R}$, as a function of (T, θ, σ), as

$$\begin{aligned}
R_{\text{Type I}}(T;\theta,\sigma) &= |\theta| \Pr(\text{Reject the status quo policy given } T | \theta \leq 0) & \\
&= -\theta P_{\theta,\sigma}\left(\hat{\theta} > T\right) & \text{if } \theta \leq 0 \\
&= -\theta \Phi\left(\tfrac{\theta - T}{\sigma}\right) & \\
R_{\text{Type II}}(T;\theta,\sigma) &= |\theta| \Pr(\text{Accept the status quo policy given } T | \theta > 0) & \\
&= \theta P_{\theta,\sigma}\left(\hat{\theta} \leq T\right) & \text{if } \theta > 0, \\
&= \theta \Phi\left(\tfrac{T - \theta}{\sigma}\right) &
\end{aligned} \tag{2.2}$$

where we invoked (2.1) in the respective second equalities and $\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2\right)$ in the respective third equalities.

Next, we define the asymmetry factors $K_1 > 0$ and $K_2 > 0$ that reflect the decision-maker's relative weight for the status quo and new policies, respectively. A higher value of K_1 (K_2) will give more weight to Type I (Type II) regret. Without losing generality, we fix $K_2 = 1$ henceforth unless noted otherwise.[4] Applying the weights, we can write the regret function as

$$R(T;\theta,\sigma) = K_1 \mathbf{1}(\theta \leq 0) R_{\text{Type I}}(T;\theta,\sigma) + \mathbf{1}(\theta > 0) R_{\text{Type II}}(T;\theta,\sigma). \tag{2.3}$$

[5] The regret $R(T;\theta,\sigma)$, which accounts for both the magnitude of the true payoff parameter θ and the probability that the estimator is realized at $\hat{\theta}$, can be interpreted as the absolute value of the expected *foregone net payoff* or the expected *net opportunity cost* by committing an error in the decision, compared with making the desirable decision.

Note, however, the function $R(T;\theta,\sigma)$ involves the population parameters (θ, σ); generally, it cannot be evaluated with only a finite sample that the analyst has in hand.We assume a consistent estimator $\hat{\sigma}$ for σ is available and provide the necessary arguments to "plug in" $\hat{\sigma}$ in place of σ below.[6] To deal with the unknown θ, we derive our optimal decision rules based on the *maximum Type I and Type II regret*, defined as follows, where the maximum is taken over the true effect-size

rules (Karlin and Rubin, 1956, Theorem 1;Hirano and Porter, 2009, Theorem 3.4;Tetenov, 2012). Note the conventional one-sided Neyman-Pearson hypothesis-testing decision rule whereby the decision rule rejects the null when $\hat{\theta}$ is "large (small) enough" also falls into this class of cutoff decision rules.

[4] (K_1, K_2) can be interpreted as reflecting either the actual deployment costs or just the psychological preference weights for the status quo policy. It turns out that setting $(K_1 = \kappa, K_2 = 1)$ for any $\kappa > 0$ leads to the identical minimax-regret decision threshold as setting $(K_1 = 1, K_2 = 1/\kappa)$; therefore, we normalize $K_2 = 1$.

[5] Note $R(T;\theta,\sigma)$ is an example of an asymmetric loss function in the decision-theoretic terminology.

[6] The intuition is basically the same as studentizing the estimators using the estimated standard errors in the asymptotic t-tests.

parameter θ:

$$\begin{aligned}\bar{R}_{\text{Type I}}(T;\sigma) &= \max_{\theta\le 0}\left\{-\theta\Phi\left(\frac{\theta-T}{\sigma}\right)\right\}\\ \bar{R}_{\text{Type II}}(T;\sigma) &= \max_{\theta>0}\left\{\theta\Phi\left(\frac{T-\theta}{\sigma}\right)\right\}.\end{aligned}\tag{2.4}$$

The maximum Type I and Type II regrets represent the worst-case net payoff losses by committing a Type I or Type II error under the null $H_0 : \theta \le 0$, thereby modeling a conservative decision maker's decision rule. Put differently, $\bar{R}_{\text{Type I}}(T;\sigma)$ is the maximum (over θ) expected foregone payoff by rejecting the null; $\bar{R}_{\text{Type II}}(T;\sigma)$ is the maximum (over θ) expected foregone payoff from rejecting the alternative.[7]

With the maximum Type I and Type II regret functions defined in (2.4), for each possible $T \in \mathbb{R}$, the analyst makes a decision regarding whether to accept the new policy by making a decision that achieves a *lower* maximum regret, that is,

$$\min\left\{K_1 \cdot \bar{R}_{\text{Type I}}(T;\sigma), \bar{R}_{\text{Type II}}(T;\sigma)\right\}.\tag{2.5}$$

Verbally, problem (2.5) is to minimize either type of maximum regret (i.e., worst-case net foregone payoff), and hence is referred to as the *minimax*-regret decision problem.[8] When (T,σ) is given, the analyst should accept the new policy and reject the status quo policy if $K_1 \cdot \bar{R}_{\text{Type I}}(T;\sigma) \le \bar{R}_{\text{Type II}}(T;\sigma)$, that is, when the maximum (over θ) regret of mistakenly rejecting the null (the status quo policy) is smaller than the maximum regret of mistakenly rejecting the alternative (the new policy).

The optimal decision threshold T^* is then defined by a T such that Type I and Type II regrets are equalized as follows:

$$K_1 \cdot \bar{R}_{\text{Type I}}(T^*;\sigma) = \bar{R}_{\text{Type II}}(T^*;\sigma).\tag{2.6}$$

[9] The decision of whether to accept the new policy can be made by comparing the realized estimate $\hat{\theta}$ with T^*: reject the new policy if $\hat{\theta} < T^*$ and accept the new policy otherwise. Or, equivalently, the

[7]Another possibility is to take the expected regrets by invoking an assumed prior distribution over θ. We discuss the expected regrets and the associated expected-regret-minimizing decision rules in Appendix C. Note, from the Bayesian perspective, the maximum Type I and Type II regrets are the respective expected regrets evaluated using the worst-case priors.

The underlying Gaussian distribution is a smooth, well-behaved distribution. In turn, the magnitude of the maximum regrets, which can be evaluated without the knowledge of θ, is very informative about the actual regrets. In Section 5, we provide the relevant evidence using simulated data.

[8]We follow the conventional formulation of the minimax-regret decision problem, which is slightly different from the problem formulation of Tetenov. However, all the results developed and presented below can also be derived from Tetenov's problem formulation.

[9]Lemma 1 of Tetenov establishes that $\bar{R}_{\text{Type I}}(\cdot;\sigma)$ and $\bar{R}_{\text{Type II}}(\cdot;\sigma)$ cross once and only once; therefore, problem (2.6) is well defined for a given σ. See Figure 2.2 below for a graphical illustration.

analyst can compare $K_1 \cdot \bar{R}_{\text{Type I}}\left(\hat{\theta};\hat{\sigma}\right)$ against $\bar{R}_{\text{Type II}}\left(\hat{\theta};\hat{\sigma}\right)$ to find the minimizer of the problem (2.5) directly.[10] Intuitively, comparing $\hat{\theta}$ with T^* can be understood analogously as comparing the realized studentized estimate with the t-statistics threshold in one-sided asymptotic hypothesis tests. Similarly, $\bar{R}_{\text{Type I}}(T;\sigma)$ evaluated at $T=\hat{\theta}, \sigma=\hat{\sigma}$ can be understood analogously as the p-value for one-sided asymptotic hypothesis tests. Again, the key difference from hypothesis tests is that our proposed minimax-regret decision rule accounts for not only the probability but also the magnitude of the net foregone payoffs from committing *either* type of error.

We developed our framework assuming $\hat{\theta}$ follows a Gaussian distribution exactly with a known variance σ^2. However, the assumption is not likely to be satisfied in most practical circumstances – $\hat{\theta}$ may follow a Gaussian distribution asymptotically, with only a consistent estimator $\hat{\sigma}$ for the unknown σ being available to the analyst. In Appendix B, we provide the argument and necessary proofs that justify the plugging-in of $\hat{\sigma}$ in place of σ and the asymptotic approximation of the optimal-threshold decision rule T^*.

One implication of our minimax-regret decision rule is that the optimal threshold $T^*=0$ when the analyst weights the maximum Type I and Type II regrets equally, that is, when $K_1=1$. In Appendix C, we prove the same optimal threshold of $T^*=0$ can be derived when the analyst minimizes average regret, where the average is taken against the assumed prior that is symmetric around zero. In Online Appendix H, we also provide the pre-calculated table for T^* with varying K_1.

Graphical Illustrations Figure 2.1 depicts $R_{\text{Type I}}(T;\theta,\sigma=1)$ and $R_{\text{Type II}}(T;\theta,\sigma=1)$ as functions of θ and T while fixing $\sigma=1$. The colors represent the value of the respective regrets. Figure 2.2 illustrates $K_1 \cdot \bar{R}_{\text{Type I}}(T^*;\sigma)$, $\bar{R}_{\text{Type II}}(T;\sigma)$, and how the optimal threshold T^* is determined. The ridges (dashed line) of Figure 2.1 trace the maximum regret across possible values of T, the height of which coincides with the maximum-regret function values depicted in the top-left panel of Figure 2.2. The projection of the ridge is illustrated in Figure 2.2, from which the optimal threshold T^* is determined by the intersection of the two lines representing maximum Type I and Type II regrets.

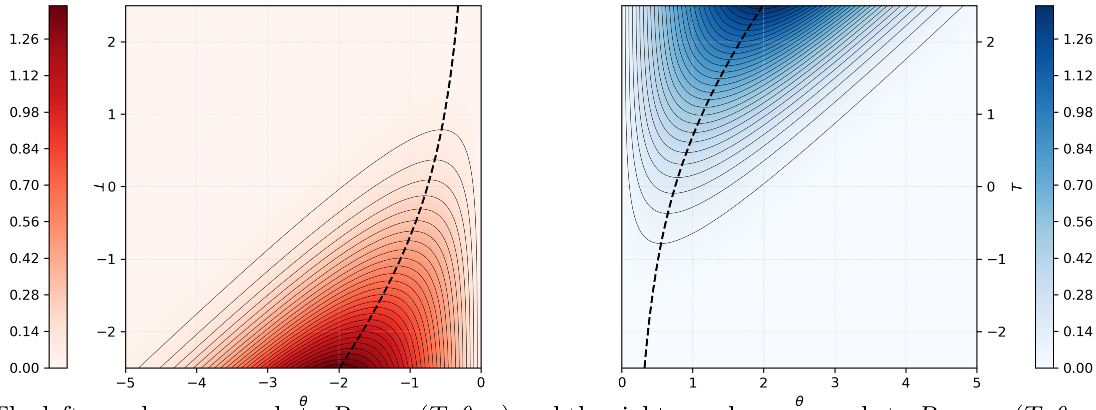
As Figure 2.1 shows, the Type I and Type II regret functions are smooth in (T,θ), and the maximum regret has an interior solution in θ across all values of T. Fixing T and deviating θ from the respective maximizer will give the respective regret values declining to zero rapidly. As such, the decision rules based on the maximum regrets are not affected "too much" by the extreme values of θ and would not be very different from those based on average regrets with a well-behaved prior.[11]

In Figure 2.2, we vary the asymmetry parameters $K_1 \in \{1,3\}$ and the magnitude of $\sigma \in \{1,2\}$. When $K_1=1$ (the left two panels), the optimal threshold T^* is zero regardless of the magnitude of

[10] We do not consider the case $\hat{\theta}=T^*$ as the event occurs with probability zero under Assumption 1.

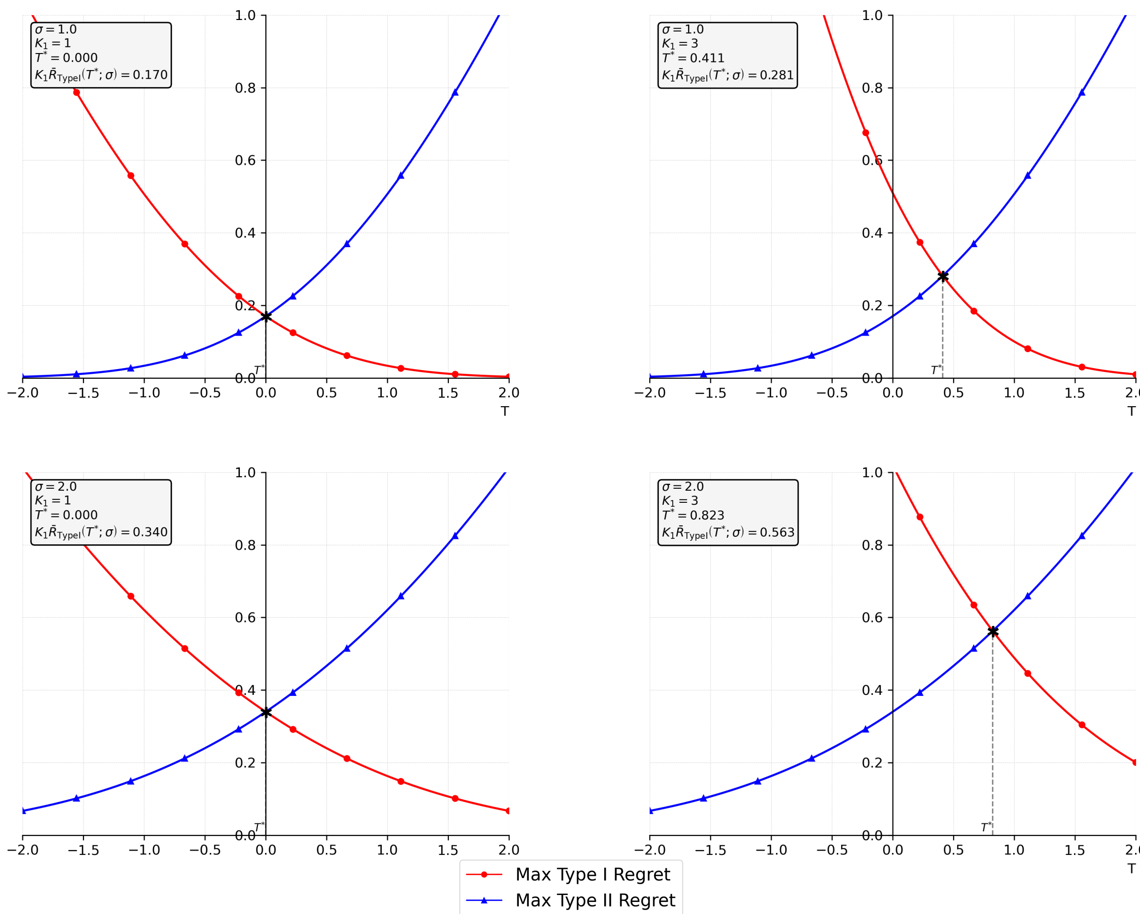
[11] In our Monte Carlo study (Section 5), we explore to what extent the maximum-regret functions can be good proxies for the regret functions.

Figure 2.1: Illustrations of Type I and Type II Regret Functions



Note. The left panel corresponds to $R_{\text{Type I}}(T;\theta,\sigma)$ and the right panel corresponds to $R_{\text{Type II}}(T;\theta,\sigma)$. The contour maps are drawn with $\sigma = 1$ and $K_1 = K_2 = 1$. The vertical axes represent T, the color represents the level of the regret, and the horizontal axes represent θ. The dotted-line trajectories trace the maximizers of θ at each level of T.

Figure 2.2: Illustrations of Max. Type I and Type II Regret Functions and the Optimal Thresholds



Note. The optimal threshold T^* is determined at the intersection of the maximum Type I and Type II regrets, marked by $*$. The horizontal axis represents T, the first argument of the respective maximum-regret functions. Parameters (σ, K_1), the optimal threshold T^*, and the max-minimax-regret value $K_1 \cdot \bar{R}_{\text{Type I}}(T^*;\sigma) = \bar{R}_{\text{Type II}}(T^*;\sigma)$ are displayed in the boxes, respectively.

σ. Hence, whenever the realized $\hat{\theta}$ is positive, the status quo policy should be rejected according to the minimax-regret decision rule. When $K_1 > 1$ (the right two panels), the optimal threshold T^* is greater than zero and varies with the level of σ.[12]

Notice the optimal threshold from our minimax-regret decision criteria is much less conservative than the hypothesis-testing counterpart. Recall the 5% significance-level threshold for one-sided hypothesis testing of $H_0 : \theta \leq 0$ rejects the null only when the realized t-value $(\hat{\theta}/\hat{\sigma})$ is larger than 1.645. The same threshold can only be rationalized with $K_1 = 102.4$, that is, when the analyst weights the (maximum) Type I regret at least 102.4 times more than the (maximum) Type II regret. We argue this decision rule is unnecessarily conservative and likely to be suboptimal for payoff maximizing. We finally note when $K_1 = 1$ and $\sigma = 1$, the maximum regret when $\hat{\theta}$ is realized right at the optimal threshold $T^* = 0$ is 0.17. This worst-case maximum regret can be calculated without evaluating $\hat{\theta}$; only a consistent estimate of σ is needed. We use this feature extensively in our efficient traffic-allocation criteria developed in Section 3.

Example: Effectiveness of a Promotion Campaign To put the maximum-regret functions into context, suppose the analyst wants to examine the effectiveness of a promotion campaign. The analyst runs a standard A/B test to measure θ. Here the outcome of interest, θ, is the difference in average revenue per user (ARPU) per day between the treatment group and the control group over the course of the campaign. The analyst sets $K_1 = 1$. After the sample collection and estimation, $\left(\hat{\theta}, \hat{\sigma}\right)$ turn out to be $\hat{\theta} = \$0.5$ and $\hat{\sigma} = \$1$. In this case, $\bar{R}_{\text{Type I}}(0.5; 1) = \$0.0814$, which is the maximum per-user expected net loss from rejecting the status quo policy if the new policy's true net effect is negative, and $\bar{R}_{\text{Type II}}(0.5; 1) = \$0.3102$, which is the maximum per-user expected foregone net benefit from sticking to the status quo policy if the new policy's true net effect is positive. Assume further that this platform has 1 million underlying users. When scaled by the number of users, the maximum Type I regret is around \$81,400 per day, and the maximum Type II regret is around \$310,200 per day. Therefore, abandoning the status quo policy in favor of the new policy is optimal. However, under the classical hypothesis tests, we would have declared $\hat{\theta}$ to be insignificant under the conventional significance levels, thereby failing to reject the status quo policy.

2.2 Sample-Collection Stopping Criteria That Controls for the Max-Minimax Regret

In this subsection, we develop the sample-collection stopping criteria based on the worst-case (i.e., maximum) minimax regret. The basic idea is to set a threshold ψ for the height of the asterisk in Figure 2.2 before starting the sample collection, and monitor only the height of the asterisk during the sample collection. The height of the asterisk declines to zero as the sample size increases, and the threshold ψ is the maximum worst-case (i.e., maximum) minimax regret that the analyst is

[12] We show the optimal T^* is in fact proportional to σ in Lemma 1 of Appendix A.

willing to tolerate. Because the analyst can monitor the height of the asterisk without monitoring $\hat{\theta}$ itself, the analyst should not "peek" at the current effect-size estimate $\hat{\theta}$ during the sample collection even if $\hat{\theta}$ can be readily calculated – $\hat{\theta}$ should be revealed to the analyst only when the height of the asterisk falls below ψ. The proposed sample-collection stopping criteria are not subject to the widespread p-hacking scheme of monitoring the estimate $\hat{\theta}$ itself or the associated p-value midway through the sample collection and rejecting the null once the p-value hits the desired significance level (see, e.g., Armitage et al., 1969; Johari et al., 2017, Forthcoming).[13]

In previous work by Tetenov (2012), only the threshold T^* from the minimization problem is of interest in the analyst's decision problem, not the level of the maximum Type I and Type II regret. However, when the analyst can effectively choose the sample size n that determines σ, as is common in the online A/B test environments, we propose monitoring the value of the following max-minimax-regret function at the optimal threshold T^* that equalizes the arguments of the min operator in (2.5), that is,

$$V(\sigma; K_1) := K_1 \cdot \bar{R}_{\text{Type I}}(T^*; \sigma) = \bar{R}_{\text{Type II}}(T^*; \sigma) . \tag{2.7}$$

The function value $V(\sigma; K_1)$ is defined by the worst-case (i.e., maximum) minimax regret, which is attained if $\hat{\theta}$ is realized exactly at the optimal threshold T^*. Put differently, $V(\sigma; K_1)$ represents the maximum net loss in payoff by committing either type of error in a decision when the analyst follows the minimax-regret decision rule subsequently.

Lemma 1 in Appendix A establishes the value of $V(\sigma; K_1)$ and the optimal threshold T^* as linear functions of σ, respectively. Specifically, $V(\sigma; K_1) = \sigma V(1; K_1)$ and $T^* = \sigma T^*_{\sigma=1}$. Therefore, Lemma 1 reduces the problem of bounding the worst-case minimax regret to the problem of pre-calculating $V(1; K_1)$ as a function of K_1. We propose a nested decision rule to first collect a large-enough sample to control $V(T^*; \sigma)$ under a certain level ψ that is the analyst's decision variable, and then compare $\hat{\theta}$ with T^* to decide whether to accept or reject the innovative new policy. Again, we assume $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2)$, where σ^2 declines to zero as the sample size n increases. Formally, the sample-collection stopping problem we described thus far can be formulated as follows:

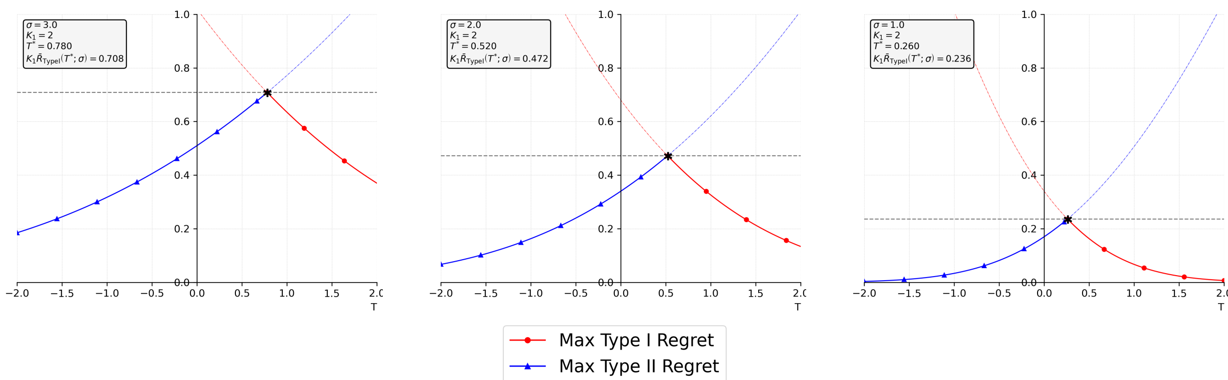
$$\begin{aligned} & \min_{n \geq 0} n \\ s.t. \quad & \sigma V(1; K_1) \leq \psi \\ & \sigma = \frac{s}{\sqrt{n}}, \end{aligned} \tag{2.8}$$

[13]Hypothesis tests intend to control the Type I error probability under some pre-set significance level. The "peeking and optional stopping" p-hacking scheme is problematic because it induces the resulting actual Type I error probability to be higher than the desired significance level. By contrast, the goal of our minimax-regret decision framework is to find a payoff-improving policy while controlling for the maximum net foregone payoff under a pre-set level. Monitoring the height of the asterisk does not undermine this goal.

where $\sigma V(1;K_1) = V(\sigma;K_1) = K_1 \cdot \bar{R}_{\text{Type I}}(T^*;\sigma) = \bar{R}_{\text{Type II}}(T^*;\sigma)$ and s is a positive constant.[14]

The following Figure 2.3 illustrates how max-minimax regret changes as σ declines to zero, that is, as the sample size increases. The coordinate where the max-minimax regret is attained is marked with asterisks, and the level of max-minimax regret is the horizontal dotted lines. The analyst may stop sampling once the level of the horizontal dotted line is below a pre-set threshold ψ. In Online Appendix H, we also provide the pre-calculated table for $V(1;K_1)$ with varying K_1. Combined with the relation $V(\sigma;K_1) = \sigma V(1;K_1)$, the level of the max-minimax regret for any combination of (σ, K_1) can be easily calculated using only the table and a calculator.

Figure 2.3: Illustrations of the Max-Minimax Regret When $K_1 = 2$ and $\sigma = 3, 2, 1$



Note. The solid lines represent the effective maximum Type I and maximum Type II regrets when the analyst follows the minimax-regret decision rule. The horizontal dotted lines represent the level of max-minimax regret. The horizontal axis represents T, the first argument of the respective maximum-regret functions. Parameters (σ, K_1), the optimal threshold T^*, and the max-minimax regret value $K_1 \cdot \bar{R}_{\text{Type I}}(T^*;\sigma) = \bar{R}_{\text{Type II}}(T^*;\sigma)$ are displayed in the boxes, respectively.

2.3 The Minimax-Regret Decision Criteria for Multi-Arm Experiments

In this subsection, we develop the minimax-regret decision criteria for multi-arm experiments. In online experiments, an analyst commonly explores multiple treatment arms simultaneously, where the analyst has to choose one best policy among all the treatment arms and the control group. The idea here is to filter and sort the policies based on the metric that is directly relevant to the payoff of interest. Along this line, we propose calculating the maximum Type I regrets for each of the proposed new policies, filtering the new policies whose effect size exceeds the respective optimal threshold, and selecting the policy that achieves the lowest maximum Type I regret. The multi-arm experiment sorting and decision procedure developed below give a *total ordering* on the set of all the policies being examined.

[14] s^2 is the variance term in $\sqrt{n}\left(\hat{\theta} - \theta\right) \to_d \mathcal{N}\left(0, s^2\right)$. Note we assumed the $\sqrt{n}$ convergence rate for simplicity in exposition. The problem can be reformulated with proper rate-of-convergence adjustments when the convergence rate is slower than $\sqrt{n}$.

Denote $\{1, 2, 3, ..., J\}$ by the different marketing policies being examined, where policy 1 denotes the status quo policy, that is, the control group. We assume the analyst sets $K_1 > 0$ as a decision variable but takes $K_2 = ... = K_J = 1$ as fixed. The assumption reflects that all the new policies are ex-ante homogeneous; however, the analyst may prefer the status quo policy ex ante because of, for example, possible deployment costs.[15] Throughout, we use the shorthand $r1f2$ for "reject policy 1 in favor of policy 2", and so on. Denote $\theta_{j,1}$ by the true population (average) treatment effect of treatment policy j compared with the baseline policy 1. Let $\hat{\theta}_{j,1} \sim \mathcal{N}\left(\theta_{j,1}, \sigma_{j,1}^2\right)$. Again, we assume consistent estimates $\{\hat{\sigma}_{j,1}\}_{j=2}^J$ for $\{\sigma_{j,1}\}_{j=2}^J$ are available to the analyst.[16]

For the null hypothesis $H_0^j : \theta_{j,1} \leq 0$, the maximum Type I regret at the current effect-size estimate $\hat{\theta}_{j,1}$ is the maximum regret of the event $r1f2$ (rejecting option 1 in favor of option 2).[17] One can then run comparisons for each of the new policies $\{2, 3, 4, ..., J\}$ with the status quo policy $\{1\}$. We propose the following algorithm:

1. Calculate the respective maximum Type I regrets $\bar{R}_{r1f2}\left(\hat{\theta}_{2,1}; \hat{\sigma}_{2,1}\right)$, $\bar{R}_{r1f3}\left(\hat{\theta}_{3,1}; \hat{\sigma}_{3,1}\right)$, ..., $\bar{R}_{r1fJ}\left(\hat{\theta}_{J,1}; \hat{\sigma}_{J,1}\right)$ and maximum Type II regrets $\bar{R}_{r2f1}\left(\hat{\theta}_{2,1}; \hat{\sigma}_{2,1}\right)$, $\bar{R}_{r3f1}\left(\hat{\theta}_{3,1}; \hat{\sigma}_{3,1}\right)$, ..., $\bar{R}_{rJf1}\left(\hat{\theta}_{J,1}; \hat{\sigma}_{J,1}\right)$.
2. Find the associated optimal thresholds $\left\{T_{j,1}^*\right\}_{j=2}^J$ for each $j \in \{2, 3, ..., J\}$, taking the weighting factor K_1 as fixed.
3. Take the index set $\mathcal{I} = \left\{j : \hat{\theta}_{j,1} > T_{j,1}^*\right\}$, i.e., the subset of the new policies that exceeds the respective optimal threshold.
   (a) If the set $\mathcal{I}$ is empty, stick with the status quo policy 1.
   (b) If the set $\mathcal{I}$ is nonempty, sort $\mathcal{I}$ by the maximum Type I regrets evaluated at $\hat{\theta}_{j,1}$ (i.e., by $\bar{R}_{r1fj}\left(\hat{\theta}_{j,1}; \hat{\sigma}_{j,1}\right)$), and then choose the new policy that achieves the lowest $\bar{R}_{r1fj}\left(\hat{\theta}_{j,1}; \hat{\sigma}_{j,1}\right)$.

The procedure provides a total ordering over set $\mathcal{I}$, leading to the unique best policy that achieves the lowest maximum Type I regret.[18]

[15] Our sorting algorithm below can be easily extended to when the analyst decides all the K_js.

[16] Note, for any reasonable estimator $\hat{\theta}_{j,1}$ for $\theta_{j,1}$, the following symmetry should hold: $\hat{\theta}_{j,1} = -\hat{\theta}_{1,j}$ with $\sigma_{j,1} = \sigma_{1,j} = \sqrt{Var\left(\hat{\theta}_{j,1}\right)}$.

[17] For example,

$$\bar{R}_{r1fj}\left(\hat{\theta}_{j,1} = 0.5; \hat{\sigma}_{j,1} = 1\right) \equiv \bar{R}^j_{\text{Type I}}\left(\hat{\theta}_{j,1} = 0.5; \hat{\sigma}_{j,1} = 1\right) = 0.0814$$
$$\bar{R}_{rjf1}\left(\hat{\theta}_{j,1} = 0.5; \hat{\sigma}_{j,1} = 1\right) \equiv \bar{R}^j_{\text{Type II}}\left(\hat{\theta}_{j,1} = 0.5; \hat{\sigma}_{j,1} = 1\right) = 0.3102.$$

[18] The policy filtering and sorting procedure we propose is based only on the maximum Type I regrets, not the maximum Type II regrets, to ensure a total ordering of the policies. In principle, one might consider the policy sorting algorithm by comparing both the maximum Type I and Type II regrets pairwise for all the policies considered. However, such a pairwise comparison and sorting may violate transitivity (e.g., $2 \succ 3$, $3 \succ 4$, and $4 \succ 2$ can occur).

The minimax-regret-based best policy selection favors the policy with higher t-values and lower $\hat{\sigma}_{j,1}$; that is, it favors a higher effect-size estimate with higher precision, which is intuitive. Notably, our minimax-regret best policy-selection rule, which also accounts for the magnitude of the foregone payoff by committing a decision error, may give a different ordering compared to those obtained by simply sorting the respective t-values or p-values that only account for the probability of committing an error. Suppose, for instance, policies 2 and 3 attained the same t-value of $1 = \hat{\theta}_{j,1}/\hat{\sigma}_{j,1}$ with $\left(\hat{\theta}_{2,1}, \hat{\sigma}_{2,1}\right) = (1,1)$, $\left(\hat{\theta}_{3,1}, \hat{\sigma}_{3,1}\right) = (2,2)$, and the optimal threshold $T^*_{2,1} = T^*_{3,1} = 0$ so both policies 2 and 3 exceeded the t-value thresholds. If the analyst simply sorts the policies according to their t-values or p-values, policies 2 and 3 should be indifferent. However, $\bar{R}_{r2f1}\left(\hat{\theta}_{2,1}; \hat{\sigma}_{2,1}\right) = 0.0334$, whereas $\bar{R}_{r3f1}\left(\hat{\theta}_{3,1}; \hat{\sigma}_{3,1}\right) = 0.0668$, and therefore, policy 2 is more favorable than policy 3 according to our minimax-regret multi-arm decision criteria. Another intuitive explanation is that our minimax-regret multi-arm policy-selection rule favors the sample evidence obtained from a larger sample (smaller $\hat{\sigma}_{j,1}$) better if the attained t-stats are the same.[19]

3 The Minimax-Regret-Decision-Rule-Based Experiment Design: The Efficient Traffic Allocation

In the previous section, we developed the minimax-regret experiment-*evaluation* scheme of two-arm and multi-arm experiments, taking the experiment *design* as given. Our novel contribution in this section is to develop the minimax-regret optimal experiment-*design* scheme, where the goal of the analyst is to minimize the wait time of sample collection until the desired level of max-minimax regret is attained. The basic idea is again to control for the height of the asterisks in Figure 2.2 in Section 2, but here, we optimize the traffic-allocation ratio to control for the highest height of the asterisks across all the treatment arms. To this end, we propose an adaptive, efficient traffic-allocation scheme in which the analyst monitors the current variance estimates of the effect-size estimators and allocates the incoming traffic accordingly.[20]

3.1 The Efficient Traffic-Allocation Probabilities across Experiment Arms

In online A/B tests, an analyst typically does not have control over the overall rate of the incoming traffic, but s/he may have control over the random-assignment probability of the incoming sample

[19] See Online Appendix G for a real-world example.

[20] Our optimal traffic-allocation scheme differs from the popular Thompson sampling in the following ways. First, the benefits/losses from the experiment sample can be ignored in our experiment setup (i.e., the "pure-exploration bandits" problem), which is different from the setup where Thompson sampling is known to work well (i.e., where the "exploration-exploitation tradeoff" for the payoff matters in the test sample). Second, our approach is frequentist in that it does not require assuming the shape of the true parameter's prior distribution. Third, we propose monitoring only the variance estimate of the effect-size estimator, not the estimate itself. Last but not least, the optimal traffic-allocation scheme is derived in close relation to the experiment-stopping criteria that explicitly control for the maximum regrets, the measure directly relevant to the payoffs undergoing experiments.

to treatment arms or the control group. That is, at any given time, the total sample size n is exogenously given, but the random-assignment probability over the experiment arms can be chosen by the analyst. As the analyst changes the assignment probability, σ in (2.8) in Section 2.2 would change. Furthermore, in multi-arm experiments, different σ's would correspond to each treatment arm.

In this section, we assume either the absence of a pretreatment-period sample, or that the analyst can utilize a size-n_0 pretreatment sample as a part of the *common* control group, as follows:

Assumption 2. *If pretreatment-period data with size $n_0 > 0$ are used for the analysis, they are part of the common control group for all treatment arms $\{2, 3, ..., J\}$.*

The assumption is innocuous to the extent that the assignment of the incoming samples to the experiment arms is randomized, which is the fundamental premise of online A/B tests. It allows for controlling for the common time trends and other covariates. However, it rules out using the same treatment-arm-specific subsample as the respective control group, and in turn, using the estimators that hinge upon a non-random assignment of the samples. For example, a difference-in-differences estimator would not work well with our efficient traffic-allocation algorithm.[21]

We seek for the efficient treatment-period random-assignment probability $\{p_1, p_2, ..., p_J\}$ that minimizes the maximum (over $k = 1, 2, ..., J$) worst-case (over T) maximum Type I regret over the pairs $\{1, 2\}$, $\{1, 3\}$, ..., $\{1, J\}$, where we denote $j = 1$ by the control group as before. Define n_k for $k = 2, 3, ..., J$ such that $p_k = \frac{n_k}{n-n_0}$ and denote $p_0 = \frac{n_0}{n-n_0}$. Denote Y by the outcome of interest and $D_j = \mathbf{1}\left(\text{Observation is treated in arm } j\right)$. Because we assume the unconfoundedness and the propensity score is assigned by the analyst so it is known, that is, $\{Y(j)\}_{j \in \mathcal{J}} \perp\!\!\!\perp D_j|\mathbf{x}$ and $\Pr\left(D_j = 1|\mathbf{x}\right)$ is known, a $\sqrt{n}$-consistent estimator $\hat{\theta}_{j,1}$ for the effect size $\theta_{j,1}$ is generally available (e.g., Horvitz and Thompson, 1952; Kitagawa and Tetenov, 2018; Athey and Wager, 2021; Farrell et al., 2021).

We also impose the following assumption in the remainder of this section about the structure of $\sigma^2_{j,1}$, the asymptotic variance of the effect-size estimator.

Assumption 3. *$\sigma^2_{j,1}$, the variance of the effect-size estimator $\hat{\theta}_{j,1}$ has the form*

$$\sigma^2_{j,1} = \frac{\eta^2_j}{n_j} + \frac{\eta^2_1}{n_0 + n_1} + o_p\left(1\right) = \frac{1}{n}\left(\frac{\eta^2_j}{p_j} + \frac{\eta^2_1}{p_0 + p_1}\right) + o_p\left(1\right), \tag{3.1}$$

where η^2_j and η^2_1 are specific to group j and 1, respectively.

$o_p\left(1\right)$ is a term that converges in probability to zero when n is large, which does not depend on the assignment probabilities $\{p_0, p_1, \ldots, p_J\}$ asymptotically. (3.1) encompasses a broad class of

[21] We note our efficient traffic-allocation algorithm would still provide an approximately efficient solution with treatment-arm-specific pretreatment samples when the pretreatment-period sample size is relatively small.

the widely used effect-size estimators that assume unconfoundedness, including but not limited to a linear regression estimator with and without covariates, a regression-adjustment estimator (e.g., Wooldridge, 2010, p.918), and any efficient semiparametric estimator (Hahn, 1998). We implicitly imposed the $\sqrt{n}$-rate of convergence for $\hat{\theta}_{j,1}$, but the results we provide below are likely to be generalized with a rate-of-convergence adjustment for a non/semiparametric estimator that has a slower rate of convergence (e.g., Wooldridge, 2010, pp. 917-918). We work out some commonly used estimators' variance formulas in Appendix D.1.

Under Assumptions 2 and 3, the analyst wants to randomly allocate the incoming sample optimally so that the maximum (across treatment arms) max-minimax regret $V(\sigma_{j,1}; K_1)$ is minimized, which translates directly to minimizing the wait times.[22] Mathematically, the analyst's optimal traffic-allocation problem is given by

$$\min_{\{p_1, p_2, \ldots, p_J\}} \left\{ \max_{j \in \{2,3,\ldots,J\}} \{\sigma_{j,1} V(1; K_1)\} \right\} \tag{3.2}$$

$$s.t. \quad n = n_0 + \sum_{j=1}^{J} (n - n_0) p_j$$

$$\sigma_{j,1}^2 = \frac{\eta_j^2}{n_j} + \frac{\eta_1^2}{n_0 + n_1},$$

where n is the total sample size at any given time and $(n - n_0)$ is the treatment-period sample size. Note $\sigma_{j,1}$ in (3.2) is determined by (3.1), a function of (p_j, p_1). Consistent estimates for η_j^2s can be plugged in during implementation.

Problem (3.2) is a J-dimensional optimization problem. However, exploiting the optimality conditions, we prove (3.2) can be effectively reduced to a one-dimensional numerical optimization problem over p_1, as follows:

$$p_1^* = \arg \min_{p_1 \in [0,1]} \left\{ \frac{\eta_1^2}{p_0 + p_1} + \frac{\sum_{j=2}^{J} \eta_j^2}{1 - p_1} \right\} \tag{3.3}$$

$$p_j^* = (1 - p_1^*) \frac{\eta_j^2}{\sum_{k=2}^{J} \eta_k^2} \quad \text{for } j > 1.$$

Problem (3.3) is a convex minimization problem with one variable, which can be easily solved numerically using any standard nonlinear optimizer. See Appendix D.2 for the derivation of (3.3) and Appendix D.3 for the solution when $\eta_1^2 = \eta_2^2 = \ldots = \eta_J^2$.

[22] We note again that the Assumptions 2 and 3 are not required if the analyst wants to only evaluate the results from a multi-arm experiment without optimizing the traffic-allocation probabilities.

3.2 Automation Algorithm of the Adaptive Randomized Traffic Allocation and Stopping of the Experiment

Using the efficient traffic-allocation and optimal stopping rules developed above, we provide the automation algorithm below. The algorithm outlined below adaptively and dynamically applies the efficient traffic-allocation rule over the course of sample collection, monitors $\sigma_{j,1}$'s continuously, and stops the experiment when the worst-case minimax regret is below the analyst's pre-set decision variable ψ.

1. The analyst is endowed with p_0, and decides (K_1, ψ).
2. Begin from assuming p_ks, for example, $p_1 = p_2 = ... = p_J$.
3. Wait until enough samples (typically at least thousands) are collected to estimate η_k^2's consistently.
4. Estimate the current $\left\{\eta_1^2, \eta_2^2, ..., \eta_J^2\right\}$.
5. Solve (3.3) to find $\{p_1^*, p_2^*, ..., p_J^*\}$ and, in turn, the target $\{n_1^*, n_2^*, ..., n_J^*\}$ at the current monitoring cycle. Note, $n_1^* < n_1$ is possible when $p_0 = \frac{n_0}{n-n_0}$ is not sufficiently small at the current n, and hence, not allocating any incoming sample to the control arm $j = 1$ at the current monitoring cycle can be optimal.
6. Calculate $\{\sigma_{1,j}\}_{j=2}^{J}$ and evaluate whether $\max_{j \in \{2,3,...,J\}} \{\sigma_{j,1} V(1; K_1)\} \leq \psi$.
   (a) If the condition is satisfied, terminate the experiment and make a decision following the steps outlined in Section 2.3.
   (b) If the condition is not satisfied, wait until more samples are collected, and go back to Step 4.

In practice, Steps 4-6 can be conducted in periodic cycles, for example, hourly, daily, every other day, and so on. Also, a smooth transition of the target $\{n_1^*, n_2^*, ..., n_J^*\}$ from the previous sample-collection cycle can be considered when Steps 4-6 are conducted with high frequency.

4 Empirical Application: A Mobile Game Company's Multi-Arm Experiment

In this section, we apply our minimax-regret decision framework to a real-world multi-arm experiment. Our finding in this section shows the decision rules based on our minimax-regret decision framework and the conventional hypothesis testing may diverge drastically.

4.1 Background and Experiment Setup

We collaborate with a large mobile gaming studio in the Asia-Pacific region. The flagship app developed by this studio is a casual gaming app with over 20 million daily active users. A major source of the company's revenue is in-app purchases, where users purchase game items using in-game currency ("Gold Coins" henceforth), which in turn are bought using real-world currency. These items give an advantage to the users in solving the level and advancing to the next level. The only way to obtain Gold Coins in the game is to purchase them using the real currency in the game store. The exchange rate between the Gold Coins and 1 USD is around 100:1,[23] and the average daily revenue from in-app Gold Coin purchases ranges from 100,000 USD to 1 million USD per day.[24]

Each game stage consists of solving a puzzle within a set number of "steps," and the company was interested in the effect of the game difficulty on the users' in-game currency spending behavior. Adjusting the difficulty of the game can lead to drastic changes in the company's revenue, and it is one of the most important levers that the company can use to maximize revenue. When the game is too easy, users will find it not challenging enough and will not buy in-app items that would help them solve the puzzles. When the game is too difficult, users will lose motivation and give up. The company therefore ran a multi-arm experiment varying the difficulty level of the game for 34 days, from 7/28/2022 to 8/30/2022.

The outcome variable of interest is the in-game currency spending. In-game currency spending is defined at the user-day level. The company was interested only in the instantaneous (i.e., daily) treatment effects, because the company can switch policies relatively frequently.[25]

The pre-intervention period is 14 days, and the intervention period is 20 days, which is a common experiment duration (e.g., Lewis and Rao, 2015; Berman and Van den Bulte, Forthcoming). The company randomly allocated users into control and treatment groups according to a 2:1:1:1 ratio. The status quo policy, which constitutes our control group, gives the hardest puzzles to the users. Each treatment group varies in the difficulty of the puzzles given, as described in Table 1. We note the company determined the traffic-allocation ratio for each group in an ad-hoc manner.

4.2 Multi-Arm Experiment Results

We compare the estimator $\hat{\theta}$s for the following regression specification across different treatment groups $j \in \{\text{Easy}, \text{Medium}, \text{Hard}\}$:

$$\text{coin\_consumption}_{it} = \gamma_{j,\text{Baseline}} \cdot \mathbf{1}(\text{post\_exp}_t) + \theta_{j,\text{Baseline}} \cdot \mathbf{1}(\text{treat}_i) \cdot \mathbf{1}(\text{post\_exp}_t) + \epsilon_{it}. \quad (4.1)$$

[23] The price of 1 Gold Coin is around 0.07-0.08 RMB, which, corresponds to 0.01 USD.

[24] Because the company is a private entity, we cannot disclose the exact number of daily users and daily Gold Coin purchases, due to the nondisclosure agreement.

[25] Measuring the long-term causal effects of the change in game difficulty on users' Gold Coin spending behavior exceeds the scope of the present research.

Table 1: Description of Treatment and Control Groups

	Group	Pre-treat n_j	Post-treat n_j	n_j	Traffic Allocation
Control	Baseline (Hardest)	1,554,851	2,135,684	3,690,535	40%
Treat	Easy	773,395	1,069,090	1,842,485	20%
	Medium	774,929	1,069,873	1,844,802	20%
	Hard	775,953	1,075,810	1,851,763	20%

Note. This table reports the description, size, and randomized traffic-allocation probabilities of the control and treatment groups, respectively.

We include $\mathbf{1}\left(\text{post\_exp}_t\right)$ to control for time fixed effects, and we do not include group fixed effects, because the users are assigned randomly into control/treatment groups. We run three separate regressions for each of the treatment groups, obtain the respective $\hat{\theta}_{j,\text{Baseline}}$, and then apply the decision criteria developed in Sections 2.1-2.3.[26] We use the weighting factor $K_1 = K_2 = 1$ for our minimax-regret decision framework; that is, we place the same weight on the maximum Type I and Type II regrets, respectively, and we provide the robustness-check results for $K_1 = 3$ in Appendix F.

Table 2: Summary of the Multi-arm Experiment Results with $K_1 = 1$

Control	Baseline (Hardest)		
Treatment	Easy	Medium	Hard
$\hat{\theta}$	0.02849	0.02430	0.04827
$\hat{\sigma}$	0.02627	0.02788	0.03230
$\hat{\theta}/\hat{\sigma}$	1.08439	0.87158	1.49417
$\bar{R}\left(T^*;\hat{\sigma}\right) \equiv V\left(\hat{\sigma};1\right)$	0.00447	0.00474	0.00549
$\bar{R}_{\text{Type I}}\left(\hat{\theta};\hat{\sigma}\right)$	0.00074	0.00119	0.00038
$\bar{R}_{\text{Type II}}\left(\hat{\theta};\hat{\sigma}\right)$	0.01432	0.01257	0.02432
Minimax Regret T^*	0	0	0
$\alpha = 0.01$ HT Threshold	0.06112	0.06486	0.07515
Minimax -Regret Decision	Accept New	Accept New	Accept New
Hypothesis-Testing Decision	Reject New	Reject New	Reject New

Note. The minimax-regret decision threshold T^*s are derived using $K_1 = K_2 = 1$. Standard errors are clustered at the user level to calculate $\hat{\sigma}$, because the same user can be sampled across multiple days.

[26] We use the linear regression estimator, following what the company used to evaluate the treatment effects, and estimate the treatment effects pairwise for each treatment group. All the results presented below are robust with insignificant group-dummy estimates when the group fixed effects are included, that is, in the usual difference-in-differences specification.

Table 2 summarizes the results. The minimax-regret optimal threshold T^*, which $\hat{\theta}$s are compared against, is zero for all the columns as derived in Section 2.1 because $K_1 = K_2 = 1$. Notice all three new policies exceed the minimax-regret optimal threshold, thereby implying either *Easy*, *Medium*, or *Hard* could be accepted if compared pairwise against the *Baseline (Hardest)*. By contrast, none of the new policies' $\hat{\theta}$ exceeds the hypothesis-testing thresholds at the 1% significance level. If the company followed the hypothesis-testing-based decision rules, the company should make no changes to the game difficulty.[27]

Our next question is which of the three new policies should be adopted in place of the *Baseline* policy. Among them, *Hard* achieves the lowest maximum Type I regret at the current $\left(\hat{\theta}, \hat{\sigma}\right)$ estimates. Therefore, the company should adopt *Hard* difficulty to minimize the maximum possible loss of revenue by in-game currency expenditure.

Our minimax-regret decision criteria provide an interpretable and decision-relevant measure to the analyst, unlike the t-value- or p-value-based decision criteria. $\bar{R}_{\text{Type I}}\left(\hat{\theta}; \hat{\sigma}\right)$, the maximum Type I regret of *Hard*, is 0.00038, implying the worst-case net expected loss from rejecting the *Baseline* in favor of *Hard*, scaled by 20 million underlying users would be only 7,600 Gold Coins (USD 76) per day. By contrast, $\bar{R}_{\text{Type II}}\left(\hat{\theta}; \hat{\sigma}\right) = 0.02432$, which is around 64 times larger than the maximum Type I regret, implying the worst-case net expected loss compared with *Hard* would be around 486,400 Gold Coins (USD 4,864) per day if the company sticks with the *Baseline* policy.

The experiment results and the decision rules presented in this section reconfirm that the decision rules based on our minimax-regret framework and the conventional hypothesis testing may diverge dramatically in practice. Specifically, the decisions diverge drastically when the t-value is positive but not "large enough." However, evaluating and comparing the performances of the decision rules is impossible without knowing the true effect sizes of the underlying data-generating process. Therefore, in the following Section 5, we run a series of simulation exercises in which the data-generating process closely mimics the experiments analyzed above and the underlying true parameter values are known. In Online Appendix G, we analyze another multi-arm experiment run by our focal company, in which, interestingly, our minimax-regret-based policy preference orderings turn out to be different from ordering the new policies by their respective p-values.

5 Monte Carlo Experiments

In this section, we validate the proposed minimax-regret decision framework by running a series of Monte Carlo simulations, in which the data-generating process closely replicates the multi-arm experiments we analyzed in the previous section. This section is divided into two parts. In the first part (Section 5.1), we evaluate and compare the performance of our minimax-regret decision rule against the hypothesis-testing decision rule, taking the sample size and traffic-allocation rules

[27] The hypothesis-testing thresholds are different across columns because they are calculated using different subsamples of treatment/control groups.

as given by what our focal company arbitrarily set during the experiment. In the second part (Section 5.2), we evaluate the performance of the efficient traffic allocation rule we proposed in Section 3, by examining the reduction of the wait time and the required sample size to reach the same max-minimax-regret threshold.[28]

5.1 Monte Carlo Study 1: Performance Comparison of the Minimax-Regret Decision Rule and Hypothesis Testing

We generate our Monte Carlo datasets in a way that closely mimics our focal mobile game company's experiment. Specifically, we take the $\hat{\theta}_j$ estimates presented in Table 2 as the true effect-size parameter $\theta_j \equiv \mu_j - \mu_1$'s. We also use the same underlying daily user size of 20,000,000 and the same traffic allocation probabilities as reported in Table 1 of Section 4.1. Because the Gold Coin consumption is nonnegative and the distribution is positively skewed, we generate the Gold Coin consumption variables from a lognormal distribution.[29]

We repeat the same data-generating process independently for 1,000 times and implement our minimax-regret decision rules and the hypothesis-testing decision rules, respectively. Because we know the true effect-size parameter during the data-generating process, we can evaluate the realized regret $R\left(T = \hat{\theta}; \theta, \sigma = \hat{\sigma}\right)$ at the current $\hat{\theta}$ and $\hat{\sigma}$ estimates, that is, at $T = \hat{\theta}$ and $\sigma = \hat{\sigma}$.

Table 3 summarizes the Monte Carlo data-generating process and reports the results from 1,000 independent trials. Table 3a summarizes the data-generating process. For all the treatments, of which θ values are positive, the correct decision should be to accept the new policy. All the true treatment effects are fairly small in terms of the relative magnitude, ranging from 2.22% to 4.43%[30] of the *Baseline* average Gold Coin consumption.

Turning to Table 3b *Average Estimates* columns and comparing the results with Table 2 in Section 4.2, we confirm the Monte Carlo data-generating process closely replicates the field experiment. Next, comparing Table 3b's *Minimax Regret* and $\alpha = 0.01$ *HT* columns, the divergences in the realized decisions across different decision rules are striking. Because the magnitudes of the true treatment effects are not large compared with the *Baseline* policy, only 5.4%-25.0% of the 1,000 trials' t-values exceeded the $\alpha = 0.01$ hypothesis-testing threshold if compared pairwise. By contrast, decisions based on our minimax-regret decision criteria achieve very good performance, designating the correct decision in 81.2%-95.6% of trials if compared pairwise.

We calculate the columns *Revenue Lost / Day* as (100% − *% Correct*)×θ, which is the ex-post

[28] Ideally, one could evaluate the performance and validate the proposed minimax-regret decision framework by implementing it in the field. However, a reliable validation would require horse-racing the performances of our minimax-regret decision framework against the hypothesis tests across hundreds of field experiments at a minimum. Due to time and resource constraints, we instead run extensive Monte Carlo experiments.

[29] To generate $X \sim \text{lognormal}\left(\mu, \sigma^2\right)$ with $\mathbb{E}[X]$ and $Var(X)$, the parameters can be set as $\mu = \ln\left(\frac{\{\mathbb{E}[X]\}^2}{\sqrt{\{\mathbb{E}[X]\}^2 + Var(X)}}\right)$, $\sigma^2 = \ln\left(1 + \frac{Var(X)}{\{\mathbb{E}[X]\}^2}\right)$.

[30] 0.02430/1.0874 = 0.0223 and 0.04827/1.0874 = 0.0443.

Table 3: Monte Carlo 1: Data-Generating Process and the Evaluation Results

(a) Setup and Data-Generating Process

	Avg. Coin Consumption	θ	Traffic Allocation
Baseline (Hardest)	1.0874	—	40%
Easy	1.1159	0.02849	20%
Medium	1.1117	0.02430	20%
Hard	1.1357	0.04827	20%

(b) Results and Performance Comparison

	Average Estimates			Minimax Regret		$\alpha = 0.01$ HT	
	$\hat{\theta}$	$\hat{\sigma}$	$\hat{\theta}/\hat{\sigma}$	% Correct	Revenue Lost / Day	% Correct	Revenue Lost / Day
Easy	0.03095	0.02833	1.05687	85.6%	$820.80	7.5%	$5,272.50
Medium	0.02533	0.02732	0.86812	81.2%	$913.68	5.4%	$4,597.56
Hard	0.04956	0.03020	1.67338	95.6%	$425.04	25.0%	$7,245.00

(c) Scaled Regrets and Scaled Max. Regrets

	Scaled Regrets			Scaled Max. Regrets		
	Scaled $R_{\text{Type I}}\left(\hat{\theta};\theta,\hat{\sigma}\right)$	Scaled $R_{\text{Type II}}\left(\hat{\theta};\theta,\hat{\sigma}\right)$	Scaled $R\left(\hat{\theta};\theta,\hat{\sigma}\right)$	Scaled $\bar{R}_{\text{Type I}}\left(\hat{\theta};\hat{\sigma}\right)$	Scaled $V\left(\hat{\sigma};1\right)$	Scaled $\bar{R}_{\text{Type II}}\left(\hat{\theta};\hat{\sigma}\right)$
Easy	0	$2,766.33	$2,766.33	$407.11	$ 963.08	$3,561.19
Medium	0	$2,328.38	$2,328.38	$506.39	$ 928.78	$3,084.04
Hard	0	$4,511.68	$4,511.68	$184.84	$1,026.68	$5,446.49

Note. The tables summarize the data-generating process and results for 1,000 independent Monte Carlo studies. Results reported in Tables 3b and 3c are averages across 1,000 trials. The sample size for each trial is 9,229,585, identical to the field-experiment data analyzed in Section 4. The *Baseline (Hardest)* group's average Gold Coin consumption is obtained from the intercept estimate of (4.1) in Section 4.2. The treatment groups' Gold Coin consumption amounts are obtained by taking the $\hat{\theta}_j$ estimates as the true parameter values. *% Correct* columns are the percentage of the correct decisions out of 1,000 independent trials. *Revenue Lost* columns report the average ex post revenue loss by making an error in decisions, scaled by 20 million underlying daily users and converting to USD. *Scaled Regrets* columns are calculated by scaling the average realized regret $R\left(\hat{\theta};\theta,\hat{\sigma}\right)$ by 20 million underlying users and converting to USD. *Scaled Max. Regrets* columns are calculated by scaling the respective maximum regret by 20 million underlying daily users and converting to USD. Conversion to USD used the 100:1 exchange ratio between Gold Coins and USD.

realized loss from making a mistake in the decision, because θ is the known, true effect size during the simulated data-generating process. The associated actual ex-post relative revenue loss from our minimax-regret decision rule, reported in *Revenue Lost* columns, is only 1/5.4-1/17.0 of the loss from the hypothesis-testing decision rule. Because we run experiments on three different new policies concurrently, among which *Hard* is the best policy, the actual lost revenue from following the hypothesis-testing decision rule would be USD 7,245.00. The results imply our focal company can achieve substantial potential gains by switching to the proposed minimax-regret decision framework.

Finally, we turn to Table 3c, where we compare the scaled regrets and scaled maximum regrets. We provide the comparison to demonstrate the maximum regrets, which can be evaluated without knowing the true parameter value θ, are very informative regarding regrets and ex-post revenue lost. *Scaled Regret* columns are calculated by scaling the average regret $R\left(T=\hat{\theta};\theta,\sigma=\hat{\sigma}\right)$ by 20,000,000 underlying users and converting the unit of regret to USD. Notice the Type I regret is zero because the true effect size θs are all positive, and the scaled Type II regret is smaller when the true effect-size parameter θ is small in magnitude. Next, comparing the *Scaled* $R_{\textit{Type II}}\left(\hat{\theta};\theta,\hat{\sigma}\right)$ column with the *Scaled* $\bar{R}_{\textit{Type II}}\left(\hat{\theta};\hat{\sigma}\right)$ shows they differ little in magnitude – the Type II regrets are 75%-80% of the maximum Type II regrets, thereby suggesting maximum regrets are good proxies for the corresponding regrets.

5.2 Monte Carlo Study 2: Efficient Traffic Allocation and Sample-Size/Wait-Time Reduction

In the field-experiment data we analyzed in Section 4, the company sets the traffic-allocation probability in an ad-hoc manner at 2 (Baseline):1 (Easy):1 (Medium):1 (High) ratio. In this section, we apply the experiment-design scheme of the efficient traffic allocation developed in Section 3 to demonstrate how much wait time could have been reduced if the company had adopted the proposed optimal traffic-allocation scheme.

The Monte Carlo setup is as follows. We leave the data-generating process summarized in Table 3a intact, except for the duration of the pretreatment period and traffic-allocation probabilities, which we now optimize. Following the field-experiment data analyzed in Section 4, we assume the analyst can randomly allocate 271,458 users daily.[31] The goal of the adaptive efficient traffic-allocation algorithm is to attain the max-minimax regret $\bar{R}\left(T^{*};\hat{\sigma}\right)$ across the three policies reported in Table 2 of Section 4.2, which is $\max\{0.00447, 0.00474, 0.00549\} = 0.00549$ ($1,098 when scaled by 20 million users and converted to USD) with the minimal amount of wait time. That is, the analyst monitors $\hat{\sigma}_{j,1}$s across the three treatment arms $j \in \{\text{Easy, Medium, Hard}\}$, with the maximum tolerable daily loss from committing either type of mistake as $1,098. We assume the analyst re-assesses the traffic-allocation probabilities every other day.[32] Because we use the linear-regression

[31] 271,458=9,229,585/34 days, where 34 days were the span of the original field experiment.

[32] The results presented are almost identical to the daily adjustment of the traffic-allocation probabilities.

estimator with covariates (4.1) throughout, η_k^2 follows the formula developed in Appendix D.1.3.

We vary the duration of the pretreatment period across 0, 1, 2, 3, 4, 5, 7, and 14 days, during which none of the samples are treated. Given the daily sample flow-in rate as fixed, a longer pretreatment period would imply the proposed optimal traffic-allocation algorithm will take longer to reach an interior optimum, thereby lowering the efficiency of the proposed optimal traffic-allocation algorithm. For each pretreatment duration we examine, we independently repeat the same data-generating process and evaluation 1,000 times. Lastly, for the benchmark, we also repeat the process with the same stopping max-minimax-regret criteria, using the ad-hoc traffic-allocation probability set by the company at 2 (Baseline):1 (Easy):1 (Medium):1 (High).

Table 4 reports the results, where the numbers reported are the averages across 1,000 independent trials. The *Wait Days* columns report the time taken to stop the experiments from the beginning of the treatment period. Comparing the *Wait Days* columns across *Efficient Allocation* and *Benchmark*, we find the substantial reduction of wait time required to achieve the same level of max-minimax regret and make the decision. Our proposed efficient traffic-allocation procedure can reduce the sample-collection time by more than 30% when it's employed without delay, that is, with no pretreatment sample – the wait time is only 14.1 days under the efficient allocation versus 20.9 days under the benchmark. Although the gain from adopting our efficient traffic-allocation procedure wanes as the pretreatment period increases, due to the increased suboptimality of the sample allocation, it still achieves a reduction in the wait time across all the pretreatment periods examined.

Our next question would be whether the remarkable reduction in wait time is obtained by sacrificing the accuracy of the decision. We find it is not. Comparing the *Minimax Regret % Correct* columns across *Efficient Allocation* and *Benchmark*, the loss of accuracy in the decision is only a few percentage points at most. Therefore, we confirm that continuously monitoring only the standard error of the effect-size estimator throughout the sample-collection stage and controlling for the max-minimax regret works well as the experiment-stopping criteria.

In sum, the efficient traffic-allocation scheme we develop would help the analysts make timely decisions without sacrificing the accuracy of the decisions.

6 Concluding Remarks

We proposed the minimax regret decision framework for online A/B tests, an integrated decision framework of multi-arm experiment evaluation and experiment design. Our minimax decision framework accounts for not only the probability of making an error in a decision, but also the magnitudes of the relevant payoffs. The real-world field-experiment data confirms that the decisions implied by our minimax regret decision criteria can be very different from what is implied by the conventional Neyman-Pearson hypothesis testing. In our Monte Carlo simulations, we show that the decisions implied by our minimax-regret decision criteria are not only different from hypothesis testing, but

Table 4: Monte Carlo 2: Performance of the Minimax-Regret Efficient Traffic-Allocation Algorithm

		Efficient Allocation					Benchmark				
	Pretreat Days	Wait Days	Base N	Treat N	Minimax Regret % Correct	$\alpha = 0.01$ HT % Correct	Wait Days	Base N	Treat N	Minimax Regret % Correct	$\alpha = 0.01$ HT % Correct
Easy				842K	77.8%	9.1%			1,132K	84.9%	8.6%
Medium	0	14.1	1,195K	951K	74.7%	8.0%	20.9	2,537K	1,132K	80.2%	3.6%
Hard				1,109K	92.6%	25.3%			1,133K	93.9%	23.7%
Easy				931K	76.2%	8.8%			1,118K	80.4%	8.0%
Medium	1	13.4	1,226K	926K	71.7%	6.9%	19.6	2,508K	1,118K	78.5%	5.6%
Hard				1,088K	90.7%	21.8%			1,118K	94.5%	22.8%
Easy				1,000K	79.2%	7.9%			1,116K	83.4%	7.3%
Medium	2	14.5	1,329K	1,102K	73.9%	8.1%	19.6	2,775K	1,116K	77.7%	4.3%
Hard				1,325K	91.5%	20.7%			1,116K	93.2%	24.8%
Easy				1,170K	80.5%	7.3%			1,079K	83.8%	6.5%
Medium	3	15.6	1,514K	1,213K	75.9%	6.3%	18.9	2,972K	1,079K	76.6%	5.1%
Hard				1,420K	93.1%	21.1%			1,079K	94.8%	21.7%
Easy				1,238K	82.3%	8.4%			1,098K	84.2%	9.2%
Medium	4	16.4	1,739K	1,296K	75.4%	6.9%	19.2	3,282K	1,098K	81.0%	4.2%
Hard				1,534K	92.6%	23.3%			1,098K	94.2%	24.4%
Easy				1,274K	82.2%	7.8%			1,140K	84.0%	8.6%
Medium	5	17.0	1,992K	1,369K	77.4%	6.9%	20.0	3,638K	1,140K	77.6%	4.7%
Hard				1,618K	92.8%	23.5%			1,140K	93.8%	22.8%
Easy				1,366K	79.8%	9.0%			1,134K	85.6%	8.6%
Medium	7	17.7	2,544K	1,436K	74.4%	6.7%	19.9	4,168K	1,134K	80.8%	5.6%
Hard				1,636K	93.0%	21.8%			1,134K	94.6%	23.3%
Easy				1,383K	80.8%	10.1%			1,102K	85.4%	6.1%
Medium	14	18.7	4,451K	1,538K	77.2%	7.4%	19.3	6,004K	1,102K	78.9%	4.9%
Hard				1,763K	94.3%	23.0%			1,102K	92.8%	21.9%

Note. The table summarizes the data-generating process and results for 1,000 independent Monte Carlo studies. *Efficient Allocation* columns correspond to the results from our efficient traffic-allocation rule, and *Benchmark* columns correspond to the benchmark where the company allocated the post-treatment traffic arbitrarily at the 2:1:1:1 ratio. *Wait Days* columns report the post-treatment duration until the experiments are stopped by meeting max-minimax-regret criteria of 0.00549. *Base N* and *Treat N* columns report the sample sizes including the pretreatment duration, where K is a shorthand for 1,000. *% Correct* columns report the percentages of the correct decisions out of 1,000 trials.

also much more accurate, especially when the magnitudes of the per-unit effect sizes are small. Finally, the efficient traffic-allocation algorithm reduces the wasted time and data remarkably without impairing the accuracy of the decisions.

We close with a caveat. Just as the Neyman-Pearson hypothesis-testing scheme may not be optimal in many contexts, we do not want to claim our minimax-regret decision framework can be or should be used universally without explicitly setting up a decision-maker's objective function and solving for the optimal solution. As such, an interesting direction for future research would consist of developing the optimal decision criteria where the decision environment being considered is different from the A/B test setup that we considered in this paper.

References

Abadie, A. and M. D. Cattaneo (2018): "Econometric methods for program evaluation," *Annual Review of Economics*, 10, 465–503.

Agrawal, R. (1995): "Sample mean based index policies with O(log n) regret for the multi-armed bandit problem," *Advances in Applied Probability*, 27, 1054–1078.

Armitage, P., C. McPherson, and B. Rowe (1969): "Repeated significance tests on accumulating data," *Journal of the Royal Statistical Society Series A*, 132, 235–244.

Ascarza, E. (2018): "Retention futility: Targeting high-risk customers might be ineffective," *Journal of Marketing Research*, LV, 80–98.

Athey, S. and S. Wager (2021): "Policy learning with observational data," *Econometrica*, 89, 133–161.

Berman, R. and C. Van den Bulte (Forthcoming): "False discovery in A/B testing," *Management Science*.

Chick, S. E. and P. Frazier (2012): "Sequential sampling with economics of selection procedures," *Management Science*, 58, 550–569.

Chick, S. E. and N. Gans (2009): "Economic analysis of simulation selection problems," *Management Science*, 55, 421–437.

Chick, S. E. and K. Inoue (2001): "New two-state and sequential procedures for selecting the best simulated system," *Operations Research*, 49, 732–743.

Dehejia, R. H. (2005): "Program evaluation as a decision problem," *Journal of Econometrics*, 125, 141–173.

Deng, A., J. Lu, and S. Chen (2016): "Continuous monitoring of A/B tests without pain: Optional stopping in Bayesian testing," in *IEEE international conference on data science and advanced analytics*, 243–252.

Durrett, R. (2019): *Probability: Theory and Examples*, Cambridge University Press, 5th ed.

Farrell, M. H., T. Liang, and S. Misra (2021): "Deep neural networks for estimation and inference," *Econometrica*, 89, 181–213.

Feit, E. M. and R. Berman (2019): "Test and roll: Profit-maximizing A/B tests," *Marketing Science*, 38, 1038–1058.

HAHN, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 66, 315–331.

HIRANO, K. AND J. R. PORTER (2009): "Asymptotics for statistical treatment rules," *Econometrica*, 77, 1683–1701.

——— (2020): *Asymptotic analysis of statistical decision rules in econometrics*, North Holland, vol. 7A of *Handbook of Econometrics*, chap. 4, 283–354.

HORVITZ, D. G. AND D. J. THOMPSON (1952): "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, 47, 663–685.

JOHARI, R., P. KOOMEN, L. PEKELIS, AND D. WALSH (2017): "Peeking at A/B tests," in *KDD 2017 Conference proceeding*, 1517–1525.

JOHARI, R., L. PEKELIS, AND D. WALSH (Forthcoming): "Always valid inference: Continuous monitoring of A/B tests," *Operations Research*.

KARLIN, S. AND H. RUBIN (1956): "The theory of decision procedures for distributions with monotone likelihood ratio," *Annals of Mathematical Statistics*, 27, 272–299.

KITAGAWA, T. AND A. TETENOV (2018): "Who should be treated? Empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616.

LAI, T. L. AND H. ROBBINS (1985): "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, 6, 4–22.

LEHMANN, E. L. AND J. P. ROMANO (2022): *Testing Statistical Hypotheses*, Springer, 4th ed.

LEWIS, R. A. AND J. M. RAO (2015): "The unfavorable economics of measuring the returns to advertising," *Quarterly Journal of Economics*, 130, 1941–1973.

MANSKI, C. F. (2000): "Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice," *Journal of Econometrics*, 95, 415–442.

——— (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246.

MCPHERSON, C. AND P. ARMITAGE (1971): "Repeated significance tests on accumulating data when the null hypothesis is not true," *Journal of the Royal Statistical Society Series A*, 134, 15–25.

MILLER, A. P. AND K. HOSANAGAR (2020): "A meta-analytic investigation of p-hacking in e-commerce experimentation," *Working Paper*.

Schwartz, E. M., E. T. Bradlow, and P. S. Fader (2017): "Customer acquisition via display advertising using multi-armed bandit experiments," *Marketing Science*, 36, 500–522.

Stoye, J. (2009): "Minimax regret treatment choice with finite samples," *Journal of Econometrics*, 151, 70–81.

Tetenov, A. (2012): "Statistical treatment choice based on asymmetric minimax regret criteria," *Journal of Econometrics*, 166, 157–165.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2nd ed.

Appendix

A Continuity of the Maximum-Regret Functions and the Optimal Threshold T^* in σ

Tetenov (2012)'s Lemma 1 shows $\bar{R}_{\text{Type I}}(T;\sigma)$ and $\bar{R}_{\text{Type II}}(T;\sigma)$ are continuous in T, respectively. Below, we show $\bar{R}_{\text{Type I}}(T;\sigma)$, $\bar{R}_{\text{Type II}}(T;\sigma)$, and T^* are continuous in σ, respectively. The continuity results established by the following Lemma 1 are used in various places throughout the paper.

Lemma 1. *Let $\sigma > 0$.*
(i) $\bar{R}_{\textit{Type I}}(T;\sigma) = \sigma \bar{R}_{\textit{Type I}}\left(\frac{T}{\sigma};1\right)$ and $\bar{R}_{\textit{Type II}}(T;\sigma) = \sigma \bar{R}_{\textit{Type II}}\left(\frac{T}{\sigma};1\right)$.
(ii) $V(\sigma;K_1) = \sigma V(1;K_1)$.
(iii) Suppose $T^*_{\sigma=1}$ solves the equation (2.6) when $\sigma = 1$. Then, for any $\sigma > 0$, $T^* = \sigma T^*_{\sigma=1}$.

Proof. Let $h = \frac{\theta}{\sigma}$. Equation (2.4) in Section 2.1 can be equivalently formulated as

$$\begin{aligned}\bar{R}_{\text{Type I}}(T;\sigma) &= \sigma \max_{h\leq 0}\left\{-h\Phi\left(h-\frac{T}{\sigma}\right)\right\}\\ \bar{R}_{\text{Type II}}(T;\sigma) &= \sigma \max_{h>0}\left\{h\Phi\left(\frac{T}{\sigma}-h\right)\right\}.\end{aligned}$$

From the above, for any $T \in \mathbb{R}$, we have

$$\bar{R}_{\text{Type I}}(T;\sigma) = \sigma \bar{R}_{\text{Type I}}\left(\frac{T}{\sigma};1\right) \tag{A.1}$$

$$\bar{R}_{\text{Type II}}(T;\sigma) = \sigma \bar{R}_{\text{Type II}}\left(\frac{T}{\sigma};1\right), \tag{A.2}$$

which shows (i).

Now, suppose the following holds for a general T^*:

$$V(\sigma;K_1) \equiv K_1 \cdot \bar{R}_{\text{Type I}}(T^*;\sigma) = \bar{R}_{\text{Type II}}(T^*;\sigma).$$

Then, by (A.1)-(A.2), we have

$$V(\sigma;K_1) = \sigma K_1 \cdot \bar{R}_{\text{Type I}}\left(\frac{T^*}{\sigma};1\right) = \sigma \bar{R}_{\text{Type II}}\left(\frac{T^*}{\sigma};1\right) \tag{A.3}$$

$$\equiv \sigma V(1;K_1), \tag{A.4}$$

which shows (ii). For (iii), because $\frac{T^*}{\sigma}$ should be the optimal threshold when $\sigma = 1$, that is, $\frac{T^*}{\sigma} = T^*_{\sigma=1}$, we get the desired conclusion that $T^* = \sigma T^*_{\sigma=1}$. □

B Argument to Exploit Asymptotic Normality of $\hat{\theta}$ and Plug-In of the Variance Estimator

In the usual Neyman-Pearson asymptotic hypothesis-testing framework, an asymptotically Gaussian test statistic $\hat{\theta}_n$ with $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \to_d \mathcal{N}\left(0, s^2\right)$ is compared with the $\mathcal{N}\left(0, \frac{s^2}{n}\right)$'s quantile when the sample size n is "large enough." Put differently, when the sample size is "large enough," the approximate distribution $\mathcal{N}\left(0, \frac{s^2}{n}\right)$ is taken as if it were exact. Furthermore, when $\sigma^2 = \frac{s^2}{n}$ is unknown and needs to be estimated, "plugging in" its consistent estimator $\hat{\sigma}^2$ yields a consistent quantile (see, e.g., Chapter 13 of Lehmann and Romano, 2022). In this section, we provide the necessary consistency results for the regret and maximum-regret functions. We introduce the subscript n in the remainder of the section to emphasize that the estimators depend on the sample size n. We establish the pointwise convergence in probability of $\bar{R}_{\text{Type II}}\left(T; \hat{\sigma}_n\right)$ to $\bar{R}_{\text{Type II}}\left(T; \sigma\right)$ when $\hat{\sigma}_n \to_p \sigma$, in the following. Assume $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \to_d \mathcal{N}\left(0, s^2\right)$ and $\sigma^2 = \frac{s^2}{n}$. Further, let s_n be a consistent estimator for s^2 and define $\hat{\sigma}_n = \frac{s_n}{\sqrt{n}}$. For pointwise convergence in the probability of $\bar{R}_{\text{Type II}}\left(T; \hat{\sigma}_n\right)$ to $\bar{R}_{\text{Type II}}\left(T; \sigma\right)$, it suffices to show that for any sequence $\hat{\sigma}_n \to \sigma$, $\bar{R}_{\text{Type II}}\left(T; \hat{\sigma}_n\right) \to \bar{R}_{\text{Type II}}\left(T; \sigma\right)$.

First, we have $\frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)/\sigma}{s_n/\sigma} \to_d \mathcal{N}\left(0, 1\right)$ because $s_n/\sigma \to_p 1$ and the Slutsky's theorem. That is, $\forall T \in \mathbb{R}$, we have

$$
\begin{aligned}
\Pr\left(\hat{\theta}_n \leq T\right) &= \Pr\left(\hat{\theta}_n - \theta \leq T - \theta\right) \\
&= \Pr\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \leq \frac{T - \theta}{\hat{\sigma}_n}\right) \\
&= \Pr\left(\sqrt{n}\left(\frac{\hat{\theta}_n - \theta}{s_n}\right) \leq \frac{T - \theta}{s_n/\sqrt{n}}\right) \\
&\to \Pr\left(\frac{\hat{\theta}_n - \theta}{\sigma} \leq \frac{T - \theta}{\sigma}\right) \\
&= \Phi\left(\frac{T - \theta}{\sigma}\right).
\end{aligned}
$$

Then, by the Continuous Mapping Theorem, it follows that $\forall \theta > 0$ and $\forall T \in \mathbb{R}$,

$$
R_{\text{Type II}}\left(T; \theta, \hat{\sigma}_n\right) = \theta \Pr\left(\hat{\theta}_n \leq T\right)\big|_{\hat{\sigma}_n} = \theta \Pr\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \leq \frac{T - \theta}{\hat{\sigma}_n}\right)
$$

converges in probability to

$$
R_{\text{Type II}}\left(T; \theta, \sigma\right) = \theta \Pr\left(\hat{\theta} \leq T\right)\big|_{\sigma} = \theta \Phi\left(\frac{T - \theta}{\sigma}\right)
$$

pointwise. Now, the remaining task is to show $\forall T \in \mathbb{R}$, $\forall \sigma > 0$, $\bar{R}_{\text{Type II}}\left(T; \hat{\sigma}_n\right) = \max_{\theta > 0} R_{\text{Type II}}\left(T; \theta, \hat{\sigma}_n\right)$

converges in probability to $\bar{R}_{\text{Type II}}(T;\sigma)$ as $\hat{\sigma}_n \to_p \sigma$. Because Lemma 1 (i) establishes the necessary continuity condition, invoking the Continuous Mapping Theorem again leads to the desired conclusion. The argument for the maximum Type I regret function is similar.

C Minimizing the Expected Regret Under a Symmetric Prior Around Zero

We derived the optimal decision criteria and solved the efficient traffic-allocation problem for two-arm and multi-arm A/B tests. Our approach has been frequentist, in that we did not specify the shape of the true/prior distribution of θ. Absent the deployment costs, and hence $K_1 = 1$, the resulting optimal decision cutoff T^* is zero. In this section, we investigate whether the same cutoff can be derived from the perspective of minimizing the *expected* regret, and what conditions need to be imposed on the prior distribution over θ.[33]

Minimizing the expected regrets with respect to an assumed prior distribution $G(\theta)$ can be of interest from both Bayesian and frequentist perspectives. From the Bayesian perspective with subjective interpretation of the prior, $G(\theta)$ may reflect the possibly biased subjective belief of the analyst. From the frequentist perspective, $G(\theta)$ represents an unknown, true distribution of the treatment effect. Such a frequentist interpretation is suitable in a circumstance where, for example, an analyst runs many different A/B tests concurrently without having prior knowledge about the effects of the new policies. Indeed, in our focal mobile game company studied in Section 4, many different experiments are live concurrently at any given time (hundreds of experiments in a year), across game features, banner advertisements, promotions, and so on.

In this section, we show that if the analyst's prior is symmetric around zero, the decision threshold $T^* = 0$ minimizes the *expected* regret as well. It implies $T^* = 0$ minimizes the expected net opportunity cost of making a decision, and thereby maximizes the expected revenue. Furthermore, we show the conventional hypothesis-testing threshold of 1.645 or 2.326 (95% and 99% Standard Gaussian quantile) can be justified in the revenue-maximizing perspective only if the analyst has a biased prior toward very negative expected gains from experiments.

The argument is as follows. Throughout this subsection, we assume the prior distribution $G(\cdot)$ has a density $g(\cdot)$. As before, we take σ as given, for which a consistent estimate $\hat{\sigma}$ will be "plugged in." Recall the definition of the regret function given by (2.3) in section 2.1:

$$R(T;\theta,\sigma) = \mathbf{1}(\theta \leq 0)\, R_{\text{Type I}}(T;\theta,\sigma) + \mathbf{1}(\theta > 0)\, R_{\text{Type II}}(T;\theta,\sigma).$$

[33] We only focus on $K_1 = 1$ in this subsection because varying K_1 and location-shifting the prior distribution lead to the same effect of shifting the optimal threshold T^*.

The objective function would be

$$\begin{aligned}
&\int R(T;\theta,\sigma)\,g(\theta)\,d\theta \\
&=\int_{-\infty}^{0} R_{\text{Type I}}(T;\theta,\sigma)\,g(\theta)\,d\theta+\int_{0}^{\infty} R_{\text{Type II}}(T;\theta,\sigma)\,g(\theta)\,d\theta. \qquad \text{(C.1)}
\end{aligned}$$

(C.1) is the expected regret, and the expected-regret-minimization problem is as follows:

$$\begin{aligned}
&\min_{T}\left\{\int_{-\infty}^{0} R_{\text{Type I}}(T;\theta,\sigma)\,g(\theta)\,d\theta+\int_{0}^{\infty} R_{\text{Type II}}(T;\theta,\sigma)\,g(\theta)\,d\theta\right\} \\
&=\min_{T}\left\{\int_{-\infty}^{0} -\theta\Phi\left(\frac{\theta-T}{\sigma}\right)g(\theta)\,d\theta+\int_{0}^{\infty}\theta\Phi\left(\frac{T-\theta}{\sigma}\right)g(\theta)\,d\theta\right\}. \qquad \text{(C.2)}
\end{aligned}$$

Under regularity conditions for differentiating under the integral sign,[34] we take the first-order and second-order condition for minimization below in Lemma 2.

Lemma 2. *Assume the regularity conditions for differentiating under integral sign holds. The first-order condition of the problem (C.2) reduces to*

$$\int_{-\infty}^{\infty}\theta\phi\left(\frac{\theta-T^{*}}{\sigma}\right)g(\theta)\,d\theta=0. \qquad \text{(C.3)}$$

Furthermore, if $T^{*}\in\mathbb{R}$ solves (C.3),

$$\frac{\partial}{\partial T}\left[\int_{-\infty}^{\infty}\theta\phi\left(\frac{\theta-T}{\sigma}\right)g(\theta)\,d\theta\right]\Bigg|_{T=T^{*}}>0; \qquad \text{(C.4)}$$

that is, the second-order condition for minimization is also satisfied at T^{*}.

Then, the following Lemma 3 establishes the conditions for the symmetry of the object $\phi\left(\frac{\theta-T}{\sigma}\right)g(\theta)$ in (C.3).

Lemma 3. *A probability density function $g(\theta)$ is symmetric around $T\in\mathbb{R}$ if and only if $h(\theta):=\phi\left(\frac{\theta-T}{\sigma}\right)g(\theta)$ is a symmetric density function around T.*

It follows that if the analyst has no prior information about θ so $g(\theta)$ can be modeled as being symmetric around zero, the optimal cutoff T^{*} must be zero, which is in line with the optimal cutoff derived from the minimax-regret decision problem in section 2.1. The following Proposition 1 provides a formal statement.

Proposition 1. *Consider the expected-regret-minimization problem (C.2). If $g(\theta)$ is a symmetric density function around zero, $T^{*}=0$ is the unique solution for the problem (C.2)*

[34] See, for example, section A.5 of Durrett (2019).

We work out a result for a Gaussian prior below. Interestingly, we find the conventional hypothesis-testing threshold implies a very negative prior.

Corollary 1. *(Gaussian Prior) Let $g(\theta)$ be a Gaussian density with mean μ_0 and variance σ_0^2. Then, it can be shown that*

$$\int_{-\infty}^{\infty} \theta\phi\left(\frac{\theta - T}{\sigma}\right) g(\theta)\, d\theta = \frac{\exp\left(-\frac{(T-\mu_0)^2}{2(\sigma^2+\sigma_0^2)}\right)\left(\mu_0\sigma^2 + \sigma_0^2 T\right)}{\sqrt{2\pi}\sqrt{\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}}\sigma_0\left(\sigma^2+\sigma_0^2\right)}.$$

It follows that the first-order condition (C.3) is satisfied when $\mu_0\sigma^2 + \sigma_0^2 T = 0$, or $T = -\frac{\sigma^2\mu_0^2}{\sigma_0^2}$; that is, the unique decision threshold that minimizes the expected regret is $T^* = -\frac{\sigma^2\mu_0^2}{\sigma_0^2}$.

Notice, to rationalize $T^* = 1.645$ or 2.326, which correspond to α =0.05 or 0.01 in one-sided z-tests, the studentized prior mean $\frac{\mu_0}{\sigma_0}$ has to be negative and large enough.

D Details on the Efficient Traffic-Allocation Problem

D.1 Examples of the Variance Estimators That Have the Form in Assumption 3

In this subsection, we provide several examples of the effect-size estimators under the unconfoundedness assumption and with known propensity scores that have the asymptotic variance of the form (3.2). As is common in online A/B test practices, we assume the analyst assigns the sample to treatments/control without taking the covariates into account. Without losing generality, we denote the control and the treatment groups by subscripts $k = 1, 2$, respectively. Denote $s_k^2(\mathbf{x}) = Var(Y|\mathbf{x}, D_k = 1)$ and $\hat{s}_k^2(\mathbf{x})$ as a $\sqrt{n_k}$-consistent estimator for $s_k^2(\mathbf{x})$. We denote $n = n_1 + n_2$, $\hat{p}_k = \frac{n_k}{n}$ with $\hat{p}_k \to p_k$ as $n \to \infty$, where p_k is the propensity score of being assigned to group k.

D.1.1 Asymptotic Variance for an Efficient Semiparametric Estimator

Hahn (1998) shows the efficient variance bound is

$$\mathbb{E}_{\mathbf{x}_i}\left[\frac{s_1^2(\mathbf{x}_i)}{p_1} + \frac{s_2^2(\mathbf{x}_i)}{p_2} + (\theta(\mathbf{x}_i) - \theta)^2\right]. \tag{D.1}$$

Note $(\theta(\mathbf{x}_i) - \theta)^2 = 0$ when $\theta(\mathbf{x}_i) = \theta$; that is, the treatment effect does not depend on the covariates. Then, the following form of the asymptotic variance estimator,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{s}_1^2(\mathbf{x}_i)}{p_1} + \frac{\hat{s}_2^2(\mathbf{x}_i)}{p_2}\right), \tag{D.2}$$

is consistent for (D.1). Therefore, it satisfies Assumption 3 with $\eta_k^2 = s_k^2$. Hahn (1998) provides an ATE estimator that achieves the lower bound (D.1).

D.1.2 Sampling-Based Frequentist Inference (Linear Regression without Covariates)

Suppose no covariate $\mathbf{x}_i$ exists. Consider the simple estimator constructed using the analogy principle (e.g., Abadie and Cattaneo, 2018):

$$\hat{\theta} = \frac{1}{n_2} \sum_{i \in \{i: i \text{ is treated}\}} Y_i - \frac{1}{n_1} \sum_{i \in \{i: i \text{ is not treated}\}} Y_i.$$

The estimator $\hat{\theta}$ coincides numerically with the OLS coefficient on the indicator variable $d_i = \mathbf{1}(i \text{ is treated})$; that is,

$$y_i = \beta_0 + \theta d_i + \epsilon_i.$$

Assuming the within-group homogeneity but possible across-group heteroskedasticity, the heteroskedasticity-robust variance estimator for $\hat{\theta}$ is

$$nVar\left(\hat{\theta}\right) = \sum_{i=1}^{n} \left(\frac{\hat{s}_1^2}{\hat{p}_1} + \frac{\hat{s}_2^2}{\hat{p}_2} \right) \to_p \frac{s_1^2}{p_1} + \frac{s_2^2}{p_2}.$$

Therefore, it satisfies Assumption 3 with $\eta_k^2 = s_k^2$. A consistent estimator for $s_k^2 = Var\left(Y|D_k = 1\right)$ can be obtained from the sample analogue; that is,

$$\hat{s}_k^2 = \frac{1}{n_k} \sum_{i \in \{i: i \text{ in group } k\}} \left(y_i - \bar{y}_k\right)^2,$$

where $\bar{y}_k$ is y_i's sample mean within group $k \in \{1, 2\}$.

D.1.3 Linear Regression with Covariates

We impose an additional assumption that the treatment effect is constant over $\mathbf{x}_i$ so that the linear regression estimators identify the corresponding average treatment effect. The $\mathbf{x}_i$ may include but is not limited to group and time fixed effects, and $d_i = 1$ only when the observation is actually treated. Now consider the linear regression model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \theta d_i + \epsilon_i.$$

It can be shown that the asymptotic variance does not depend on the annihilator matrix $\mathbf{M}_{\mathbf{x}} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ as long as the rank of $\mathbf{X}$ does not grow with n (see Online Appendix E for the proof) and has the form

$$nVar\left(\hat{\theta}\right) \to_p \frac{s_1^2}{p_1} + \frac{s_2^2}{p_2}. \tag{D.3}$$

Note (D.3) satisfies Assumption 3 with $\eta_k^2 = s_k^2$. A consistent estimator for $s_k^2 = Var\left(y_i | D_k = 1\right)$ can be obtained from the regression residuals:

$$\hat{s}_k^2 = \frac{1}{n_k} \sum_{i \in \{i: i \text{ in group } k\}} \hat{u}_i^2 = \frac{1}{n_k} \sum_{i \in \{i: i \text{ in group } k\}} \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}} - \hat{\theta} d_i\right)^2.$$

D.1.4 Regression-Adjustment Estimators

Wooldridge (2010, pp. 917-918) works out the variance formula for the regression-adjustment estimators. Specifically, consider the following data-generating process:

$$\begin{aligned} Y_{1,i} &= m_1\left(\mathbf{x}_i, \boldsymbol{\beta}_1\right) + \epsilon_{1,i} \\ Y_{2,i} &= m_2\left(\mathbf{x}_i, \boldsymbol{\beta}_2\right) + \epsilon_{2,i}, \end{aligned}$$

with known functional forms of $m_k\left(\mathbf{x}_i, \boldsymbol{\beta}_k\right)$s. Note the simplest form is $m_k\left(\mathbf{x}_i, \boldsymbol{\beta}_j\right) = \mathbf{x}_i' \boldsymbol{\beta}_k$. The average treatment effect is estimated by first estimating $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ at $\left(\sqrt{n_1}, \sqrt{n_2}\right)$ rates, respectively, and then, plugging in $\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right)$:

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{Y}_{2,i}\left(\hat{\boldsymbol{\beta}}_2\right) - \hat{Y}_{1,i}\left(\hat{\boldsymbol{\beta}}_1\right) \right\}.$$

The corresponding asymptotic variance estimator is given by

$$\begin{aligned} & \left[\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_1} m_1\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_1\right)\right]' \widehat{AVar}\left(\hat{\boldsymbol{\beta}}_1\right) \left[\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_1} m_1\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_1\right)\right] \\ + & \left[\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_2} m_2\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_2\right)\right]' \widehat{AVar}\left(\hat{\boldsymbol{\beta}}_2\right) \left[\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_2} m_2\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_2\right)\right]. \end{aligned}$$

Under the usual regularity conditions, $\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_k} m_k\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_k\right) \rightarrow_p \mathbb{E}_{\mathbf{x}_i}\left[\nabla_{\boldsymbol{\beta}_k} m_k\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_k\right)\right]$. Invoking the continuous mapping theorem, the relative convergence rate of the two terms boils down to the respective convergence rates of $\widehat{AVar}\left(\hat{\boldsymbol{\beta}}_1\right)$ and $\widehat{AVar}\left(\hat{\boldsymbol{\beta}}_2\right)$, which are $O_p\left(\frac{1}{n_1}\right)$ and $O_p\left(\frac{1}{n_2}\right)$. Define η_k^2 as the corresponding probability limit as follows:

$$n_k \left[\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_k} m_k\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_k\right)\right]' \widehat{AVar}\left(\hat{\boldsymbol{\beta}}_k\right) \left[\frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}_k} m_k\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_k\right)\right] \rightarrow_p \eta_k^2$$

as $n_k \rightarrow \infty$ with $n_k / n \rightarrow p_k \in (0, 1)$. Then, the regression-adjustment estimator satisfies Assumption 3.

D.2 Simplification of the Efficient Traffic Allocation Problem to (3.3)

We begin by characterizing the optimal traffic-allocation probabilities for treatment arms except for the control arm, that is, for $j = 2, 3, ..., J$.

Lemma 4. *For each of the treatment arms except for the status quo policy, the optimal traffic-allocation probability for the treatment arms are proportional to η_j^2, that is, $p_j^* \propto \eta_j^2$ for $j > 1$.*

Proof. Recall we have $V(\sigma_{j,1}; K_1) = \sigma_{j,1} V(1; K_1)$ from Lemma 1 in Appendix A. Note, for the optimal solution,

$$\sigma_{j,1} V(1; K_1) = \sigma_{k,1} V(1; K_1)$$

must hold because the minimax problem (3.2) achieves the optimal solution when all the arguments are the same. Invoking (3.1) for $j \neq k > 1$ and equating, we get

$$\frac{\eta_1^2}{n_0 + n_1^*} + \frac{\eta_j^2}{n_j^*} = \frac{\eta_1^2}{n_0 + n_1^*} + \frac{\eta_k^2}{n_k^*}$$

for the optimal post-treatment sample size $\left\{n_j^*\right\}_{j=2}^{J}$. Reducing the above and combining with $(n - n_0) p_j^* = n_j^*$, the conclusion follows. □

Derivation of (3.3) Denote $p_0 = \frac{n_0}{n - n_0}$. Invoking Lemma 4 as the constraints with $n_k = (n - n_0) p_k$, the original problem (3.2) can be equivalently formulated as follows:

$$\min_{(p_1, p_j)} \left\{ \frac{\eta_1^2}{p_0 + p_1} + \frac{\eta_j^2}{p_j} \right\} \tag{D.4}$$

$$s.t. \quad \sum_{j=1}^{J} p_j = 1 \tag{D.5}$$

$$\frac{\eta_j^2}{p_j} = \frac{\eta_k^2}{p_k} \quad \text{for } j \neq 1, k \neq 1. \tag{D.6}$$

Take any $2 \leq j \leq J$. Using the equivalent relation (D.6), define $\nu = \frac{p_j}{\eta_j^2}$ for all $j > 1$. The minimization problem can then be reformulated as follows:

$$\min_{p_1 \in [0,1]} \left\{ \frac{\eta_1^2}{p_0 + p_1} + \frac{1}{\nu} \right\}. \tag{D.7}$$

Invoking (D.5), we have

$$\nu \sum_{j=2}^{J} \eta_j^2 = 1 - p_1.$$

Rearranging and substituting back into (D.7) gives

$$\min_{p_1 \in [0,1]} \left\{ \frac{\eta_1^2}{p_0 + p_1} + \frac{\sum_{k=2}^{J} \eta_k^2}{1 - p_1} \right\}. \tag{D.8}$$

Using $\nu \eta_j^2 = p_j^*$ for all $j > 1$ with $1 - p_1 = \sum_{k=2}^{J} p_k = \nu \sum_{k=2}^{J} \eta_k^2$, the relation

$$p_j^* = (1 - p_1^*) \frac{\eta_j^2}{\sum_{k=2}^{J} \eta_k^2} \qquad \forall j > 1$$

follows at the optimum, as desired.

D.3 Illustration of the Solution When $\eta_1^2 = ... = \eta_J^2$

We illustrate the solution when $\eta_1^2 = ... = \eta_J^2$. Solving the problem (D.8) gives

$$p_1^* = \begin{cases} \frac{\left(\sqrt{J-1}-1\right) - p_0\left((J-1) - \sqrt{J-1}\right)}{J-2} & \text{if } J \geq 3 \\ \frac{1-p_0}{2} & \text{if } J = 2 \end{cases}$$
$$p_j^* = \frac{1}{J-1}\left(1 - p_1^*\right). \tag{D.9}$$

If $p_1^* < 0$, implying the problem does not have an interior solution because p_0 is not sufficiently small, set $p_1^* = 0$ and $p_j^* = \frac{1}{J-1}$. Note $p_0 \equiv \frac{n_0}{n-n_0}$ declines to zero as n increases, and the optimal traffic-allocation problem (D.8) will have an interior solution with a large-enough sample.

Appendix for Online Publication

E The Proofs Omitted in Appendices C and D.1.3

E.1 Proofs Omitted in Appendix C

E.1.1 Proof of Lemma 2

Proof. Using $\Phi(x) = 1 - \Phi(-x)$, the problem (C.2) can be rewritten as

$$\min_{T} \left\{ \int_{-\infty}^{0} -\theta \Phi\left(\frac{\theta - T}{\sigma}\right) dG(\theta) + \int_{0}^{\infty} \theta \left\{ 1 - \Phi\left(\frac{\theta - T}{\sigma}\right) \right\} dG(\theta) \right\}.$$

Differentiating w.r.t. T, scaling with σ, and invoking the first-order condition at the optimum T^* yields

$$\begin{aligned} & \int_{-\infty}^{\infty} \theta \phi\left(\frac{\theta - T^*}{\sigma}\right) dG(\theta) = 0 \\ \Rightarrow \quad & \int_{-\infty}^{\infty} \theta \phi\left(\frac{\theta - T^*}{\sigma}\right) g(\theta)\, d\theta = 0, \end{aligned} \tag{E.1}$$

where the second line follows because we assumed $G(\theta)$ has a density with respect to the Lebesgue measure. Expression (E.1) can be rewritten as $\mathbb{E}_h[\theta]\,|_{T=T^*} = 0$, where $\mathbb{E}_h[\cdot]$ denotes the expectation of a random variable against the measure

$$h(x)\, dx = \phi\left(\frac{x - T}{\sigma}\right) g(x)\, dx. \tag{E.2}$$

Note, $h(\cdot)$ is a mixture of Gaussian density with the mixing weight $g(\cdot)$.

For the second-order condition, we have

$$\begin{aligned} \frac{\partial}{\partial T}\left[\int_{-\infty}^{\infty} \theta \phi\left(\frac{\theta - T}{\sigma}\right) g(\theta)\, d\theta\right] &= \int_{-\infty}^{\infty} \theta \phi\left(\frac{\theta - T}{\sigma}\right) g(\theta)\, d\theta \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \theta(\theta - T)\, \phi\left(\frac{\theta - T}{\sigma}\right) g(\theta)\, d\theta. \end{aligned} \tag{E.3}$$

We claim the right-hand side of (E.3) is strictly positive when evaluated at $T = T^*$. To show this, first denote

$$\mathbb{E}_h[\theta(\theta - T)] = \int_{-\infty}^{\infty} \theta(\theta - T)\, h(\theta)\, d\theta,$$

where $h(\cdot)$ is given by (E.2). Then, consider the transformation $\varphi(\theta) = \theta(\theta - T)$. Applying the Jensen's inequality gives

$$\varphi(\mathbb{E}_h[\theta]) = \mathbb{E}_h[\theta](\mathbb{E}_h[\theta] - T) < \mathbb{E}_h[\theta(\theta - T)] = \mathbb{E}_h[\varphi(\theta)],$$

where the inequality is strict because $\varphi(\theta)$ is a strictly convex function in θ and $h(\cdot)$ is a nondegenerate density. When T satisfies the first-order condition, that is, when $T = T^*$, $\mathbb{E}_h[\theta] = 0$ by the first-order condition (E.1). Therefore,

$$\mathbb{E}_h[\theta(\theta - T)]\big|_{T=T^*} > \mathbb{E}_h[\theta](\mathbb{E}_h[\theta] - T)\big|_{T=T^*} = 0 \cdot (-T^*) = 0,$$

as desired. □

E.1.2 Proof of Lemma 3

Proof. Recall the standard Gaussian density $\phi\left(\frac{\theta-T}{\sigma}\right)$ is symmetric around T and its support is the entire real line, so we denote $\phi\left(\frac{\theta-T}{\sigma}\right) = q(\theta)$, where $q(T+\delta) = q(T-\delta)$ holds for any $\delta \in \mathbb{R}$.

($\Rightarrow$) Assume $g(\cdot)$ is symmetric around T. We only need to show the $h(\theta) = q(\theta)g(\theta)$ is a symmetric density around T. Let $\theta = T - \delta$. We have

$$\begin{aligned} h(T-\delta) &= q(T-\delta)g(T-\delta) \\ &= q(T+\delta)g(T+\delta) \\ &= h(T+\delta). \end{aligned}$$

($\Leftarrow$) Assume $h(\cdot)$ is symmetric around T. For the sake of contradiction, suppose a $g(\cdot)$ and a δ exist such that $g(T+\delta) \neq g(T-\delta)$ with $g(T+\delta) > 0$ and $g(T-\delta) > 0$. Take such a $\{g(\cdot), T, \delta\}$. Then,

$$\begin{aligned} h(T-\delta) &= q(T-\delta)g(T-\delta) \\ &= q(T+\delta)g(T-\delta) \\ &\neq q(T+\delta)g(T+\delta) \\ &= h(T+\delta), \end{aligned}$$

which is a contradiction to the assumption that $h(\cdot)$ is symmetric around T. □

E.1.3 Proof of Proposition 1

Proof. $T^* = 0$ being a solution for the problem (C.2) follows from Lemma 2 that $h(\theta) = \phi\left(\frac{\theta-T}{\sigma}\right)g(\theta)$ is a symmetric density around 0 if and only if $T = 0$, and therefore, $T = 0$ satisfies the first-order necessary condition (C.3) and the second-order sufficient condition (C.4).

The proof for uniqueness of T^* is as follows. If $g(\theta)$ is a symmetric density around zero, $h(\theta) := \phi\left(\frac{\theta-T}{\sigma}\right)g(\theta)$ is a symmetric density around zero by Lemma 3, and therefore, $\int_{-\infty}^{\infty} \theta\phi\left(\frac{\theta-T^*}{\sigma}\right)g(\theta)\,d\theta = 0$, where $T^* = 0$ is a solution. Suppose, for the sake of contradiction, that another solution $T_1 \neq 0$

exists such that

$$\int_{-\infty}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta = 0. \tag{E.4}$$

Without loss of generality, let $T_1 > 0$.

Let us first break down (E.4) as

$$\int_{-\infty}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta = \int_{-\infty}^{0} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta + \int_{0}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta. \tag{E.5}$$

Invoking the symmetry of $\phi\left(\frac{\theta-T_1}{\sigma}\right) g(\theta)$ around zero, the absolute value of each term in the right-hand side of (E.5) must be the same; that is,

$$\left|\int_{-\infty}^{0} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta\right| = \int_{0}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta. \tag{E.6}$$

Now observe the left-hand side of (E.6) becomes

$$\begin{aligned}\left|\int_{-\infty}^{0} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta\right| &= \int_{-\infty}^{0} -\theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta \\ &= \int_{0}^{\infty} u\phi\left(\frac{u + T_1}{\sigma}\right) g(u)\, du,\end{aligned}$$

where we substituted $u = -\theta$. Then, the equality (E.6) can be rewritten as

$$\int_{0}^{\infty} \theta\phi\left(\frac{\theta + T_1}{\sigma}\right) g(\theta)\, d\theta = \int_{0}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta. \tag{E.7}$$

We claim the equality (E.7) cannot be true, and in fact, the following strict inequality holds:

$$\int_{0}^{\infty} \theta\phi\left(\frac{\theta + T_1}{\sigma}\right) g(\theta)\, d\theta < \int_{0}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta. \tag{E.8}$$

The reason is that $\phi\left(\frac{\theta+T_1}{\sigma}\right) < \phi\left(\frac{\theta-T_1}{\sigma}\right)$ for any $\theta > 0$. To see why, let $T_1 > 0$. For any $\delta \in (0, T_1]$, we have

$$\phi\left(\frac{\delta + T_1}{\sigma}\right) < \phi\left(\frac{T_1}{\sigma}\right) < \phi\left(\frac{\delta - T_1}{\sigma}\right) = \phi\left(\frac{T_1 - \delta}{\sigma}\right)$$

because the function $\phi(x)$ is decreasing for $x > 0$. For any $\delta > T_1$,

$$\phi\left(\frac{\delta + T_1}{\sigma}\right) < \phi\left(\frac{\delta - T_1}{\sigma}\right)$$

because $\delta + T_1 > \delta - T_1 > 0$. It follows from (E.8) that

$$\left| \int_{-\infty}^{0} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta \right| < \int_{0}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta,$$

and therefore,

$$\int_{-\infty}^{\infty} \theta\phi\left(\frac{\theta - T_1}{\sigma}\right) g(\theta)\, d\theta > 0. \tag{E.9}$$

(E.9) is a contradiction to (E.4). □

E.2 The Proof Omitted in Appendix D.1.3

Consider the linear regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \theta d_i + \epsilon_i.$$

We assume all the conditions for θ to identify the ATE are satisfied. We further assume $\mathbf{x}_i$ includes the column of 1's, $\mathbb{E}[|\mathbf{x}_i|]$ is bounded, and $d_i \in \{0,1\}$. Let $p = \Pr(d_i = 1)$. As usual, we assume $s_i^2 = O(1)$ and is bounded away from zero, that is, $\exists \underline{B}, \bar{B} > 0$ such that $\max_{1 \le i \le n}\left(s_i^2\right) \le \bar{B}$ and $\min_{1 \le i \le n}\left(s_i^2\right) \ge \underline{B}$, where $\underline{B}$ and $\bar{B}$ do not depend on n. Finally, we also assume $(\mathbf{x}_i, d_i)$ is i.i.d. across i. Note we allow the conditional heteroskedasticity so that $Var(\epsilon_i|\mathbf{x}_i, d_i) = s_i^2$.

Denote $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{d}}$ as the demeaned variables of $\mathbf{X}$ and $\mathbf{d}$, respectively. We know the demeaned version gives the identical variance when the error term is homoskedastic (up to degree-of-freedom adjustments), and hence, $\sigma^2 \left(\mathbb{E}_{\mathbf{d}}\left[\mathbf{d}'\mathbf{M}_{\mathbf{x}}\mathbf{d}|\mathbf{X}\right]\right)^{-1} = \sigma^2\left(\mathbb{E}_{\tilde{\mathbf{d}}}\left[\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}|\tilde{\mathbf{X}}\right]\right)^{-1}$, which, in turn, establishes $\mathbf{d}'\mathbf{M}_{\mathbf{x}}\mathbf{d} = \tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$. Note $\tilde{\mathbf{X}}$ does not include the column of 1's.

Our main object of interest, the heteroskedasticity-robust variance estimator, is

$$\begin{aligned} Var\left(\hat{\theta}|\mathbf{X},\mathbf{d}\right) &= \left(\mathbf{d}'\mathbf{M}_{\mathbf{x}}\mathbf{d}\right)^{-1}\mathbf{d}'\mathbf{M}_{\mathbf{x}}\mathbf{V}\mathbf{M}_{\mathbf{x}}\mathbf{d}\left(\mathbf{d}'\mathbf{M}_{\mathbf{x}}\mathbf{d}\right)^{-1} \\ &= \left(\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)^{-1}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\left(\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)^{-1}, \end{aligned} \tag{E.10}$$

where

$$V_{ij} = \begin{cases} s_i^2 & \text{diagonal entries} \\ 0 & \text{otherwise.} \end{cases}$$

Preliminary Results We first establish two preliminary results. First is the characterization of $\tilde{\mathbf{d}}'\tilde{\mathbf{d}}$. Without losing generality, sort $\mathbf{d}$ such that $(1,1,...,1,0,0,...,0)$, where the first n_1 elements are 1 and the remaining $n - n_1$ elements are 0. Then, we have

$$\begin{aligned}
\tilde{\mathbf{d}}'\tilde{\mathbf{d}} &= \sum_{i=1}^{n_1} \left(1 - \frac{n_1}{n}\right)^2 + \sum_{i=n_1+1}^{n} \left(-\frac{n_1}{n}\right)^2 \\
&= n_1 \left(1 - \frac{n_1}{n}\right)^2 + (n - n_1) \left(\frac{n_1}{n}\right)^2 \\
&= n_1 - 2\frac{n_1^2}{n} + n_1 \frac{n_1^2}{n^2} + n\frac{n_1^2}{n^2} - n_1 \frac{n_1^2}{n^2} \\
&= \frac{n^2 n_1 - n n_1^2}{n^2} \\
&= n_1 - \frac{n_1^2}{n}.
\end{aligned} \tag{E.11}$$

Second is the characterization of $\mathbb{E}_{\mathbf{d}} \left[\tilde{\mathbf{d}}\tilde{\mathbf{d}}' | \tilde{\mathbf{X}}\right]$. Note

$$\tilde{d}_i = \begin{cases} 1 - p & \text{with prob. } p \\ -p & \text{with prob. } 1 - p. \end{cases}$$

Therefore, the expectation of the diagonal entries are

$$\begin{aligned}
\mathbb{E}\left[\tilde{d}_i \tilde{d}_i | \tilde{\mathbf{X}}\right] &= (1 - p)^2 p + (-p)^2 (1 - p) \\
&= p (1 - p).
\end{aligned}$$

The expectation of the non-diagonal entries are

$$\begin{aligned}
\mathbb{E}\left[\tilde{d}_i \tilde{d}_j | \tilde{\mathbf{X}}\right] &= ((1 - p) p)^2 + 2 (1 - p) p (-p) (1 - p) + ((-p) (1 - p))^2 \\
&= 0.
\end{aligned}$$

Using this, we obtain

$$\mathbb{E}_{\mathbf{d}} \left[\tilde{\mathbf{d}}\tilde{\mathbf{d}}' | \tilde{\mathbf{X}}\right] = p (1 - p) \mathbf{I}. \tag{E.12}$$

$\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$ Part of (E.10) In this part of the proof, we claim

$$\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} \to_p p (1 - p) \tag{E.13}$$

as $n \to \infty$. We begin by decomposing

$$
\begin{aligned}
\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} &= \tilde{\mathbf{d}}'\left(\mathbf{I}-\mathbf{P}_{\tilde{\mathbf{x}}}\right)\tilde{\mathbf{d}} \\
&= \frac{1}{n}\tilde{\mathbf{d}}'\tilde{\mathbf{d}} - \frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}.
\end{aligned}
$$

First, we have

$$
\begin{aligned}
\mathbb{E}_{\tilde{\mathbf{d}}}\left[\tilde{\mathbf{d}}'\tilde{\mathbf{d}}\right] &= np\left(1-p\right)^2 + n\left(1-p\right)p^2 = np\left(1-p\right) \\
\frac{1}{n}\tilde{\mathbf{d}}'\tilde{\mathbf{d}} &= \frac{1}{n}\sum_{i=1}^{n}\tilde{d}_i^2 \to_p \mathbb{E}_{\tilde{d}}\left[\tilde{d}_i^2\right] = \frac{1}{n}\mathbb{E}_{\tilde{\mathbf{d}}}\left[\tilde{\mathbf{d}}'\tilde{\mathbf{d}}\right] = p\left(1-p\right),
\end{aligned} \tag{E.14}
$$

where we used (E.11) in the second line. Next,

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}|\tilde{\mathbf{X}}\right] &= \mathbb{E}\left[tr\left(\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)|\tilde{\mathbf{X}}\right] \\
&= \mathbb{E}\left[tr\left(\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\tilde{\mathbf{d}}'\right)|\tilde{\mathbf{X}}\right] \\
&= tr\left(\mathbf{P}_{\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{d}}\left[\tilde{\mathbf{d}}\tilde{\mathbf{d}}'|\tilde{\mathbf{X}}\right]\right) \\
&= p\left(1-p\right)tr\left(\mathbf{P}_{\tilde{\mathbf{x}}}\right) \\
&= p\left(1-p\right)r_{\tilde{\mathbf{x}}},
\end{aligned}
$$

where we used (E.12) in the fourth equality. Because $\mathbb{E}\left[\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}|\tilde{\mathbf{X}}\right]$ does not depend on $\tilde{\mathbf{X}}$, $\mathbb{E}\left[\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}|\tilde{\mathbf{X}}\right] = \mathbb{E}\left[\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right] = p\left(1-p\right)r_{\tilde{\mathbf{x}}}$.

Now, we can express $\mathbb{E}\left[\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right] = \mathbb{E}\left[\sum_{i=1}^{n}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)\right] = \sum_{i=1}^{n}\mathbb{E}\left[\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)\right]$. Because we assumed i.i.d. of $(\mathbf{x}_i, d_i)$, we have

$$
\sum_{i=1}^{n}\mathbb{E}\left[\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)\right] = n\mathbb{E}\left[\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)\right].
$$

Then, we finally obtain

$$
\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)\right\} \to_p \mathbb{E}\left[\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)'\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{d}_i\tilde{\mathbf{x}}_i\right)\right] = \frac{1}{n}p\left(1-p\right)r_{\tilde{\mathbf{x}}} = 0 \tag{E.15}
$$

as $n \to \infty$.

$\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$ Part of (E.10) In this part of the proof, we claim

$$\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} \to_p = \frac{1}{n}\left(1-\frac{n_1}{n}\right)^2\sum_{i=1}^{n_1} s_i^2 + \frac{1}{n}\left(\frac{n_1}{n}\right)^2\sum_{i=n_1+1}^{n} s_i^2 \tag{E.16}$$

as $n \to \infty$.

First, decompose $\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$ as follows:

$$\begin{aligned}\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} &= \frac{1}{n}\tilde{\mathbf{d}}'\left(\mathbf{I}-\mathbf{P}_{\tilde{\mathbf{x}}}\right)\mathbf{V}\left(\mathbf{I}-\mathbf{P}_{\tilde{\mathbf{x}}}\right)\tilde{\mathbf{d}} \\ &= \frac{1}{n}\tilde{\mathbf{d}}'\mathbf{V}\tilde{\mathbf{d}} - \frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{V}\tilde{\mathbf{d}} - \frac{1}{n}\tilde{\mathbf{d}}'\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} + \frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}.\end{aligned}$$

Next, we claim $\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{V}\tilde{\mathbf{d}}$, $\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$, and $\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$ are all $o_p(1)$, so the three terms do not matter asymptotically. To see this, because we assumed s_i^2 is bounded, we have

$$\begin{aligned}\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} &= \frac{1}{n}\begin{pmatrix} s_1^2\tilde{d}_1 \\ s_2^2\tilde{d}_2 \\ \vdots \\ s_n^2\tilde{d}_n\end{pmatrix}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} \\ &\leq \frac{B}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} \\ &\to_p 0\end{aligned}$$

as $n \to \infty$ by invoking (E.15). The argument for $\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}$ is similar, because

$$\begin{aligned}\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} &\leq \frac{B}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\mathbf{I}\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} \\ &= \frac{B}{n}\tilde{\mathbf{d}}'\mathbf{P}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}} \\ &\to_p 0\end{aligned}$$

as $n \to \infty$, by recognizing all the entries of $B\mathbf{I}-\mathbf{V}$ are nonnegative, and hence, $B\mathbf{I}-\mathbf{V}$ is positive semidefinite.

The last term is $\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{V}\tilde{\mathbf{d}}$, which turns out to be the only term that matters asymptotically:

$$\begin{aligned}\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{V}\tilde{\mathbf{d}} &= \frac{1}{n}\sum_{i=1}^{n_1} s_i^2\left(1-\frac{n_1}{n}\right)^2 + \frac{1}{n}\sum_{i=n_1+1}^{n} s_i^2\left(-\frac{n_1}{n}\right)^2 \\ &= \frac{1}{n}\left(1-\frac{n_1}{n}\right)^2\sum_{i=1}^{n_1} s_i^2 + \frac{1}{n}\left(\frac{n_1}{n}\right)^2\sum_{i=n_1+1}^{n} s_i^2.\end{aligned}$$

The Expression of the Asymptotic Variance Applying the results (E.13) and (E.16), we have the following:

$$\begin{aligned}
nVar\left(\hat{\theta}|\mathbf{X},\mathbf{d}\right) &= n\left(\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)^{-1}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\left(\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)^{-1} \\
&= \left(\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)^{-1}\left(\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{V}\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)\left(\frac{1}{n}\tilde{\mathbf{d}}'\mathbf{M}_{\tilde{\mathbf{x}}}\tilde{\mathbf{d}}\right)^{-1} \\
&= \frac{\left(\frac{1}{n}\left(1-\frac{n_1}{n}\right)^2\sum_{i=1}^{n_1}s_i^2+\frac{1}{n}\left(\frac{n_1}{n}\right)^2\sum_{i=n_1+1}^{n}s_i^2+o_p(1)\right)}{\left(\frac{1}{n}\tilde{\mathbf{d}}'\tilde{\mathbf{d}}+o_p(1)\right)^2} \\
&= \frac{\left(\frac{1}{n}\left(1-\frac{n_1}{n}\right)^2\sum_{i=1}^{n_1}s_i^2+\frac{1}{n}\left(\frac{n_1}{n}\right)^2\sum_{i=n_1+1}^{n}s_i^2+o_p(1)\right)}{\left(\frac{n_1}{n}\left(1-\frac{n_1}{n}\right)+o_p(1)\right)^2}, \qquad \text{(E.17)}
\end{aligned}$$

where we invoked (E.11) in (E.17). When n is large, $o_p(1)$ terms become negligible, $\frac{n_1}{n}\approx p$, and thus, (E.17) becomes

$$\begin{aligned}
&\frac{\left(p(1-p)^2\left(\frac{1}{n_1}\sum_{i=1}^{n_1}s_i^2\right)+(1-p)p^2\left(\frac{1}{n-n_1}\sum_{i=n_1+1}^{n}s_i^2\right)\right)}{p^2(1-p)^2} \\
=&\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}s_i^2}{p}+\frac{\frac{1}{n-n_1}\sum_{i=n_1+1}^{n}s_i^2}{1-p}. \qquad \text{(E.18)}
\end{aligned}$$

Note the denominators are well defined with a large n and n_1 because s_i^2s are bounded below and above.

Now assume the grouped heteroskedasticity so that $\frac{1}{n_1}\sum_{i=1}^{n_1}s_i^2=s_1^2$ and $\frac{1}{n-n_1}\sum_{i=n_1+1}^{n}s_i^2=s_0^2$. Then, (E.18) becomes

$$nVar\left(\hat{\theta}|\mathbf{X},\mathbf{d}\right)\rightarrow_p\frac{s_1^2}{p}+\frac{s_0^2}{1-p},$$

as desired. The law of iterated expectation then gives the desired result for $nVar\left(\hat{\theta}\right)$.

F Robustness Checks Using Different K_1

Table 5 summarizes the multi-arm-experiment decision results for $K_1 = 3$ and $K_2 = 1$. Our results presented in Section 4 are robust when $K_1 = 3$. In fact, K_1 has to be as large as 11 to stick with the *Baseline (Hardest)* compared with *Medium* with the t-value of 0.8715.[35] Note, unlike when $K_1 = 1$, the minimax-regret optimal threshold T^* differs across columns because different subsamples are involved in calculating T^*.

Table 5: Summary of the Multi-arm Trial Evaluation Results at $K_1 = 3$

Control	Baseline (Hardest)		
Treatment	Easy	Medium	Hard
$\hat{\theta}$	0.02849	0.02430	0.04827
$\hat{\sigma}$	0.02627	0.02788	0.03230
$\hat{\theta}/\hat{\sigma}$	1.08439	0.87158	1.49417
$\bar{R}(T^*;\hat{\sigma}) \equiv V(\hat{\sigma};1)$	0.00447	0.00474	0.00549
$\bar{R}_{\text{Type I}}\left(\hat{\theta};\hat{\sigma}\right)$	0.00074	0.00119	0.00038
$\bar{R}_{\text{Type II}}\left(\hat{\theta};\hat{\sigma}\right)$	0.01432	0.01257	0.02432
Minimax Regret T^*	0.01081	0.01147	0.01329
$\alpha = 0.01$ HT Threshold	0.06112	0.06486	0.07515
Minimax Regret Decision	Accept New	Accept New	Accept New
Hypothesis Test Decision	Reject New	Reject New	Reject New

Note. The minimax-regret decision threshold T^*s are derived using $K_1 = 3$ and $K_2 = 1$. Variance estimators are clustered at the user level to calculate $\hat{\sigma}$.

[35] See Table 9 of Online Appendix H, where $T^* = 0.8519$ when $(K_1 = 10, \hat{\sigma} = 1)$ and $T^* = 0.8861$ when $(K_1 = 11, \hat{\sigma} = 1)$. Comparing $\hat{\theta}/\hat{\sigma} = 0.87158$ with the relation $T^* = \sigma T^*_{\sigma=1}$ given in Lemma 1 in Appendix A yields the conclusion.

G Empirical Application on the Mobile Game Company's Retention Experiments

We analyze another multi-arm experiment run by our focal company. The analysis presented in this section highlights that the minimax-regret-based ordering of the new policies can be different from ordering the policies based on their t-values or p-values.

Another source of the company's revenue besides users' Gold Coin purchase is from advertisers for every ad that has been served (pay per impression). The company was interested in finding whether frequent exposure to advertisements may affect retention of the customers, and therefore, it ran a multi-arm experiment across different ad-cooldown durations. The experiment ran from October 2021 to November 2021, and terminated when a sample size of 10 million had been collected. The treatment status and result for each sample point is recorded at the user-day level. The pre-treatment sample is 2.8 million, and the post-treatment sample is 7.2 million.

The experiment consists of five treatment groups against one control group. The traffic was allocated into the six groups with different probabilities. Each group varies in the duration of advertisement cooldown, as summarized in Table 6. An ad-cooldown duration of c minutes implies that after the user watches an ad, no ad will be shown for the next c minutes. The format of the ad is as follows: when a user fails a level, an icon pops up giving users the option to opt in to watching a short video ad. After the user watches the ad, she will receive a reward that helps her pass the failed level. These ads are called incentivized ads and are the most common ad format in the mobile gaming industry. The company's concern is two-fold: (i) Rewards from watching the ads will make the game too easy for the users, and therefore, users would churn; (ii) even though users opt in to watching ads, the presence of an ad option may negatively affect the user's experience.

The status quo policy, which constitutes our control group, has $c =$1,800 minutes. The treatment groups manipulate the intensity of advertising through varying cooldown durations. We note the size of each experiment group was determined in an ad-hoc manner. The outcome variable of interest is retention. Retention is defined at the user-day level, where $\text{retention}_{it} = 1$ if user i logs in to the game at least once on day t; otherwise, $\text{retention}_{it} = 0$. Although revenue from advertisements is important, we do not consider it in this section, because the company's focus was on user's retention.

Table 6: Description of Treatment and Control Groups

Group		Cooldown Duration	Pretreat n_j	Post-treat n_j	n_j	Traffic Allocation
Control	Baseline	1,800 min	1,911,728	4,915,872	6,827,600	70.25%
	OFF	No ads	410,802	1,056,348	1,467,150	12.75%
	ON1	Zero cooldown	119,868	308,232	428,100	4.25%
Treat	ON2	7.5 min	119,378	306,972	426,350	4.25%
	ON3	90 min	118,188	303,912	422,100	4.25%
	ON4	1,080 min	120,036	308,664	428,700	4.25%

Note. This table provides the description, size, and randomized traffic-allocation probabilities of the control and treatment groups, respectively.

We compare the difference-in-differences estimator $\hat{\theta}$s for the following specification across different treatment groups denoted by j:

$$\text{retention}_{it} = \alpha_{j,\text{Baseline}} \cdot \mathbf{1}\left(\text{treat}_i\right) + \gamma_{j,\text{Baseline}} \cdot \mathbf{1}\left(\text{post\_exp}_t\right) + \theta_{j,\text{Baseline}} \cdot \mathbf{1}\left(\text{treat}_i\right) \cdot \mathbf{1}\left(\text{post\_exp}_t\right) + \epsilon_{it}. \tag{G.1}$$

We run five separate difference-in-differences regressions for each of the treatment groups, obtain the respective $\hat{\theta}_{j,\text{Baseline}}$, and then apply the decision criteria developed in Sections 2.1-2.3.[36] We use the weighting factor $K_1 = K_2 = 1$ for our minimax-regret decision framework.

[36]Because our goal in this section is to analyze the existing multi-arm experiment data, not the optimal traffic allocation, we can use the treatment-arm-specific pretreatment data and a difference-in-differences estimator. We also note $\alpha_{j,\text{Baseline}}$ estimates are insignificant and the results are similar when we do not include the treatment-group fixed effects during estimation.

Table 7: Summary of the Multi-arm Trial Results at $K_1 = 1$

Control	Baseline	Baseline	Baseline	Baseline	Baseline
Treatment	OFF	ON1	ON2	ON3	ON4
$\hat{\theta}$	0.001958	–0.003378	0.003856	–0.000975	0.001684
$\hat{\sigma}$	0.001330	0.002317	0.002334	0.002367	0.002323
$\hat{\theta}/\hat{\sigma}$	1.472259	–1.457737	1.652475	–0.412028	0.725003
$\bar{R}\left(T^*; \hat{\sigma}\right)$	0.000226	0.000394	0.000397	0.000402	0.000395
$\bar{R}_{\text{Type I}}\left(\hat{\theta}; \hat{\sigma}\right)$	0.000016	0.001698	0.000019	0.000667	0.000129
$\bar{R}_{\text{Type II}}\left(\hat{\theta}; \hat{\sigma}\right)$	0.000985	0.000029	0.001964	0.000222	0.000910
Minimax Regret T^*	0	0	0	0	0
$\alpha = 0.01$ HT Threshold	0.003093	0.005391	0.005429	0.005508	0.005403
Minimax -Regret Decision	Accept New	Reject New	Accept New	Reject New	Accept New
Hypothesis-Testing Decision	Reject New	Reject New	Reject New	Reject New	Reject New

Note. The minimax-regret decision threshold T^*s are derived using $K_1 = K_2 = 1$. Variance estimators are clustered at the user level to calculate $\hat{\sigma}$.

Table 7 summarizes the results. Note three of the new policies {*OFF*, *ON2*, *ON4*} exceed the minimax-regret optimal threshold, thereby implying either *OFF*, *ON2*, or *ON4* could be accepted if compared pairwise with the *Baseline*. By contrast, none of the new policies' $\hat{\theta}$ exceed the hypothesis-testing thresholds at the 1% significance level. Among the three policies, *OFF* and *ON2* achieve similar maximum Type I regret at the current $\left(\hat{\theta}, \hat{\sigma}\right)$ estimates, with *OFF* being slightly preferred to *ON2*. Therefore, taking the sample as given fixed, the company should just turn off the advertisements to minimize the maximum possible loss of retention. Note the t-value ($\hat{\theta}/\hat{\sigma} = 1.47226$) of *OFF* is slightly smaller than that of *ON2* (1.652475). Hence, our maximum-regret-based criteria give different preference orderings of the new policies than comparing the t-values or the associated p-values pairwise.

H Tables for Max-Minimax-Regret Level and Minimax-Regret Decision Threshold

Table 8: Max-Minimax Regret Level $V(\sigma; K_1)$ Table with $\sigma = 1$

K_1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.1700	0.1782	0.1860	0.1933	0.2003	0.2069	0.2132	0.2193	0.2251	0.2307
2	0.2361	0.2412	0.2463	0.2511	0.2558	0.2604	0.2648	0.2691	0.2733	0.2774
3	0.2813	0.2852	0.2890	0.2927	0.2963	0.2998	0.3033	0.3066	0.3099	0.3132
4	0.3164	0.3195	0.3225	0.3255	0.3285	0.3314	0.3342	0.3370	0.3398	0.3425
5	0.3451	0.3477	0.3503	0.3528	0.3553	0.3578	0.3602	0.3626	0.3650	0.3673
6	0.3696	0.3719	0.3741	0.3763	0.3785	0.3806	0.3827	0.3848	0.3869	0.3889
7	0.3910	0.3930	0.3949	0.3969	0.3988	0.4007	0.4026	0.4045	0.4063	0.4082
8	0.4100	0.4118	0.4135	0.4153	0.4170	0.4187	0.4204	0.4221	0.4238	0.4254
9	0.4271	0.4287	0.4303	0.4319	0.4335	0.4350	0.4366	0.4381	0.4396	0.4412
K_1	0	1	2	3	4	5	6	7	8	9
10	0.4427	0.4570	0.4702	0.4825	0.4941	0.5049	0.5151	0.5248	0.5340	0.5427
20	0.5511	0.5591	0.5667	0.5741	0.5811	0.5880	0.5945	0.6009	0.6070	0.6129
30	0.6187	0.6243	0.6297	0.6350	0.6401	0.6451	0.6500	0.6547	0.6593	0.6638
40	0.6682	0.6725	0.6767	0.6809	0.6849	0.6888	0.6927	0.6965	0.7002	0.7039
50	0.7074	0.7109	0.7144	0.7178	0.7211	0.7244	0.7276	0.7307	0.7338	0.7369
60	0.7399	0.7429	0.7458	0.7487	0.7515	0.7543	0.7570	0.7598	0.7624	0.7651
70	0.7677	0.7702	0.7728	0.7753	0.7777	0.7802	0.7826	0.7849	0.7873	0.7896
80	0.7919	0.7942	0.7964	0.7986	0.8008	0.8030	0.8051	0.8072	0.8093	0.8114
90	0.8134	0.8154	0.8175	0.8194	0.8214	0.8233	0.8253	0.8272	0.8291	0.8309
$1/K_1$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.1700	0.1620	0.1550	0.1487	0.1430	0.1379	0.1333	0.1290	0.1250	0.1214
2	0.1180	0.1149	0.1119	0.1092	0.1066	0.1042	0.1018	0.0997	0.0976	0.0956
3	0.0938	0.0920	0.0903	0.0887	0.0871	0.0857	0.0842	0.0829	0.0816	0.0803
4	0.0791	0.0779	0.0768	0.0757	0.0747	0.0736	0.0727	0.0717	0.0708	0.0699
5	0.0690	0.0682	0.0674	0.0666	0.0658	0.0651	0.0643	0.0636	0.0629	0.0623
6	0.0616	0.0610	0.0603	0.0597	0.0591	0.0586	0.0580	0.0574	0.0569	0.0564
7	0.0559	0.0553	0.0549	0.0544	0.0539	0.0534	0.0530	0.0525	0.0521	0.0517
8	0.0512	0.0508	0.0504	0.0500	0.0496	0.0493	0.0489	0.0485	0.0482	0.0478
9	0.0475	0.0471	0.0468	0.0464	0.0461	0.0458	0.0455	0.0452	0.0449	0.0446
$1/K_1$	0	1	2	3	4	5	6	7	8	9
10	0.0443	0.0415	0.0392	0.0371	0.0353	0.0337	0.0322	0.0309	0.0297	0.0286
20	0.0276	0.0266	0.0258	0.0250	0.0242	0.0235	0.0229	0.0223	0.0217	0.0211
30	0.0206	0.0201	0.0197	0.0192	0.0188	0.0184	0.0181	0.0177	0.0174	0.0170
40	0.0167	0.0164	0.0161	0.0158	0.0156	0.0153	0.0151	0.0148	0.0146	0.0144
50	0.0141	0.0139	0.0137	0.0135	0.0134	0.0132	0.0130	0.0128	0.0127	0.0125
60	0.0123	0.0122	0.0120	0.0119	0.0117	0.0116	0.0115	0.0113	0.0112	0.0111
70	0.0110	0.0108	0.0107	0.0106	0.0105	0.0104	0.0103	0.0102	0.0101	0.0100
80	0.0099	0.0098	0.0097	0.0096	0.0095	0.0094	0.0094	0.0093	0.0092	0.0091
90	0.0090	0.0090	0.0089	0.0088	0.0087	0.0087	0.0086	0.0085	0.0085	0.0084

Note. This table presents the max-minimax-regret values when $\sigma = 1$. For instance, $V(1; K_1) = 0.6297$ when $K_1 = 32$ and $V(\sigma; K_1) = 0.0544$ when $K_1 = 1/7.3$. For $\sigma \neq 1$, the relation $V(\sigma; K_1) = \sigma V(1; K_1)$ shown in Lemma 1 in Appendix A can be used. The first two blocks are $V(1; K_1)$s when $K_1 \geq 1$; the last two blocks are $V(1; K_1)$s when $0 < K_1 \leq 1$. For the last two blocks, the max-minimax-regret values for $1/K_1$ are calculated.

Table 9: Minimax-Regret Decision Threshold T^* Table with $\sigma = 1$

K_1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.0000	0.0358	0.0685	0.0986	0.1264	0.1523	0.1765	0.1993	0.2207	0.2410
2	0.2602	0.2784	0.2958	0.3124	0.3283	0.3435	0.3582	0.3722	0.3858	0.3988
3	0.4114	0.4236	0.4354	0.4468	0.4579	0.4686	0.4791	0.4892	0.4991	0.5087
4	0.5180	0.5271	0.5360	0.5447	0.5532	0.5615	0.5695	0.5775	0.5852	0.5928
5	0.6002	0.6075	0.6146	0.6216	0.6284	0.6351	0.6417	0.6482	0.6546	0.6608
6	0.6670	0.6730	0.6789	0.6848	0.6905	0.6961	0.7017	0.7072	0.7126	0.7179
7	0.7231	0.7283	0.7333	0.7384	0.7433	0.7482	0.7530	0.7577	0.7624	0.7670
8	0.7715	0.7760	0.7805	0.7848	0.7892	0.7934	0.7977	0.8018	0.8059	0.8100
9	0.8140	0.8180	0.8220	0.8259	0.8297	0.8335	0.8373	0.8410	0.8447	0.8483

K_1	0	1	2	3	4	5	6	7	8	9
10	0.8519	0.8861	0.9171	0.9456	0.9719	0.9963	1.0190	1.0404	1.0604	1.0793
20	1.0972	1.1142	1.1304	1.1458	1.1606	1.1747	1.1882	1.2012	1.2137	1.2257
30	1.2373	1.2486	1.2594	1.2699	1.2801	1.2899	1.2995	1.3088	1.3178	1.3266
40	1.3351	1.3435	1.3516	1.3595	1.3672	1.3748	1.3822	1.3894	1.3964	1.4033
50	1.4100	1.4167	1.4231	1.4295	1.4357	1.4418	1.4478	1.4536	1.4594	1.4651
60	1.4706	1.4761	1.4815	1.4867	1.4919	1.4971	1.5021	1.5070	1.5119	1.5167
70	1.5214	1.5261	1.5307	1.5352	1.5396	1.5440	1.5483	1.5526	1.5568	1.5610
80	1.5651	1.5691	1.5731	1.5771	1.5810	1.5848	1.5886	1.5924	1.5961	1.5997
90	1.6033	1.6069	1.6105	1.6140	1.6174	1.6208	1.6242	1.6276	1.6309	1.6341

$1/K_1$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.0000	–0.0358	–0.0685	–0.0986	–0.1264	–0.1523	–0.1765	–0.1993	–0.2207	–0.2410
2	–0.2602	–0.2784	–0.2958	–0.3124	–0.3283	–0.3435	–0.3582	–0.3722	–0.3858	–0.3988
3	–0.4114	–0.4236	–0.4354	–0.4468	–0.4579	–0.4686	–0.4791	–0.4892	–0.4991	–0.5087
4	–0.5180	–0.5271	–0.5360	–0.5447	–0.5532	–0.5615	–0.5695	–0.5775	–0.5852	–0.5928
5	–0.6002	–0.6075	–0.6146	–0.6216	–0.6284	–0.6351	–0.6417	–0.6482	–0.6546	–0.6608
6	–0.6670	–0.6730	–0.6789	–0.6848	–0.6905	–0.6961	–0.7017	–0.7072	–0.7126	–0.7179
7	–0.7231	–0.7283	–0.7333	–0.7384	–0.7433	–0.7482	–0.7530	–0.7577	–0.7624	–0.7670
8	–0.7715	–0.7760	–0.7805	–0.7848	–0.7892	–0.7934	–0.7977	–0.8018	–0.8059	–0.8100
9	–0.8140	–0.8180	–0.8220	–0.8259	–0.8297	–0.8335	–0.8373	–0.8410	–0.8447	–0.8483

$1/K_1$	0	1	2	3	4	5	6	7	8	9
10	–0.8519	–0.8861	–0.9171	–0.9456	–0.9719	–0.9963	–1.0190	–1.0404	–1.0604	–1.0793
20	–1.0972	–1.1142	–1.1304	–1.1458	–1.1606	–1.1747	–1.1882	–1.2012	–1.2137	–1.2257
30	–1.2373	–1.2486	–1.2594	–1.2699	–1.2801	–1.2899	–1.2995	–1.3088	–1.3178	–1.3266
40	–1.3351	–1.3435	–1.3516	–1.3595	–1.3672	–1.3748	–1.3822	–1.3894	–1.3964	–1.4033
50	–1.4100	–1.4167	–1.4231	–1.4295	–1.4357	–1.4418	–1.4478	–1.4536	–1.4594	–1.4651
60	–1.4706	–1.4761	–1.4815	–1.4867	–1.4919	–1.4971	–1.5021	–1.5070	–1.5119	–1.5167
70	–1.5214	–1.5261	–1.5307	–1.5352	–1.5396	–1.5440	–1.5483	–1.5526	–1.5568	–1.5610
80	–1.5651	–1.5691	–1.5731	–1.5771	–1.5810	–1.5848	–1.5886	–1.5924	–1.5961	–1.5997
90	–1.6033	–1.6069	–1.6105	–1.6140	–1.6174	–1.6208	–1.6242	–1.6276	–1.6309	–1.6341

Note. This table presents the $\hat{\theta}$ thresholds for the minimax-regret decision rule when $\sigma = 1$. This table can also be understood as the cutoff-threshold values for the studentized estimate $\hat{\theta}/\hat{\sigma}$ due to the relation $T^* = \sigma T^*_{\sigma=1}$ shown in Lemma 1 in Appendix A. For instance, $T^* = 1.2594$ when $K_1 = 32$ and $T^* = -0.7384$ when $K_1 = 1/7.3$. The first two blocks are T^*s when $K_1 \geq 1$; the last two blocks are T^*s when $0 < K_1 \leq 1$. For the last two blocks, the threshold values for $1/K_1$ are calculated. The first two and last two blocks are symmetric around zero because setting $K_1 = 1/x$ while fixing $K_2 = 1$ yields the identical threshold to fixing $K_1 = 1$ and setting $K_2 = x$.