Adaptive A/B Testing Under Nonstationary Dynamics Using State-Space Models

Junzhe Shao
University of California, Berkeley

Waverly Wei
University of Southern California

Jingshen Wang
University of California, Berkeley

Abstract

A/B testing is central to evaluating how modifications to products, services, and user experiences impact user outcomes. Yet in practice, experiments rarely occur in stationary environments: seasonality, feature launches, and dynamically evolved user demographics make the underlying treatment effects shift over time. Conventional fixed-allocation designs fail to adapt to this nonstationarity, relying on static treatment allocations that potentially compromise estimation efficiency and lead to inefficient use of experimental resources. Response-adaptive randomization (RAR) design provides a natural alternative, adaptively allocating participants over time based on accrued information. In this work, we propose a methodology framework that addresses these challenges. On the one hand, we model period-level treatment arm means as autoregressive state-space processes and develop a Kalman smoother estimator for the time-averaged treatment effect that exploits temporal dependence. On the other hand, we propose an RAR design that accommodates nonstationarity by incorporating state uncertainty via predicted Kalman variances. Our theoretical analysis establishes asymptotic normality of both a naive and a smoother-based estimator, proves that the smoother strictly dominates the naive estimator in asymptotic variance under correct specification, compares relative efficiency, and enables the construction of anytime-valid confidence sequences for continuous monitoring. Simulation studies demonstrate that our method is significantly more efficient than a benchmark time-averaging estimator and fixed allocation strategy, particularly under treatment effect drift and variance imbalance.

Proceedings of the 29[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

1 INTRODUCTION

Online A/B tests are rarely run in stationary environments. Business cycles, seasonality, product launches, and novelty effects can all make the size of treatment effects shift during a test. In large-scale experimentation, such nonstationary factors can move metrics enough to flip conclusions if not modeled appropriately (Kohavi et al., 2013, 2020; Sadeghi et al., 2021). Conventional A/B designs do not fully exploit the sample in these settings: static variance-minimization strategies like Neyman allocation ignore temporal structure and fixed-horizon analyses become inefficient under continuous monitoring. However, nonstationarity in the context of experimentation like novelty effects that decay over time, often follows a predictable pattern like an AR(1) process. When such temporal structure is present, it represents an *opportunity* rather than merely a challenge: the predictability can be leveraged for more efficient estimation of the treatment effect.

Remark. The notion of nonstationarity in experimentation differs from the classical definition of nonstationarity in time series/stochastic processes because the object that changes over time is different. In time series, nonstationarity refers to instability in the probabilistic properties of the stochastic process itself, such as its mean, variance, or autocovariance structure. In experimentation, by contrast, nonstationarity often refers to systematic time variation in the response to treatment or in the experimental environment. Thus, the concern in experimentation is not necessarily that the data-generating process lacks a stable distributional structure, but rather that the treatment effect or outcome dynamics evolve over time.

An important stream of algorithms in this space comes from the multi-armed bandit literature. Research on nonstationary bandits possibly dates back to the work

of Whittle (1988), and since then a wide range of approaches have been developed. Some methods rely on a variation budget framework, which allows the problem to be approximated by linear bandits when the underlying drift is modest (Besbes et al., 2014, 2019). Others address nonstationarity by down-weighting older observations or restarting after detecting shifts, as in Discounted-UCB and Sliding-Window UCB (Garivier and Moulines, 2008, 2011; Trovo et al., 2020; Fiandri et al., 2025). A further line of work involves change-point aware strategies, which explicitly trigger resets when distributional shifts are detected (Liu et al., 2018; Wei and Srivastava, 2018; Cao et al., 2019). Beyond these examples, the literature on nonstationary bandits is vast and many algorithms and variants exist that cannot be fully covered here (Cheung et al., 2021; Jia et al., 2023; Raj and Kalyani, 2017; Mellor and Shapiro, 2013; Zeng et al., 2025). Each of these approaches comes with theoretical regret guarantees and is designed to be robust under specific assumptions. However, these methods primarily target cumulative reward maximization rather than estimation and inference.

A more structural perspective models treatment effects as latent time series. If arm means evolve according to a state-space process, dynamic inference tools such as the Kalman filter can track the underlying state and guide decisions (Liu et al., 2023; Hong et al., 2020; Trella et al., 2025; Chen et al., 2023).

On the other hand, response-adaptive randomization (RAR) is widely studied in statistics and econometrics with different goals from bandits. RAR designs update allocation as outcomes accrue, aiming to increase the power of testing treatment effects while preserving valid inference and Type I error control (Hu and Hu, 2012; Robertson et al., 2020; Zhang and Rosenberger, 2006; Hahn et al., 2011; Hadad et al., 2021; Pallmann et al., 2018; Kato et al., 2025).

Our work is in the spirit of RAR: we adapt assignment probabilities over time to minimize the asymptotic variance of the final treatment-effect estimator. This contrasts with bandit algorithms that maximize cumulative reward or minimize regret. An important distinction is that our primary design goal is to *maximize power at a fixed type I error level* for a prespecified estimand, rather than to minimize regret or eliminate arms. A direct regret-based comparison is therefore not the appropriate metric for our setting; correct inference and variance minimization of the final estimator are the primary goals. In Wu et al. (2025) a related framework is developed in continuous time, with smooth time trends and timestamps entering directly into the estimators. Their continuous-time formulation is quite different from the classic sequential adaptive A/B testing setting where experiments are naturally indexed in discrete periods (user arrivals, days, weeks) and design/inference are typically phrased in batched frameworks. We focus on the discrete-time batched setting that is standard in sequential A/B testing, and our contribution is to build a discrete-time state-space model for arm means and propose a Kalman-smoother-based response-adaptive Neyman allocation.

Contributions. We develop a methodology for adaptive experimentation in nonstationary environments that unifies state-space modeling, response-adaptive allocation, and valid inference. First, we model period-level arm means as autoregressive state-space processes and derive a Kalman smoother estimator for the time-averaged treatment effect that exploits temporal dependence. Second, we design an RAR scheme that incorporates state uncertainty, via predicted Kalman filter/smoother variances to optimize allocation in real time. Third, we establish asymptotic normality for both the naive and smoother estimators, prove that the smoother strictly dominates the naive estimator in asymptotic variance, and provide a decomposition quantifying the gain from temporal modeling. We also prove that the model-adjusted Neyman allocation uniquely minimizes the smoother asymptotic variance. Fourth, we construct confidence sequences valid under continuous monitoring. Finally, simulations confirm that the smoother estimator combined with adaptive allocation substantially improves efficiency over naive time-averaging and fixed allocation.

2 NOTATION, PROBLEM STATEMENT, AND ASSUMPTIONS

We consider a sequential experiment over periods $t = 1, \ldots, T$. For each period t and binary arm $d \in \{0, 1\}$, let $Y_{it}(d)$ denote unit i's potential outcome if assigned to arm d.At each discrete period t, a batch of n_t units arrives and a fraction $\pi_t(d)$ are assigned to arm d.

Under the Neyman–Rubin potential outcome framework (Splawa-Neyman et al., 1923; Rubin, 1974), the observed outcome is

$$Y_{it} = D_{it}\, Y_{it}(1) + (1 - D_{it})\, Y_{it}(0), \tag{1}$$

where $D_{it} \in \{0, 1\}$ is the assignment indicator. Let $\bar{Y}_{d,t}$ denote the sample mean of outcomes for arm d in period t. The time-varying treatment effect is $\tau_t = \mu_t(1) - \mu_t(0)$, where $\mu_t(d)$ is the period-t mean outcome under arm d.

We model the period means as noisy observations of a

latent state:

$$\bar{Y}_{d,t} = \mu_t(d) + \bar{\varepsilon}_{d,t}, \quad \mathbb{E}[\bar{\varepsilon}_{d,t}] = 0, \tag{2}$$

$$\mathrm{Var}(\bar{\varepsilon}_{d,t}) = r_{d,t} = \frac{\sigma_d^2}{n_t \pi_t(d)}, \tag{3}$$

where $\pi_t(1) = \pi_t \in (0,1)$ is the propensity for arm 1, $\pi_t(0) = 1 - \pi_t$, and n_t is the period sample size. The sample mean $\bar{Y}_{d,t}$ is an unbiased estimator of $\mu_t(d)$ with conditional variance $\sigma_d^2/(n_t\pi_t(d))$. Equation (2) is thus the observation equation of a state-space model: each period provides a noisy snapshot of the latent arm mean, with precision governed by the batch size n_t and the adaptive allocation π_t. The batched formulation allows us to treat each period as a noisy snapshot of a time-varying latent mean and to explicitly model temporal dependence.

Remark. We begin in a no-covariate setting to isolate the time-series aspect; Section 5 discusses extensions to covariate-adjusted estimators.

The latent arm means $\mu_t(d)$ evolve according to a state-space model (Shumway and Stoffer, 2017):

$$\mu_t(d) = f(\mu_{1:t-1}(d)) + \eta_{d,t}, \quad \eta_{d,t} \sim \mathcal{N}(0, q_{d,t}). \tag{4}$$

We focus on linear dynamics; the following examples represent cases of practical interest.

Example 1: AR(1) latent state. The latent states evolve as

$$\mu_t(d) = \beta_d\, \mu_{t-1}(d) + \eta_{d,t}, \quad \eta_{d,t} \sim \mathcal{N}(0, q_d), \quad |\beta_d| < 1. \tag{5}$$

This is similar to a novelty effect where an initial treatment impact decays toward a long-run level. We can allow the innovations $\eta_{0,t}$ and $\eta_{1,t}$ to be correlated, with $\mathrm{Cov}(\eta_{1,t}, \eta_{0,t}) = \sigma_{10}$, and model it as a vector form of the state-space model: $\boldsymbol{\mu}_t = (\mu_t(1), \mu_t(0))^\top$. Then the latent state is:

$$\boldsymbol{\mu}_t = F\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad F = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_0 \end{pmatrix}, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, Q), \tag{6}$$

where

$$Q = \begin{pmatrix} q_1 & \sigma_{10} \\ \sigma_{10} & q_0 \end{pmatrix} \succeq 0. \tag{7}$$

Figure 1 illustrates a realization of the treatment effect trajectory τ_t and its running average $\bar{\tau}_t$ under this model.

Example 2: Seasonal/Periodic latent state. To capture day-of-week, time-of-day, or other periodic patterns with period S, a convenient linear-Gaussian specification is a local (mean-reverting) level plus seasonal component for each arm:

$$\mu_t(d) = \ell_t(d) + s_t(d).$$

mean-reverting dynamics:

$$\ell_t(d) = \beta_d\, \ell_{t-1}(d) + \xi_{d,t}, \quad \xi_{d,t} \sim \mathcal{N}\left(0, q_d^{(\ell)}\right), \quad |\beta_d| < 1. \tag{8}$$

Seasonal dynamics (structural, period S):

$$s_t(d) = -\sum_{j=1}^{S-1} s_{t-j}(d) + \omega_{d,t}, \quad \omega_{d,t} \sim \mathcal{N}\left(0, q_d^{(s)}\right) \tag{9}$$

with the standard identifiability constraint $\sum_{j=0}^{S-1} s_{t-j}(d) = 0$ for all t. Stacking states for both arms yields a linear system still estimable by the Kalman filter. Figure 2 illustrates the treatment effect trajectory under this model.

Target estimand. Our primary target is the running averaged treatment effect (ATE):

$$\bar{\tau}_T = \frac{1}{T}\sum_{t=1}^{T} \big(\mu_t(1) - \mu_t(0)\big). \tag{10}$$

This parameter represents the average treatment effect over the observed time horizon. The asymptotic analysis of this estimand will require the long-run ATE, defined as $\tau_\infty = \lim_{T\to\infty} \bar{\tau}_T$. To simplify discussion, we assume the limit exists. For AR(1) series and other stationary series the limit indeed exists and the concept can be relaxed to consider Cesàro limit to accommodate higher-order slowly moving effect. For totally nonstationary time series like random walk, studying the running average is not practical.

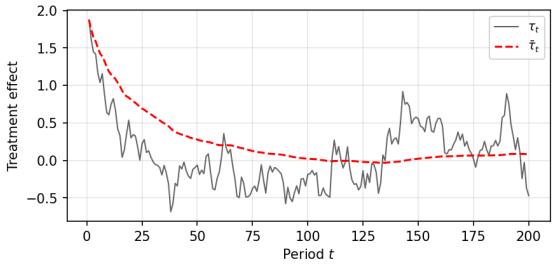

Figure 1: Example 1: AR(1) treatment effect trajectory τ_t and running average $\bar{\tau}_t$.

We now state the key assumptions. Let $(\mathcal{F}_t)_{t\geq 0}$ denote the filtration generated by all observations through period t.

Assumption 2.1 (Consistency / SUTVA)**.** The observed outcome equals the potential outcome under the realized treatment, with no interference across units or periods.

Assumption 2.2 (Adaptive allocation)**.** The propensity π_t and period size n_t are determined by past data: both are functions of $\mathcal{F}_{t-1}$. Conditional on the past, treatment assignment is randomized independently of contemporaneous potential outcomes.

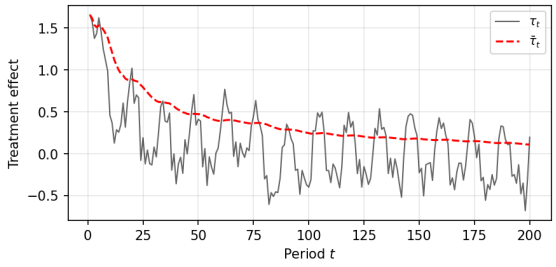

Figure 2: Example 2: AR(1) latent state with seasonal component.

Assumption 2.3 (Positivity)**.** There exists $\varepsilon \in (0, 1/2]$ such that $\varepsilon \le \pi_t \le 1 - \varepsilon$ a.s. for all t.

Assumption 2.4 (Sampling noise)**.** Conditional on $(\mathcal{F}_{t-1}, \mu_t)$, $\bar{Y}_{d,t} = \mu_t(d) + \bar{\varepsilon}_{d,t}$ with mean-zero, finite-variance errors that are independent across arms and of the state innovations.

Assumption 2.5 (State-space model)**.** The latent arm means evolve according to a stable linear-Gaussian state-space model. Innovations are mean-zero with finite covariance, independent of $\mathcal{F}_{t-1}$, and the initial state has finite second moments.

Parametric state-space modeling of time series is well-established in both statistics (Shumway and Stoffer, 2017) and the bandit literature (Liu et al., 2023; Hong et al., 2020). When temporal patterns like novelty effects or seasonality are approximately predictable, the linear-Gaussian specification provides a principled way to borrow strength across periods. These assumptions are mostly practical and necessary to apply the martingale central limit theorem and the Kalman filter recursions.

Assumption 2.6 (Regularity)**.** (i) There exists $\delta > 0$ and $C < \infty$ such that $\mathbb{E}[|\bar{\varepsilon}_{d,t}|^{2+\delta} \mid \mathcal{F}_{t-1}] \le C$ and $\mathbb{E}[|\eta_{d,t}|^{2+\delta}] \le C$ for all t, d. (ii) The state transition is stable ($|\beta_d| < 1$) and $\sup_t \mathbb{E}\|\mu_t\|^{2+\delta} < \infty$. (iii) Period sizes satisfy $0 < n_t < \infty$.

Problem statement. Given Assumptions 2.1–2.6, an *allocation rule* is a predictable process $\{\pi_t\}_{t\ge1}$ with $\pi_t \in [\varepsilon, 1-\varepsilon]$ and π_t $\mathcal{F}_{t-1}$-measurable. Our goals are:

1. *(Design)* Choose $\{\pi_t\}$ to minimize the asymptotic variance of an estimator of the running ATE $\bar{\tau}_T = \frac{1}{T}\sum_{t=1}^{T}(\mu_t(1) - \mu_t(0))$.
2. *(Estimation)* Construct an efficient estimator $\hat{\tau}_T$ that is consistent for $\bar{\tau}_T$.
3. *(Inference)* Provide a consistent estimator of the asymptotic variance and Wald-type confidence intervals that remain valid under response-adaptive allocation.

3 ESTIMATORS

As a starting point, we first need to introduce two classes of estimators for the average treatment effect $\bar{\tau}_T$ as the response adaptive design are related. The first is a simple naive estimator (or difference-in-means) that serves as a natural benchmark without knowledge of any nonstationary structure. The second leverages the state-space structure through Kalman filtering and smoothing to potentially achieve variance reduction.

3.1 Naive Time-Average Estimator

Definition 3.1 (Naive estimator)**.**

$$\hat{\tau}_T^{\text{Naive}} = \frac{1}{T}\sum_{t=1}^{T}\big(\bar{Y}_{1,t} - \bar{Y}_{0,t}\big). \tag{11}$$

This estimator treats each period independently and makes no use of the state-space structure.

3.2 Kalman Filter and Smoother

The Kalman filter (Welch, 1997; Shumway and Stoffer, 2017) provides recursive state estimation for the online phase of the experiment. For each arm d, initialized with prior $(\mu_{d,0|0}, P_{d,0|0})$, the filter iterates:

Predict:

$$\mu_{d,t|t-1} = \beta_d\, \mu_{d,t-1|t-1}, \tag{12}$$

$$P_{d,t|t-1} = \beta_d^2\, P_{d,t-1|t-1} + q_d. \tag{13}$$

Innovation:

$$\nu_{d,t} = \bar{Y}_{d,t} - \mu_{d,t|t-1}. \tag{14}$$

Kalman gain:

$$K_{d,t} = \frac{P_{d,t|t-1}}{P_{d,t|t-1} + r_{d,t}}. \tag{15}$$

Update:

$$\mu_{d,t|t} = \mu_{d,t|t-1} + K_{d,t}\,\nu_{d,t}, \tag{16}$$

$$P_{d,t|t} = (1 - K_{d,t})\,P_{d,t|t-1}. \tag{17}$$

Here $\bar{Y}_{d,t}$ is the observation, $r_{d,t}$ is the observation variance, and q_d is the state evolution variance. The filter is used during the experiment to compute predictive variances $P_{d,t|t-1}$ that guide allocation decisions.

After data collection, the Kalman smoother refines these estimates by incorporating future observations. The backward pass, for $t = T-1, \ldots, 1$, computes:

$$J_{d,t} = P_{d,t|t}\,\beta_d \,/\, P_{d,t+1|t}, \tag{18}$$

$$\mu_{d,t|T} = \mu_{d,t|t} + J_{d,t}\big(\mu_{d,t+1|T} - \mu_{d,t+1|t}\big), \tag{19}$$

$$P_{d,t|T} = P_{d,t|t} + J_{d,t}^2\big(P_{d,t+1|T} - P_{d,t+1|t}\big), \tag{20}$$

with $\mu_{d,T|T}$ and $P_{d,T|T}$ from the forward filter. The smoother MSE satisfies $P_{d,t|T} \le P_{d,t|t}$ for all t.

Definition 3.2 (Kalman smoother ATE estimator)**.**

$$\hat{\tau}_T^{\text{Smooth}} = \frac{1}{T}\sum_{t=1}^{T}\big(\mu_{1,t|T} - \mu_{0,t|T}\big). \tag{21}$$

4 STUDY DESIGN AND STATISTICAL INFERENCE

To apply response-adaptive design with Neyman allocation, once we specify the explicit form of the estimator for our causal estimand, we must establish the asymptotic distribution of the naive running-average treatment effect estimator. We denote the allocation trajectory as the sequence of allocation probabilities π_t:

$$\Pi = \{\pi_t;\, t = 1, \dots, T\}. \tag{22}$$

Theorem 4.1 (Asymptotic distribution of the naive estimator)**.** *Under Assumptions 2.1–2.6,*

$$\sqrt{T}\big(\hat{\tau}_T^{\text{Naive}} - \bar{\tau}_T\big) \xrightarrow{d} \mathcal{N}\big(0,\, V_{\text{Naive}}(\Pi)\big),$$

where

$$V_{\text{Naive}}(\Pi) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sigma_1^2}{n_t\pi_t} + \frac{\sigma_0^2}{n_t(1-\pi_t)}\right). \tag{23}$$

Proof. The error $\bar{\varepsilon}_{1,t} - \bar{\varepsilon}_{0,t}$ forms a martingale difference array with conditional variance $r_{1,t} + r_{0,t}$. The result follows from the martingale CLT under the Lindeberg condition, which is implied by Assumption 2.6. See Appendix B for details. □

Note that this estimator ignores the potential dynamics of the latent state. Then we can apply the results from response adaptive randomization to minimize the asymptotic variance.

4.1 Allocation Strategies

Definition 4.2 (Periodwise Neyman allocation)**.** The classical Neyman allocation minimizes the per-period variance of $\bar{Y}_{1,t} - \bar{Y}_{0,t}$:

$$\pi_t^{\text{Neyman}} = \frac{\hat{\sigma}_{1,t-1}}{\hat{\sigma}_{0,t-1} + \hat{\sigma}_{1,t-1}}, \tag{24}$$

where $\hat{\sigma}_{d,t-1}$ is estimated from all available historical data. In practice σ_0, σ_1 are unobserved and we need to estimate them. Let $n_{d,s} \equiv \sum_{i=1}^{n_s}\mathbf{1}\{D_{is} = d\}$ and

$$\bar{Y}_{d,s} \equiv \frac{1}{n_{d,s}}\sum_{i:D_{i,s}=d} Y_{i,s}. \tag{25}$$

An estimator of the measurement variance σ_d is the pooled within-period standard deviation:

$$\hat{\sigma}_{d,t-1} = \left\{\frac{\sum_{s=1}^{t-1}\sum_{i:D_{i,s}=d}(Y_{i,s} - \bar{Y}_{d,s})^2}{\sum_{s=1}^{t-1}(n_{d,s}-1)}\right\}^{1/2}. \tag{26}$$

Denote $r_{1,t} = \hat{\sigma}_{1,t-1}^2/(n_t\pi_t)$ and $r_{0,t} = \hat{\sigma}_{0,t-1}^2/(n_t(1-\pi_t))$.

To account for temporal structure, we propose an allocation that directly minimizes the smoother asymptotic variance:

Definition 4.3 (Model-adjusted Neyman allocation)**.** The model-adjusted Neyman allocation minimizes $V_{\text{Smooth}}(\pi; \hat{\sigma}_{1,t-1}, \hat{\sigma}_{0,t-1})$ with respect to π:

$$\pi_t^{\text{Smooth}} = \frac{f_1(\pi_t^{\text{Smooth}})\,\hat{\sigma}_{1,t-1}}{f_1(\pi_t^{\text{Smooth}})\,\hat{\sigma}_{1,t-1} + f_0(\pi_t^{\text{Smooth}})\,\hat{\sigma}_{0,t-1}}, \tag{27}$$

where $f_d(\pi) = q_d \,/\, (q_d + r_d(\pi)\,(1-\beta_d)^2)$ is the model information fraction for arm d. This fixed-point equation is solved by iteration from $\pi = 0.5$.

4.2 Asymptotic Distribution of the Smoother Estimator

Theorem 4.4 (Asymptotic distribution of the smoother estimator)**.** *Under Assumptions 2.1–2.6, suppose the Kalman filter reaches steady state. Then*

$$\sqrt{T}\big(\hat{\tau}_T^{\text{Smooth}} - \bar{\tau}_T\big) \xrightarrow{d} \mathcal{N}\big(0,\, V_{\text{Smooth}}(\Pi)\big),$$

where, under constant allocation π with $n_t \equiv n$,

$$V_{\text{Smooth}}(\pi) = \sum_{d\in\{0,1\}}\frac{q_d\, r_d(\pi)}{q_d + r_d(\pi)(1-\beta_d)^2}, \tag{28}$$

with $r_1(\pi) = \sigma_1^2/(n\pi)$ and $r_0(\pi) = \sigma_0^2/(n(1-\pi))$.

Proof sketch. Under the linear-Gaussian model, the smoother average $\hat{\tau}_T^{\text{Smooth}}$ is a linear function of Gaussian observations, hence exactly Gaussian for every T. Its variance equals the mean squared error of the optimal linear estimator of the time average $\bar{\tau}_T$. In steady state, the per-arm contribution converges to the spectral density of the smoothing error at frequency zero, which yields (28). The full derivation is in Appendix B. □

4.3 Efficiency Comparison

Theorem 4.5 (Efficiency and optimal allocation)**.** *Under Assumptions 2.1–2.6:*

(i) For any allocation Π, $V_{\text{Smooth}}(\Pi) \le V_{\text{Naive}}(\Pi)$, with equality if and only if $\beta_d = 1$ or $r_d = 0$ for each arm.

(ii) The Neyman allocation is optimal for the naive estimator: $V_{\text{Naive}}(\Pi^{\text{Neyman}}) \le V_{\text{Naive}}(\Pi)$ *for all* Π.

(iii) The model-adjusted allocation is optimal for the smoother estimator: $V_{\text{Smooth}}(\Pi^{\text{Smooth}}) \le V_{\text{Smooth}}(\Pi)$ *for all* Π. *The minimizer* π^{Smooth} *is unique (by strict convexity of* V_{Smooth}*) and satisfies the fixed-point equation of Definition 4.3.*

Proof. (i) For each arm d, $q_d r_d/(q_d + r_d(1-\beta_d)^2) \le r_d$ simplifies to $r_d(1-\beta_d)^2 \ge 0$. (ii) Strict convexity of $V_{\text{Naive}}(\pi)$. (iii) See Appendix B for the derivation via strict convexity and the first-order condition. □

Remark (Variance decomposition and limiting behavior). A convenient way to understand the smoother variance is through the per-arm decomposition

$$\frac{1}{V_{\text{Smooth},d}} = \frac{1}{r_d} + \frac{(1-\beta_d)^2}{q_d}.$$

The first term is information from direct observations (identical to the semiparametric bound); the second is additional information from the AR(1) temporal constraint. This decomposition clarifies when the smoother provides the most benefit and how the model-adjusted allocation π^{Smooth} differs from classical Neyman:

- When $\beta_d \to 1$ (random walk), the model information $(1-\beta_d)^2/q_d \to 0$ and the smoother approaches the naive estimator. No gain from temporal modeling.
- When $n \to \infty$ (large batches), $r_d \to 0$, the data information dominates, and $\pi^{\text{Smooth}} \to \pi^{\text{Neyman}}$. The model correction vanishes.
- The smoother gain is largest when β_d is small (strong mean-reversion) and $r_d \gg q_d$ (noisy observations with strong temporal signal).

In practice, we clip $\pi_t \in [\varepsilon, 1-\varepsilon]$ to enforce positivity.

4.4 Confidence Sequences

For continuous monitoring, we construct confidence sequences that remain valid at all stopping times.

Definition 4.6 (Confidence sequence)**.** A $(1-\alpha)$ confidence sequence for $\bar{\tau}_t$ is a sequence of intervals $\{I_t\}_{t=1}^{\infty}$ such that $\mathbb{P}(\bar{\tau}_t \in I_t \text{ for all } t \ge 1) \ge 1-\alpha$.

Theorem 4.7 (Asymptotic confidence sequence)**.** *Under Assumptions 2.1–2.6, the intervals*

$$I_t = \left[\hat{\tau}_t \pm \sqrt{\frac{2(\eta^2 \hat{S}_t + 1)}{t^2 \eta^2} \log\left(\frac{\sqrt{\eta^2 \hat{S}_t + 1}}{\alpha}\right)}\right], \quad (29)$$

where $\hat{S}_t = \sum_{j=1}^{t}(P_{1,j|T} + P_{0,j|T})$ *uses the smoother MSE and* $\eta > 0$ *is a tuning parameter, form an asymptotic* $(1-\alpha)$ *confidence sequence. The tuning parameter* η *trades off tightness of the bound at early stopping times versus robustness at longer horizons.*

Proof. See Appendix B. The construction follows the mixture martingale approach of Waudby-Smith et al. (2024), adapted to the smoother variance process. □

5 EXTENSIONS AND DISCUSSIONS

5.1 Unknown Parameters and Misspecification

Beyond the simplified setup used above for theoretical clarity, practical applications must allow for unknown parameters and possible misspecification. A convenient frequentist approach is the EM algorithm (Dempster et al., 1977; Shumway and Stoffer, 2017): in the E-step, expectations of the sufficient statistics are computed using the Kalman smoother, and in the M-step the unknown parameters are updated by maximizing the observed-data likelihood. The same framework extends naturally to richer specifications: all standard linear state-space models—including regression components, exogenous inputs, and seasonal states—are handled with the (standard) Kalman filter/smoother. When the measurement or transition equations are nonlinear, the methodology carries over by replacing the Kalman filter with the extended Kalman filter, which linearized around current state estimates to deliver approximate filtering and parameter updates. These extensions make the approach robust to modest misspecification and broaden its applicability across a wide range of empirical settings. Under mild misspecification, the smoother estimator remains consistent for $\bar{\tau}_T$ because it is a weighted average of the data; misspecification primarily affects efficiency rather than validity, and does not inflate Type I error. See Appendix C for the full algorithm.

5.2 Incorporating Covariates

In the no-covariate setting studied in Sections 3–4, the naive difference-in-means estimator is already semiparametrically efficient: there is no extra $(\tau(X)-\tau)^2$ term in the influence function that could be exploited by AIPW-type estimators (Kato et al., 2025). Our Kalman smoother framework provides additional gains by exploiting temporal structure, which is orthogonal to covariate adjustment. When covariates are available, our framework is compatible with augmented IPW (AIPW/A2IPW) estimators: one can plug covariate-adjusted period-level estimates into the state-space

model and then aggregate across time. We now describe this extension.

Baseline covariate information is important for A/B testing. We can extend the framework to incorporate baseline covariates X_i for each unit i. First consider the case where X_i can be partitioned into $\{\mathcal{X}_j\}_{j=1}^J$. Let $p(\mathcal{X}_j) = P(X_i \in \mathcal{X}_j)$ be the proportion of units with covariates level X_j.

Stationary covariate distributions. If the distribution of covariates across the population is stable over time, i.e., $p_t(X_i \in \mathcal{X}_j) = p(X_i \in \mathcal{X}_j)$ for all t, we can decompose the overall ATE at time t into a weighted sum of subgroup ATE for each subgroup j:

$$\tau_t = \sum_{j=1}^{J} \tau_{tj} \cdot p_j, \tag{30}$$

where $\tau_{tj} = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid X_i \in \mathcal{X}_j]$. In this scenario, the nonstationarity discussion primarily shifts to the CATEs. We can model the potential outcomes for each subgroup j similarly to the overall potential outcomes. The latent conditional means $\mu_{t,j}(d)$ can evolve according to subgroup-specific local-level models:

$$\mu_{t,j}(d) = \beta_j \, \mu_{t-1,j}(d) + \eta_{t,j}(d), \quad d \in \{0, 1\}. \tag{31}$$

The estimation proceeds similarly by running Kalman filters for each subgroup j to estimate $\hat{\mu}_{t,j}(d)$ and thus $\hat{\tau}_{t,j} = \hat{\mu}_{t,j}(1) - \hat{\mu}_{t,j}(0)$. The overall ATE estimator is then:

$$\hat{\tau}_t = \sum_{j=1}^{J} \hat{\tau}_{t,j} \cdot \hat{p}_j. \tag{32}$$

Nonstationary covariate distributions. A more complex scenario arises when the proportion of units belonging to subgroup X_j changes over time, denoted by p_{tj}. The ATE decomposition now explicitly depends on time-varying proportions:

$$\tau_t = \sum_{j=1}^{J} \tau_{tj} \cdot p_{tj}. \tag{33}$$

Here, the overall ATE τ_t can exhibit nonstationarity due to two sources: τ_{tj} can be nonstationary if $\mu_{t,j}(d)$ follows a nonstationary process (e.g., trends, shifts), and the proportions p_{tj} themselves might follow a nonstationary process. The tricky part is that we have finite N at each stage to estimate such process. Furthermore, there might be interactions. The evolution of the latent means $\mu_{t,j}(d)$ could potentially be influenced by the changing composition of the population. For example, the state equation could be modified to include $p_{t,j}$:

$$\mu_{t,j}(d) = f(\mu_{t-1,j}(d), \mathbf{p}_t) + \eta_{t,j}(d), \tag{34}$$

where $\mathbf{p}_t = (p_{t1}, \ldots, p_{tJ})^\top$ of proportions at time t. We have further state equation for $\mathbf{p}_t$:

$$\mathbf{p}_t = g(\mathbf{p}_{1:t-1}) + \boldsymbol{\zeta}_t. \tag{35}$$

The estimation framework remains similar, requiring subgroup-specific Kalman filters. However, the ATE estimator now uses the time-specific proportions: $\hat{\tau}_t = \sum_{j=1}^{J} \hat{\tau}_{t,j} \cdot \hat{p}_{tj}$. Adaptive allocation becomes more complex. Optimizing allocation to minimize the variance of $\hat{\tau}_t$ must account for both the uncertainty in estimating $\tau_{t,j}$ (reflected in $P_{t|t,j}(d)$) and the potentially changing weights $p_{t+1,j}$ using the time series equation. The estimation bias should be considered and final Neyman allocation will have a more complicated form so in practice we usually do not move to this step and can decide by early exploration stage of the design.

6 SIMULATION

We now present a simulation study designed to illustrate the behavior of the allocation strategies and estimators introduced earlier. The simulation parameters are calibrated to mimic the characteristics of a real batched A/B testing dataset from the ASOS digital experiments (available at `https://osf.io/64jsb/`). In particular, Experiment 08bcc2 in that dataset exhibits a clear AR(1)-like novelty effect in the treatment arm, motivating the state-space specification below. Further details on the real data characteristics are provided in the appendix. The data-generating process uses $\beta_0 = \beta_1 = 0.7$, $q_0 = 0.01$, $q_1 = 0.015$, and observation variances $\sigma_0^2 = 1.0$, $\sigma_1^2 = 16.0$. True initial states are $\mu_0(0) = 3.0$ and $\mu_0(1) = 8.0$. Each period has $n = 30$ units and the horizon is $T = 200$ periods, with an initial burn-in of $T_0 = 10$ periods under balanced allocation.

Three allocation strategies are compared: a fixed rule with $\pi_t \equiv 0.5$, the classical Neyman allocation based on variance estimates, and the model-adjusted Neyman allocation that incorporates the predicted uncertainty from the Kalman filter. For each strategy, we compute both the naive estimator of the time-averaged treatment effect and the Kalman smoother estimator. To evaluate uncertainty, we conduct Monte Carlo experiments with repeated draws of the individual-level outcomes, holding the underlying state paths fixed across replications.

The first diagnostic focuses on the realized assignment probabilities over time. Figure 3 shows that while the fixed strategy remains at 0.5, the Neyman and model-adjusted rules adjust allocations dynamically, with the model-adjusted rule reflecting additional sensitivity to predictive state uncertainty.

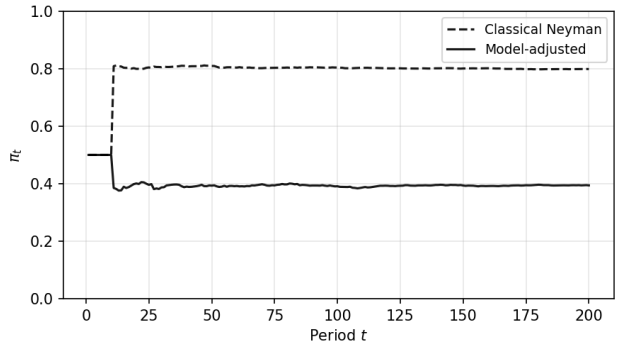

Figure 3: Allocation probability π_t: Classical Neyman (dashed) vs. model-adjusted (solid).

Next, we compare the running averages of the estimated treatment effect with the true underlying average. Figure 4 shows the Neyman allocation and Figure 5 shows the model-adjusted allocation. In both cases, the Kalman smoother estimator (green) tracks the true running average (red) more closely than the naive estimator (blue dashed), which exhibits noticeably larger fluctuations.

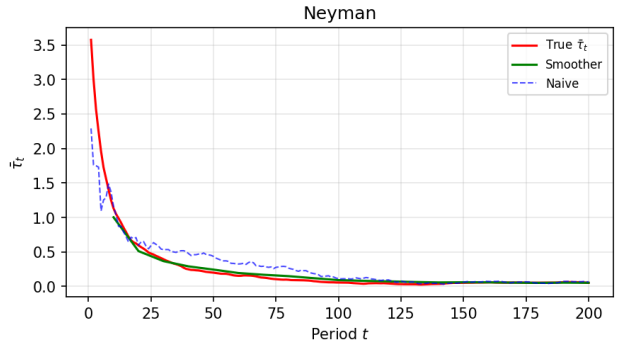

Figure 4: Running average treatment effect under Neyman allocation. Red: true $\bar{\tau}_t$; green: smoother; blue dashed: naive.

Figure 6 summarizes the Monte Carlo standard deviation of the two estimators under both allocation strategies. For each of 500 Monte Carlo replications, we compute the estimation error of the naive running average and the Kalman smoother (evaluated at checkpoints every 10 periods) relative to the true running average $\bar{\tau}_t$. The scaled quantity $\sqrt{t} \times \text{SD}$ is plotted; under standard asymptotics this converges to the square root of the per-period asymptotic variance. Four curves are shown: naive and smoother estimators, each under Neyman (dashed) and model-adjusted (solid) allocation. The smoother (green) consistently achieves lower scaled SD than the naive (blue), confirming the dominance result of Theorem 4.5(i). The model-adjusted allocation (solid) further reduces the smoother's variance compared to Neyman (dashed), consistent with Theorem 4.5(iii). The naive estimator under model-adjusted allocation has slightly higher variance than under Neyman, reflecting the expected trade-off: the model-adjusted allocation sacrifices naive efficiency to optimize the smoother.

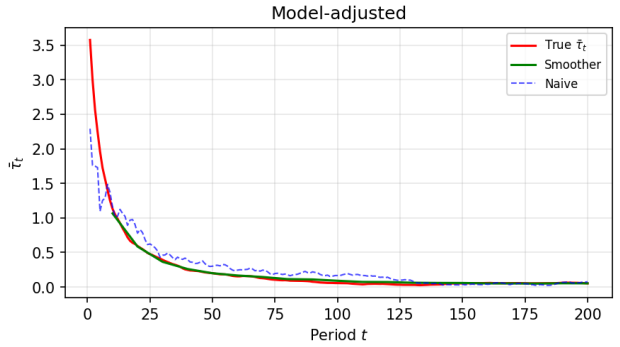

Figure 5: Running average treatment effect under model-adjusted allocation.

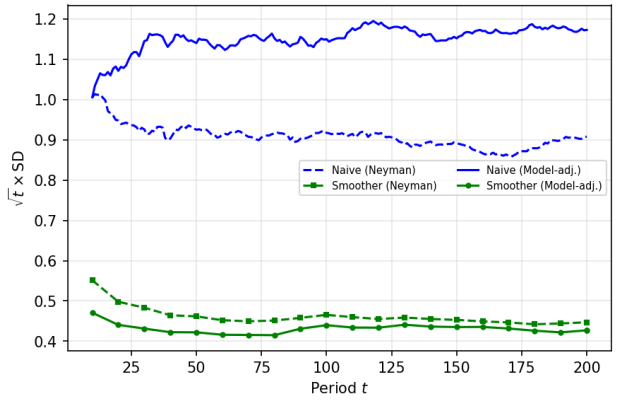

Figure 6: Monte Carlo $\sqrt{t} \times \text{SD}$ of estimators. Blue: naive; green: smoother. Dashed: Neyman allocation; solid: model-adjusted.

Finally, Figure 7 demonstrates the anytime-valid confidence sequence constructed from Theorem 4.7 under the Neyman allocation combined with the smoother estimator. We notice that the confidence interval is monotone with respect to $\hat{S}_t$ so the model-adjusted Neyman Π^{Smooth} also leads to the narrowest always valid CI.

Overall, the simulations show that dynamic allocation strategies combined with Kalman smoothing yield more stable estimates of the treatment effect, with reduced variance compared to naive averaging. Confidence sequences constructed from these estimators remain well-calibrated under continuous monitoring, confirming the practical value of the proposed methodology.

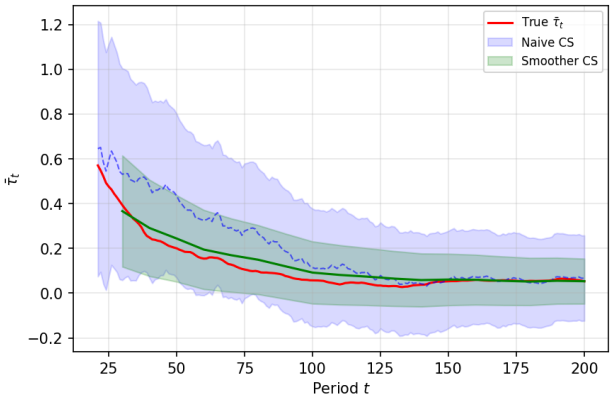

Figure 7: Anytime-valid confidence sequence for the treatment effect.

7 CONCLUSION

This paper has investigated the problem of time-averaged treatment effect estimation in dynamic environments where arm means evolve according to a linear state-space model and observations are subject to both temporal variation and adaptive allocation. We proposed a Kalman-smoother-based estimator that explicitly accounts for temporal dynamics. A key contribution is the model-adjusted Neyman allocation strategy that incorporates smoother variance and admits a unique fixed-point solution that generalizes classical stationary Neyman allocation through the model information fraction.

For inference, we presented asymptotic normal results with efficiency comparisons and developed asymptotic confidence sequences that ensure time-uniform validity under broad conditions. This enables the guidance of efficient design and continuous monitoring without inflating type I error rates, addressing the sequential nature of adaptive experimentation. The framework naturally extends to several different settings including seasonal models, unknown parameters via the EM algorithm, and covariate-stratified designs.

For practical implementation, we envision the following workflow. After an initial exploration phase under a simple allocation (e.g., 50/50 or Neyman), the experimenter can inspect the aggregated time series of $\bar{Y}_{d,t}$: if novelty effects or seasonal patterns are clearly visible and approximately predictable, switching to the state-space-based design can capture that additional structure and improve power. If the trajectory is very irregular or dominated by abrupt shifts, a purely nonparametric design (e.g., time-averaging combined with A2IPW-style estimators) may be preferable. At the same time, if the time series looks completely unpredictable like a random walk, then the overall ATE $\bar{\tau}_T$ may be ill-defined or converge too slowly, so even nonparametric estimators face fundamental limitations. When state uncertainty tracking is clear after an exploration stage, the model-adjusted Neyman allocation provides significant benefits in efficiency. For final reporting, Kalman-smoother-based ATE estimation and confidence sequences ensure valid and efficient uncertainty quantification under adaptivity.

We emphasize that our approach should be viewed as complementary to, rather than a replacement for, fully nonparametric methods. Our contribution is to explore the direction of explicitly utilizing temporal nonstationarity, when it is present and predictable, to potentially reduce variance beyond what model-free methods achieve. When temporal structure is absent or unpredictable, nonparametric approaches remain the appropriate choice. Our method is thus a feasible alternative that can be considered when the experimenter has reason to believe that treatment effects follow a structured temporal pattern.

References

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic Multi-Armed-Bandit Problem with Non-stationary Rewards. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL `https://proceedings.neurips.cc/paper_files/paper/2014/hash/91ba7292e5388b90b58d0b839a7f19ec-Abstract.html`.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Optimal Exploration–Exploitation in a Multi-armed Bandit Problem with Non-stationary Rewards. *Stochastic Systems*, 9(4):319–337, December 2019. ISSN 1946-5238. doi: 10.1287/stsy.2019.0033. URL `https://pubsonline.informs.org/doi/10.1287/stsy.2019.0033`.

Yang Cao, Zheng Wen, Branislav Kveton, and Yao Xie. Nearly Optimal Adaptive Procedure with Change Detection for Piecewise-Stationary Bandit. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 418–427. PMLR, April 2019. URL `https://proceedings.mlr.press/v89/cao19a.html`.

Qinyi Chen, Negin Golrezaei, and Djallel Bouneffouf. Non-Stationary Bandits with Auto-Regressive Temporal Dependency, December 2023. URL `http://arxiv.org/abs/2210.16386`. arXiv:2210.16386 [cs] version: 3.

Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to Optimize under Non-Stationarity, July 2021. URL `http://arxiv.org/abs/1810.03024`. arXiv:1810.03024 [cs].

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 0035-9246. URL https://www.jstor.org/stable/2984875.

Marco Fiandri, Alberto Maria Metelli, and Francesco Trovò. Sliding-Window Thompson Sampling for Non-Stationary Settings, June 2025. URL http://arxiv.org/abs/2409.05181. arXiv:2409.05181 [stat].

Aurélien Garivier and Eric Moulines. On Upper-Confidence Bound Policies for Non-Stationary Bandit Problems, May 2008. URL http://arxiv.org/abs/0805.3415. arXiv:0805.3415 [math].

Aurélien Garivier and Eric Moulines. On Upper-Confidence Bound Policies for Switching Bandit Problems. In Jyrki Kivinen, Csaba Szepesvári, Esko Ukkonen, and Thomas Zeugmann, editors, *Algorithmic Learning Theory*, pages 174–188, Berlin, Heidelberg, 2011. Springer. ISBN 978-3-642-24412-4. doi: 10.1007/978-3-642-24412-4_16.

Vitor Hadad, David A. Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences*, 118(15):e2014602118, April 2021. doi: 10.1073/pnas.2014602118. URL https://www.pnas.org/doi/10.1073/pnas.2014602118.

Jinyong Hahn, Keisuke Hirano, and Dean Karlan. Adaptive Experimental Design Using the Propensity Score. *Journal of Business & Economic Statistics*, 29(1):96–108, January 2011. ISSN 0735-0015. doi: 10.1198/jbes.2009.08161. URL https://doi.org/10.1198/jbes.2009.08161. _eprint: https://doi.org/10.1198/jbes.2009.08161.

Peter Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980. ISBN 978-0-12-319350-6. Google-Books-ID: xxbvAAAAMAAJ.

Joey Hong, Branislav Kveton, Manzil Zaheer, Yinlam Chow, Amr Ahmed, Mohammad Ghavamzadeh, and Craig Boutilier. Non-Stationary Latent Bandits, December 2020. URL http://arxiv.org/abs/2012.00386. arXiv:2012.00386 [cs].

Yanqing Hu and Feifang Hu. Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics*, 40(3):1794–1815, June 2012. ISSN 0090-5364, 2168-8966. doi: 10.1214/12-AOS983. URL https://projecteuclid.org/journals/annals-of-statistics/volume-40/issue-3/Asymptotic-properties-of-covariate-adaptive-randomization/10.1214/12-AOS983.full.

Su Jia, Qian Xie, Nathan Kallus, and Peter I. Frazier. Smooth Non-stationary Bandits. In *Proceedings of the 40th International Conference on Machine Learning*, pages 14930–14944. PMLR, July 2023. URL https://proceedings.mlr.press/v202/jia23c.html.

Masahiro Kato, Takuya Ishihara, Junya Honda, and Yusuke Narita. Efficient Adaptive Experimental Design for Average Treatment Effect Estimation, February 2025. URL http://arxiv.org/abs/2002.05308. arXiv:2002.05308 [stat].

Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1168–1176, Chicago Illinois USA, August 2013. ACM. ISBN 978-1-4503-2174-7. doi: 10.1145/2487575.2488217. URL https://dl.acm.org/doi/10.1145/2487575.2488217.

Ron Kohavi, Diane Tang, and Ya Xu. *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press, April 2020. ISBN 978-1-108-72426-5. Google-Books-ID: TFjPDwAAQBAJ.

Fang Liu, Joohyun Lee, and Ness Shroff. A change-detection based framework for piecewise-stationary multi-armed bandit problem. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18, pages 3651–3658, New Orleans, Louisiana, USA, February 2018. AAAI Press. ISBN 978-1-57735-800-8.

Yueyang Liu, Benjamin Van Roy, and Kuang Xu. Nonstationary Bandit Learning via Predictive Sampling. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, pages 6215–6244. PMLR, April 2023. URL https://proceedings.mlr.press/v206/liu23e.html.

Joseph Mellor and Jonathan Shapiro. Thompson Sampling in Switching Environments with Bayesian Online Change Point Detection, February 2013. URL http://arxiv.org/abs/1302.3721. arXiv:1302.3721 [cs].

Philip Pallmann, Alun W. Bedding, Babak Choodari-Oskooei, Munyaradzi Dimairo, Laura Flight, Lisa V. Hampson, Jane Holmes, Adrian P. Mander, Lang'o Odondi, Matthew R. Sydes, Sofía S. Villar, James M. S. Wason, Christopher J. Weir, Graham M. Wheeler, Christina Yap, and Thomas Jaki. Adaptive designs in clinical trials: why use them, and how to run and report them.

BMC Medicine, 16(1):29, December 2018. ISSN 1741-7015. doi: 10.1186/s12916-018-1017-7. URL https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1017-7.

Vishnu Raj and Sheetal Kalyani. Taming Non-stationary Bandits: A Bayesian Approach, July 2017. URL http://arxiv.org/abs/1707.09727. arXiv:1707.09727 [stat].

David S. Robertson, Kim May Lee, Boryana C. Lopez-Kolkovska, and Sofia S. Villar. Response-adaptive randomization in clinical trials: from myths to practical considerations, May 2020. URL http://arxiv.org/abs/2005.00564. arXiv:2005.00564 [stat] version: 1.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, October 1974. ISSN 1939-2176, 0022-0663. doi: 10.1037/h0037350. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/h0037350.

Soheil Sadeghi, Somit Gupta, Stefan Gramatovici, Jiannan Lu, Hao Ai, and Ruhan Zhang. Novelty and Primacy: A Long-Term Estimator for Online Experiments, February 2021. URL http://arxiv.org/abs/2102.12893. arXiv:2102.12893 [cs].

Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Springer International Publishing, Cham, 2017. ISBN 978-3-319-52451-1 978-3-319-52452-8. doi: 10.1007/978-3-319-52452-8. URL http://link.springer.com/10.1007/978-3-319-52452-8.

Jerzy Splawa-Neyman, D. M. Dabrowska, and T. P. Speed. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472, 1923. ISSN 0883-4237, 2168-8745. doi: 10.1214/ss/1177012031. URL https://projecteuclid.org/journals/statistical-science/volume-5/issue-4/On-the-Application-of-Probability-Theory-to-Agricultural-Experiments-Essay/10.1214/ss/1177012031.full.

Anna L. Trella, Walter Dempsey, Asim H. Gazi, Ziping Xu, Finale Doshi-Velez, and Susan A. Murphy. Non-Stationary Latent Auto-Regressive Bandits, February 2025. URL http://arxiv.org/abs/2402.03110. arXiv:2402.03110 [cs] version: 3 TLDR: Latent AR LinUCB (LARL), an online linear contextual bandit algorithm that does not rely on the non-stationary budget, but instead forms good predictions of reward means by implicitly predicting the latent state, is developed.

Francesco Trovo, Stefano Paladino, Marcello Restelli, and Nicola Gatti. Sliding-window thompson sampling for non-stationary settings. *Journal of Artificial Intelligence Research*, 68:311–364, 2020. URL https://www.jair.org/index.php/jair/article/view/11407.

Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H. Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences, March 2024. URL http://arxiv.org/abs/2103.06476. arXiv:2103.06476 [math].

Lai Wei and Vaibhav Srivastava. On Abruptly-Changing and Slowly-Varying Multiarmed Bandit Problems, April 2018. URL http://arxiv.org/abs/1802.08380. arXiv:1802.08380 [stat].

Greg Welch. An Introduction to the Kalman Filter. 1997.

P. Whittle. Restless bandits: activity allocation in a changing world. *Journal of Applied Probability*, 25(A):287–298, January 1988. ISSN 0021-9002, 1475-6072. doi: 10.2307/3214163. URL https://www.cambridge.org/core/journals/journal-of-applied-probability/article/abs/restless-bandits-activity-allocation-in-a-changing-world/DDEB5E22AFFEFF50AA97ADC96B71AE35.

Yuhang Wu, Zeyu Zheng, Guangyu Zhang, Zuohua Zhang, and Chu Wang. Nonstationary A/B Tests: Optimal Variance Reduction, Bias Correction, and Valid Inference. *Management Science*, 71(6):4707–4727, June 2025. ISSN 0025-1909. doi: 10.1287/mnsc.2022.01205. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2022.01205.

Sihan Zeng, Sujay Bhatt, Alec Koppel, and Sumitra Ganesh. Partially Observable Contextual Bandits With Linear Payoffs. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, April 2025. doi: 10.1109/ICASSP49660.2025.10890035. URL https://ieeexplore.ieee.org/document/10890035. ISSN: 2379-190X.

Lanju Zhang and William F. Rosenberger. Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics*, 62(2):562–569, June 2006. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2005.00496.x.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]

(b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]

(c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes/No/Not Applicable]

2. For any theoretical claim, check if you include:

(a) Statements of the full set of assumptions of all theoretical results. [Yes]

(b) Complete proofs of all theoretical results. [Yes]

(c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

(a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]

(b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]

(c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

(a) Citations of the creator If your work uses existing assets. [/Not Applicable]

(b) The license information of the assets, if applicable. [Not Applicable]

(c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

(d) Information about consent from data providers/curators. [Not Applicable]

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Materials

A USEFUL LEMMAS

Lemma A.1 (Martingale CLT (Hall and Heyde, 1980)). *Let $\{(X_{n,k}, \mathcal{F}_{n,k}) : 1 \le k \le k_n,\ n \ge 1\}$ be a triangular array of square-integrable martingale differences. Define $S_n = \sum_{k=1}^{k_n} X_{n,k}$ and $V_n^2 = \sum_{k=1}^{k_n} \mathbb{E}[X_{n,k}^2 \mid \mathcal{F}_{n,k-1}]$. If*

1. *$V_n^2 \xrightarrow{\mathbb{P}} \sigma^2 \in (0, \infty)$; and*
2. *for every $\varepsilon > 0$, $\sum_{k=1}^{k_n} \mathbb{E}[X_{n,k}^2\, \mathbf{1}\{|X_{n,k}| > \varepsilon\} \mid \mathcal{F}_{n,k-1}] \xrightarrow{\mathbb{P}} 0$,*

then $S_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

B DETAILED PROOFS

Throughout this section, the Lindeberg condition is implied by the bounded Lyapunov condition in Assumption 2.6.

B.1 Proof of Theorem 4.1

Proof. Define $\Delta_t = \bar{\varepsilon}_{1,t} - \bar{\varepsilon}_{0,t}$, so that $\sqrt{T}(\hat{\tau}_T^{\text{Naive}} - \bar{\tau}_T) = T^{-1/2} \sum_{t=1}^{T} \Delta_t$. Since π_t is $\mathcal{F}_{t-1}$-measurable and sampling errors are conditionally mean-zero:

$$\mathbb{E}[\Delta_t \mid \mathcal{F}_{t-1}] = 0,$$

making $\{\Delta_t, \mathcal{F}_t\}$ a martingale difference array. The conditional variance is

$$\mathbb{E}[\Delta_t^2 \mid \mathcal{F}_{t-1}] = r_{1,t} + r_{0,t} = \frac{\sigma_1^2}{n_t \pi_t} + \frac{\sigma_0^2}{n_t(1-\pi_t)},$$

using independence across arms. Under Assumption 2.6, bounded $(2+\delta)$ moments yield the Lindeberg condition:

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\big[\Delta_t^2\, \mathbf{1}\{|\Delta_t| > \epsilon \sqrt{T}\} \mid \mathcal{F}_{t-1}\big] \to 0$$

for every $\epsilon > 0$. The martingale CLT (Lemma A.1) gives the result. □

B.2 Proof of Theorem 4.4

Proof. We establish the asymptotic variance of the Kalman smoother estimator $\hat{\tau}_T^{\text{Smooth}} = T^{-1} \sum_{t=1}^{T} (\mu_{1,t|T} - \mu_{0,t|T})$.

Step 1: Gaussianity. Under the linear-Gaussian state-space model, the smoother estimates $\mu_{d,t|T} = \mathbb{E}[\mu_t(d) \mid Y_{1:T}]$ are linear functions of the observations $\{Y_{1:T}\}$. Therefore $\hat{\tau}_T^{\text{Smooth}}$ is exactly Gaussian for every finite T, and it suffices to compute its variance.

Step 2: Unbiasedness. By the tower property, $\mathbb{E}[\mu_{d,t|T}] = \mathbb{E}[\mathbb{E}[\mu_t(d) \mid Y_{1:T}]] = \mathbb{E}[\mu_t(d)]$, so $\mathbb{E}[\hat{\tau}_T^{\text{Smooth}}] = \bar{\tau}_T$.

Step 3: Variance computation. Define the smoothing error $e_{d,t|T} = \mu_{d,t|T} - \mu_t(d)$. Since the two arms have independent state dynamics conditional on the allocation:

$$\text{Var}\big(\hat{\tau}_T^{\text{Smooth}} - \bar{\tau}_T\big) = \frac{1}{T^2} \sum_{d \in \{0,1\}} \sum_{s,t=1}^{T} \text{Cov}(e_{d,s|T},\, e_{d,t|T}).$$

The inner double sum equals $\mathbf{1}^\top \Sigma_{d|T} \mathbf{1}$, where $\Sigma_{d|T}$ is the $T \times T$ posterior covariance matrix of $(\mu_1(d), \ldots, \mu_T(d))$ given $Y_{1:T}$.

Step 4: Spectral limit. Under constant allocation π and after the Kalman filter reaches steady state (guaranteed by $|\beta_d| < 1$), the observation and state processes are jointly stationary. The spectral density of the signal $\mu_t(d)$ is

$$S_\mu(\omega) = \frac{q_d}{|1 - \beta_d\, e^{-i\omega}|^2},$$

and the spectral density of the smoothing error is

$$S_e^{\text{smooth}}(\omega) = \frac{S_\mu(\omega) \cdot r_d}{S_\mu(\omega) + r_d}.$$

The long-run variance per arm is

$$\lim_{T\to\infty} T \cdot \text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} e_{d,t|T}\right) = S_e^{\text{smooth}}(0) = \frac{q_d\, r_d}{q_d + r_d(1-\beta_d)^2},$$

where at $\omega = 0$: $S_\mu(0) = q_d/(1-\beta_d)^2$ and $S_e^{\text{smooth}}(0) = q_d(1-\beta_d)^{-2} \cdot r_d \big/ \big(q_d(1-\beta_d)^{-2} + r_d\big) = q_d r_d \big/ \big(q_d + r_d(1-\beta_d)^2\big)$.

Summing over both arms yields $V_{\text{Smooth}}(\pi)$ as stated. □

B.3 Proof of Theorem 4.5

Proof. **(i) Smoother dominates naive.** For each arm d, we need

$$\frac{q_d\, r_d}{q_d + r_d(1-\beta_d)^2} \le r_d.$$

This simplifies to $q_d \le q_d + r_d(1-\beta_d)^2$, i.e., $r_d(1-\beta_d)^2 \ge 0$. This always holds, with equality iff $\beta_d = 1$ or $r_d = 0$.

(ii) Neyman optimality. Fix t and consider $V_t(\pi) = \sigma_1^2/(n_t\pi) + \sigma_0^2/(n_t(1-\pi))$. Since $V_t''(\pi) = 2\sigma_1^2/(n_t\pi^3) + 2\sigma_0^2/(n_t(1-\pi)^3) > 0$, the function is strictly convex. The first-order condition yields $\pi^* = \sigma_1/(\sigma_0 + \sigma_1)$, the Neyman allocation. This minimizes V_{Naive} period by period and hence in the time average. □

B.4 Derivation of the Model-Adjusted Neyman Allocation

The model-adjusted allocation (Definition 4.3) minimizes

$$V_{\text{Smooth}}(\pi) = \sum_{d\in\{0,1\}} \frac{r_d(\pi)}{1 + a_d\, r_d(\pi)}, \quad a_d = \frac{(1-\beta_d)^2}{q_d},$$

with $r_1(\pi) = \sigma_1^2/(n\pi)$ and $r_0(\pi) = \sigma_0^2/(n(1-\pi))$.

Taking the derivative and using $dr_1/d\pi = -\sigma_1^2/(n\pi^2)$ and $dr_0/d\pi = \sigma_0^2/(n(1-\pi)^2)$:

$$\frac{dV_{\text{Smooth}}}{d\pi} = \frac{-\sigma_1^2/(n\pi^2)}{(1+a_1 r_1)^2} + \frac{\sigma_0^2/(n(1-\pi)^2)}{(1+a_0 r_0)^2}.$$

Setting this to zero and letting $f_d(\pi) = 1/(1 + a_d r_d(\pi)) = q_d/(q_d + r_d(\pi)(1-\beta_d)^2)$:

$$\frac{f_1(\pi)\,\sigma_1}{\pi} = \frac{f_0(\pi)\,\sigma_0}{1-\pi},$$

which rearranges to the fixed-point equation in Definition 4.3. Since $V_{\text{Smooth}}(\pi)$ is strictly convex on $(0,1)$, this fixed point is unique and the iteration converges from any starting value.

B.5 Proof of Theorem 4.7

We follow the approach of Proposition 2.5 in Waudby-Smith et al. (2024), adapted to the Kalman smoother setting. The key insight is that the smoother errors satisfy the same sub-Gaussian property needed for the mixture martingale construction.

Proof. **Step 1: Setup and error decomposition.** For the smoother estimator, define the period-t smoothing error for the treatment effect:

$$\Delta_t = (\mu_{1,t|T} - \mu_t(1)) - (\mu_{0,t|T} - \mu_t(0)).$$

Under independent arms in the linear-Gaussian model, the smoother errors satisfy:

$$\mathbb{E}[\Delta_t \mid Y_{1:T}, \mathcal{F}_{t-1}] = 0, \quad \mathrm{Var}(\Delta_t \mid Y_{1:T}) = P_{1,t|T} + P_{0,t|T} \equiv \sigma_t^2.$$

Moreover, since the model is Gaussian, Δ_t/σ_t is sub-Gaussian with parameter 1. Let $S_t = \sum_{j=1}^t \sigma_j^2$ denote the cumulative variance process. Define the standardized innovations $G_j = \Delta_j/\sigma_j$, which are independent standard normal under the linear-Gaussian model.

Step 2: Exponential supermartingale. For any fixed $\lambda \in \mathbb{R}$, define the process

$$\widetilde{M}_t(\lambda) = \exp\left\{\sum_{j=1}^t \left(\lambda\sigma_j G_j - \frac{\lambda^2\sigma_j^2}{2}\right)\right\}.$$

Since G_j are independent sub-Gaussian with parameter 1, the moment generating function satisfies $\mathbb{E}[e^{\lambda\sigma_j G_j} \mid \mathcal{F}_{j-1}] \le e^{\lambda^2\sigma_j^2/2}$, so $\widetilde{M}_t(\lambda)$ is a nonnegative supermartingale with $\mathbb{E}[\widetilde{M}_t(\lambda)] \le 1$.

Step 3: Gaussian mixture. Mixing over $\Lambda \sim \mathcal{N}(0, \eta^2)$ produces the process

$$\widetilde{M}_t = \int_{\lambda\in\mathbb{R}} \widetilde{M}_t(\lambda)\, dF(\lambda) = \exp\left\{\frac{\eta^2\left(\sum_{j=1}^t \sigma_j G_j\right)^2}{2(S_t\eta^2+1)}\right\} \cdot (S_t\eta^2+1)^{-1/2},$$

which is obtained by completing the square in the Gaussian integral. Since $\widetilde{M}_t(\lambda)$ is a supermartingale for each λ and the mixture is a convex combination, $\widetilde{M}_t$ is also a nonnegative supermartingale with $\mathbb{E}[\widetilde{M}_t] \le 1$.

Step 4: Ville's inequality. By Ville's inequality, $\mathbb{P}(\widetilde{M}_t \ge 1/\alpha \text{ for some } t \ge 1) \le \alpha$. Inverting $\widetilde{M}_t < 1/\alpha$ yields the half-width bound: for all $t \ge 1$,

$$\left|\frac{1}{t}\sum_{j=1}^t \sigma_j G_j\right| < \sqrt{\frac{2(S_t\eta^2+1)}{t^2\eta^2}\log\frac{\sqrt{S_t\eta^2+1}}{\alpha}}$$

with probability at least $1-\alpha$, simultaneously over all t.

Step 5: Plugging in consistent variance estimator. When S_t is unknown, we replace it with the predictable estimator $\hat{S}_t = \sum_{j=1}^t (P_{1,j|T} + P_{0,j|T})$ from the Kalman smoother. Since the smoother MSE $P_{d,t|T}$ converges to its steady-state value under Assumption 2.6, we have $\hat{S}_t/S_t \to 1$ almost surely. By standard arguments for plug-in confidence sequences (see Proposition 2.5 of Waudby-Smith et al. (2024)), the intervals

$$I_t = \left[\hat{\tau}_t \pm \sqrt{\frac{2(\eta^2\hat{S}_t+1)}{t^2\eta^2}\log\frac{\sqrt{\eta^2\hat{S}_t+1}}{\alpha}}\right]$$

form an asymptotic $(1-\alpha)$ confidence sequence that remains valid under adaptive sampling. □

C EM Algorithm for Unknown Parameters

When $\Theta = (\mu_{0,d}, \sigma_d, q_d, \beta_d;\, d = 0, 1)$ is unknown, the EM algorithm provides maximum likelihood estimation (Dempster et al., 1977; Shumway and Stoffer, 2017).

Algorithm 1 EM Algorithm for State-Space Model

Require: Observations $\{(\bar{Y}_{d,t})\}_{t=1}^{T}$; initial $\Theta^{(0)}$; tolerance $\varepsilon > 0$.
for $j = 1, 2, \ldots$ **do**
  E-step: Run Kalman filter and smoother with $\Theta^{(j-1)}$ to obtain $\{\mu_{d,t|T}, P_{d,t|T}\}$.
  Compute $\mathcal{Q}(\Theta \mid \Theta^{(j-1)}) = \mathbb{E}[\log L(\Theta) \mid Y_{1:T}, \Theta^{(j-1)}]$.
  M-step: $\Theta^{(j)} \leftarrow \arg\max_{\Theta} \mathcal{Q}(\Theta \mid \Theta^{(j-1)})$.
  $\ell^{(j)} \leftarrow \log L(\Theta^{(j)})$ ▷ observed-data log-likelihood
  if $|\ell^{(j)} - \ell^{(j-1)}|/|\ell^{(j-1)}| < \varepsilon$ **then break**
  end if
end for
return $\Theta^{(j)}$

For the AR(1) model with parameters $\Theta = (\beta_d, q_d, \sigma_d^2, \mu_{0,d})$, the M-step updates have closed-form solutions. Define the smoothed sufficient statistics:

$$A_d = \sum_{t=1}^{T} \mathbb{E}[\mu_t(d)^2 \mid Y_{1:T}] = \sum_{t=1}^{T} \big(\mu_{d,t|T}^2 + P_{d,t|T}\big), \tag{36}$$

$$B_d = \sum_{t=1}^{T} \mathbb{E}[\mu_t(d)\mu_{t-1}(d) \mid Y_{1:T}], \tag{37}$$

$$C_d = \sum_{t=1}^{T} \mathbb{E}[\mu_{t-1}(d)^2 \mid Y_{1:T}]. \tag{38}$$

The cross-covariance $\mathbb{E}[\mu_t(d)\mu_{t-1}(d) \mid Y_{1:T}]$ is obtained from the smoother via $P_{d,t,t-1|T} = J_{d,t-1} P_{d,t|T}$. Then the M-step updates are:

$$\beta_d^{(j+1)} = B_d / C_d, \tag{39}$$

$$q_d^{(j+1)} = \frac{1}{T}\big(A_d - 2\beta_d^{(j+1)} B_d + (\beta_d^{(j+1)})^2 C_d\big), \tag{40}$$

$$(\sigma_d^2)^{(j+1)} = \frac{1}{T} \sum_{t=1}^{T} \big[(\bar{Y}_{d,t} - \mu_{d,t|T})^2 + P_{d,t|T}\big] \cdot n_t \pi_t(d), \tag{41}$$

$$\mu_{0,d}^{(j+1)} = \mu_{d,0|T}. \tag{42}$$

Under standard regularity conditions, the MLE satisfies $\sqrt{T}(\hat{\Theta}_T - \Theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\Theta_0)^{-1})$, where $\mathcal{I}(\Theta) = \lim_{T\to\infty} T^{-1}\mathbb{E}[-\partial^2 \log L(\Theta)/\partial\Theta\partial\Theta^\top]$.

D ADDITIONAL SIMULATION RESULTS

Real data motivation. Our simulation parameters are calibrated from the ASOS digital experiments dataset (`https://osf.io/64jsb/`), a publicly available collection of batched A/B tests with no covariate information. In particular, Experiment 08bcc2 in that dataset exhibits a clear AR(1)-like novelty effect: the treatment arm mean decays smoothly toward a long-run level over approximately 200 periods with batch sizes of order $n \approx 30$, and the variance asymmetry between treatment and control arms is substantial. The parameters $\beta = 0.7$, $q_0 = 0.01$, $q_1 = 0.015$, $\sigma_0^2 = 1$, $\sigma_1^2 = 16$ are chosen to approximate these characteristics. The script is available at `https://github.com/JunzheShao98/aistats2026-2106`

Results for $\beta = 0.9$. We present additional simulation results for $\beta_0 = \beta_1 = 0.9$, with all other parameters identical to the main text ($q_0 = 0.01$, $q_1 = 0.015$, $\sigma_0^2 = 1$, $\sigma_1^2 = 16$, $n = 30$, $T = 200$).

Under this weaker mean-reversion, the smoother still dominates the naive estimator, but the efficiency gain is smaller than in the $\beta = 0.7$ case. This is consistent with Theorem 4.5: the model information $I_{\text{model}} = (1-\beta)^2/q$ decreases as β increases, so the smoother advantage diminishes for near-unit-root processes.

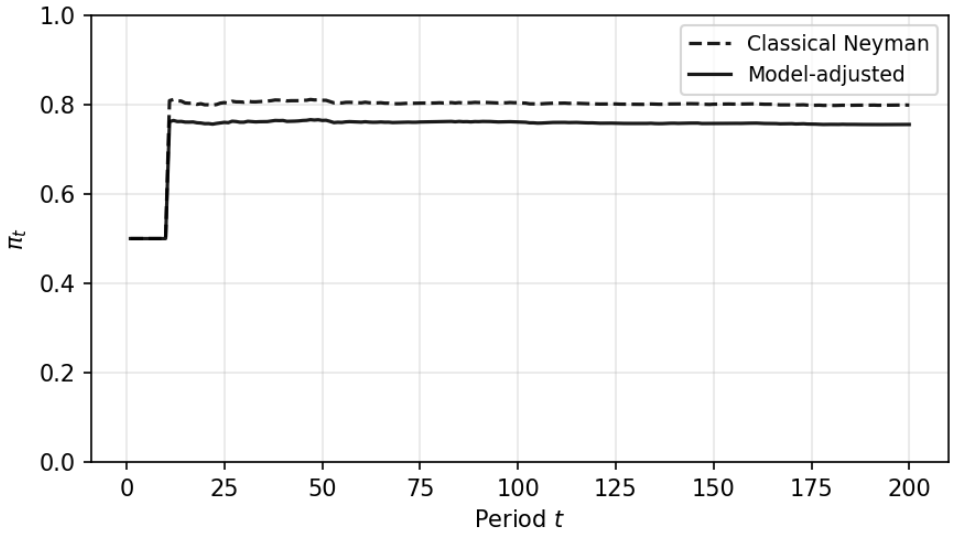

Figure 8: Allocation probability π_t under different rules ($\beta = 0.9$).

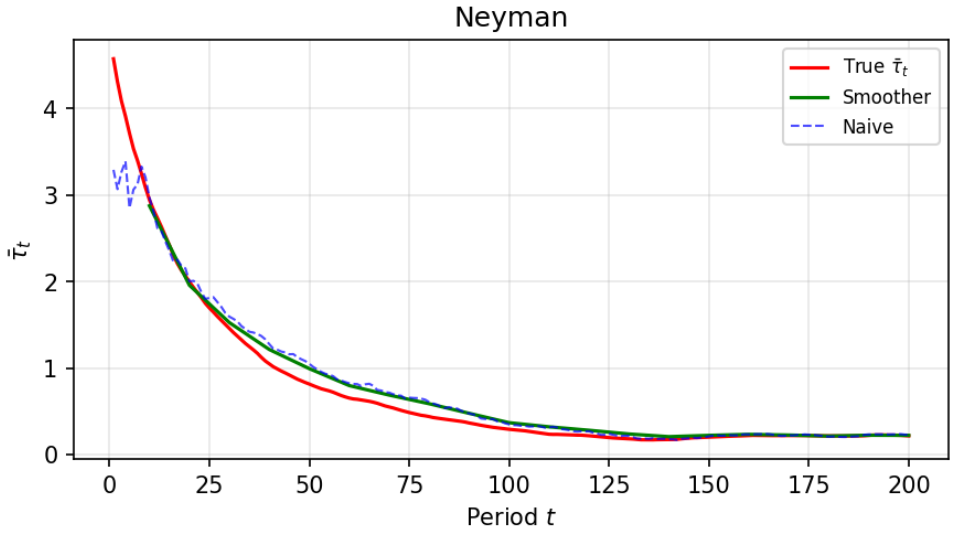

Figure 9: Running average treatment effect under Neyman allocation ($\beta = 0.9$).

D.1 Type I Error under Continuous Monitoring

We verify the anytime validity of the confidence sequence by computing cumulative miscoverage rates under continuous monitoring ("peeking"). For each Monte Carlo replication, we track whether the confidence interval has *ever* failed to cover the true $\bar{\tau}_t$ at any time up to t. The cumulative miscoverage at time t is the fraction of replications where coverage was violated at some time $s \leq t$. A valid confidence sequence should keep this rate below the nominal level α, while a naive fixed-horizon CI (constructed from the CLT at each t without correction for multiple comparisons) will exhibit inflated cumulative miscoverage under peeking.

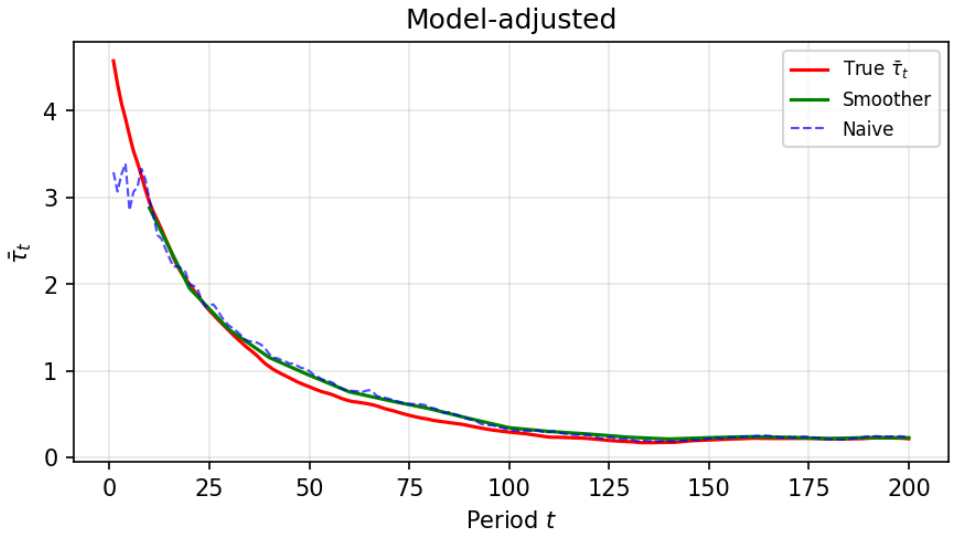

Figure 10: Running average treatment effect under model-adjusted allocation ($\beta = 0.9$).

E DISCUSSION: NEYMAN REGRET

A natural non-asymptotic performance criterion for adaptive allocation is the *Neyman regret*, defined as the gap between the variance achieved by the adaptive policy and the variance under the oracle allocation:

$$\text{Regret}_T = V_{\text{Smooth}}(\hat{\Pi}_T) - V_{\text{Smooth}}(\pi^{\text{Smooth}}),$$

where $\hat{\Pi}_T$ is the allocation trajectory produced by the adaptive rule and π^{Smooth} is the model-adjusted Neyman allocation from Theorem 4.5. Under the AR(1) state-space model, the plug-in estimates $\hat{\sigma}_{d,t-1}$ converge to σ_d and the fixed-point iteration converges, so the adaptive allocation satisfies $\hat{\pi}_t \to \pi^{\text{Smooth}}$ almost surely. The strict convexity of $V_{\text{Smooth}}(\pi)$ (Theorem 4.5) then implies that the excess variance $\text{Regret}_T \to 0$ as $T \to \infty$.

A quantitative non-asymptotic bound on the convergence rate—whether logarithmic in T as in recent work on Neyman regret for stationary designs, or $O(\sqrt{T})$ due to the additional state estimation error—is an important direction for future work.

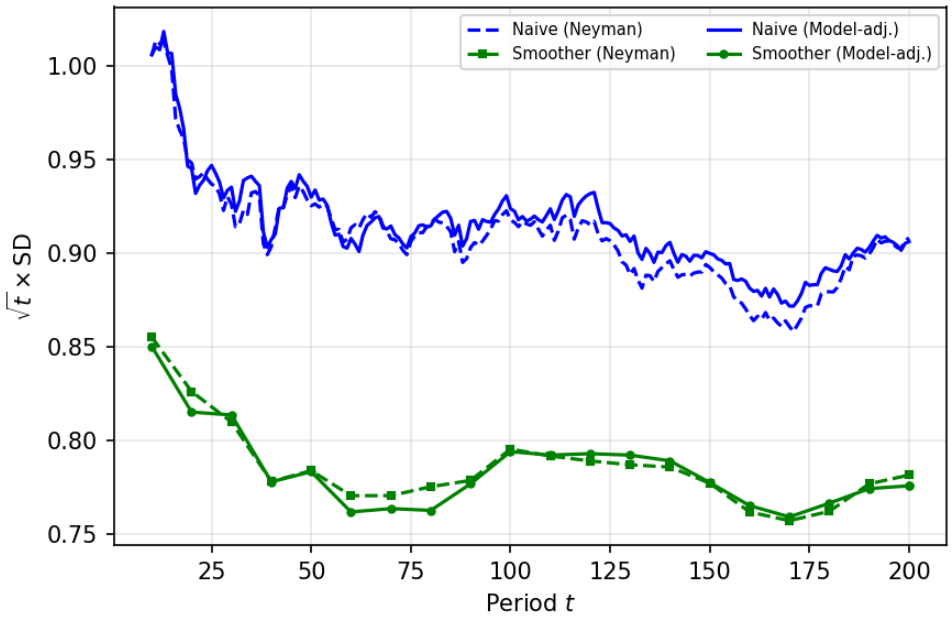

Figure 11: Monte Carlo standard deviation of estimators under Neyman and model-adjusted allocations ($\beta = 0.9$).

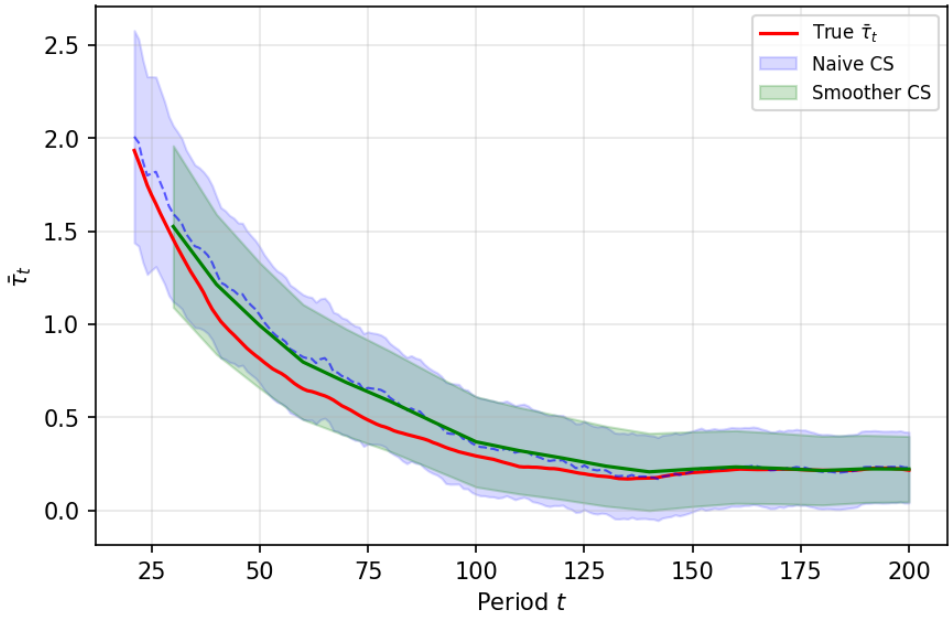

Figure 12: Anytime-valid confidence sequence for the treatment effect ($\beta = 0.9$).

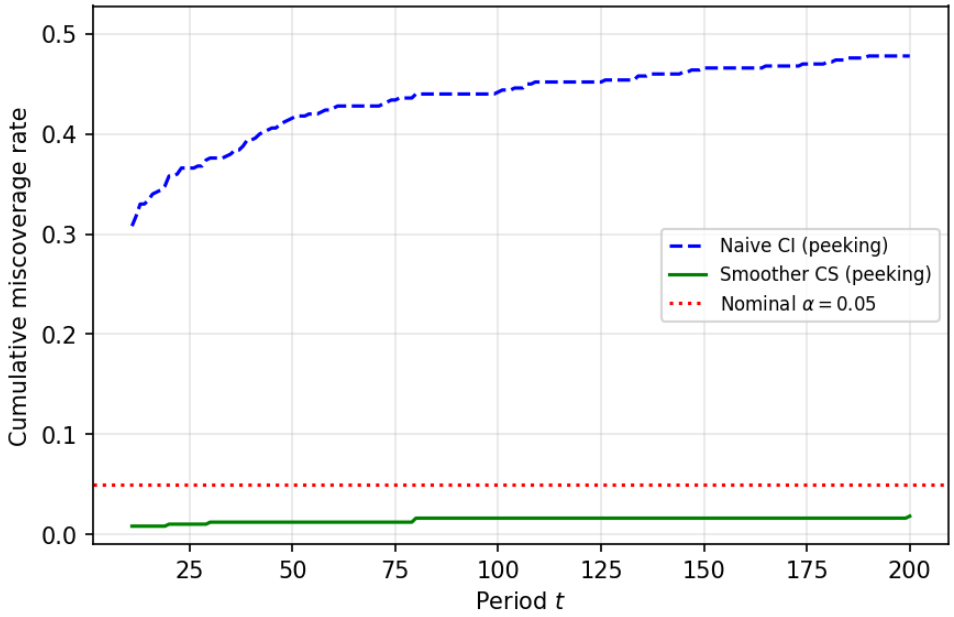

Figure 13: Cumulative miscoverage under peeking ($\beta = 0.7$). The naive CLT-based CI exhibits inflated miscoverage when peeking at multiple time points, while the anytime-valid CS from Theorem 4.7 remains below the nominal level $\alpha = 0.05$.

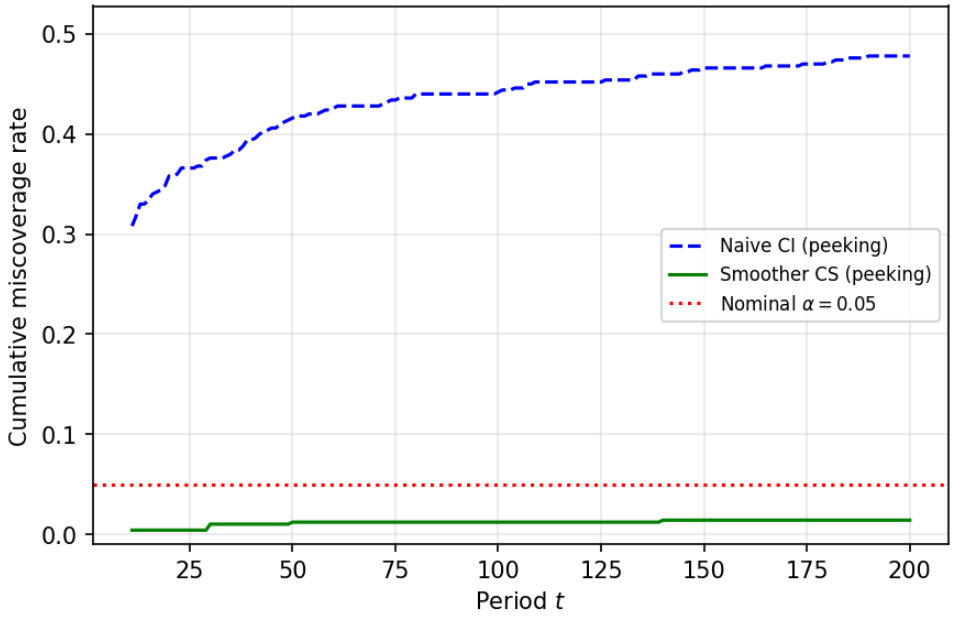

Figure 14: Cumulative miscoverage under peeking ($\beta = 0.9$).