Econometric Theory, **40**, 2024, 1211–1252.
doi:10.1017/S0266466623000051

FUNCTIONAL SEQUENTIAL TREATMENT ALLOCATION WITH COVARIATES

ANDERS BREDAHL KOCK
University of Oxford

DAVID PREINERSTORFER
University of St.Gallen

BEZIRGEN VELIYEV
Aarhus University

We consider a sequential treatment problem with covariates. Given a realization of the covariate vector, instead of targeting the treatment with highest conditional expectation, the decision-maker targets the treatment which maximizes a general *functional* of the conditional potential outcome distribution, e.g., a conditional quantile, trimmed mean, or a socioeconomic functional such as an inequality, welfare, or poverty measure. We develop expected regret lower bounds for this problem and construct a near minimax optimal sequential assignment policy.

1. INTRODUCTION

An expanding literature is concerned with statistical treatment rules, important contributions including Chamberlain (2000), Manski (2004), Dehejia (2005), Hirano and Porter (2009), Stoye (2009), Bhattacharya and Dupas (2012), Stoye (2012), Tetenov (2012), Manski and Tetenov (2016), Kitagawa and Tetenov (2018), Kitagawa and Tetenov (2019), Manski (2019), Athey and Wager (2021), and Zhou, Athey, and Wager (2022) (cf. also the overview in Hirano and Porter, 2019). These references all study a nonsequential setup, in which the dataset has been sampled before the policymaker enters the picture. The data could be observational or the outcome of a randomized controlled trial. In any case, the task of the policymaker in the nonsequential case is to construct policies with good properties given the data at hand. Sequential policies, on the other hand, allow the policymaker to enter already in the sampling phase and thus to fine-tune the treatment program as outcomes are observed. This opens the opportunity to gradually target the sampling effort where it is most useful, by monitoring the outcomes of the treatments as these are observed. There has been

We are grateful to the Editor, a Co-Editor, and three anonymous referees for their valuable comments. Address correspondence to Bezirgen Veliyev, Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark; e-mail: bveliyev@econ.au.dk.

a recent appreciation of this opportunity in the context of treatment allocation as witnessed by the works of Kasy and Sautmann (2021) and Kock, Preinerstorfer, and Veliyev (2022), both articles building on the literature on multi-armed bandits. The work of Currie and MacLeod (2020) studying the factors influencing medical doctors' optimal prescription strategies of antidepressants also draws on ideas from sequential decision-making.

The classical multi-armed bandit literature considers a policymaker who attempts to assign subjects to the treatment with the highest expected outcome, and without observing covariates. Two strands of developments of this idealized sequential setting that are particularly relevant for socioeconomic problems have attracted much attention: (i) Allowing the decision-maker to incorporate a vector of covariates in the assignment of each subject (cf. Woodroofe, 1979; Yang and Zhu, 2002; Rigollet and Zeevi, 2010; Perchet and Rigollet, 2013). This is important as not incorporating individual-specific heterogeneity into the policy does not exploit that individuals may react differently to every treatment. (ii) Problems where instead of targeting the outcome distribution with highest expectation, the decision-maker is interested in targeting another functional of the outcome distribution (cf. Sani, Lazaric, and Munos, 2012; Maillard, 2013; Tran-Thanh and Yu, 2014; Zimin, Ibsen-Jensen, and Chatterjee, 2014; Vakili and Zhao, 2016; Kock and Thyrsgaard, 2018; Cassel, Mannor, and Zeevi, 2018; Vakili, Boukouvalas, and Zhao, 2019; Ma et al., 2020). Of particular relevance for the present article is the recent paper Kock et al. (2022). In that article, a theory for sequential assignment problems was built for a policymaker targeting a general functional of the outcome distributions of the treatments. The functionals covered include quantiles, moment-based functionals, and a large range of inequality, welfare, and poverty measures. However, a setting without covariates was considered.[1] That targets other than the expected outcome are relevant in economic decision-making has recently been underscored by Bitler, Gelbach, and Hoynes (2006) or Rostek (2010).[2]

The only article we are aware of to provide regret bounds on a policy for a target other than the conditional expectation and in the presence of covariates is Kock and Thyrsgaard (2018). This paper has two limitations: First, it considers the special class of functionals which can be written as a Lipschitz-continuous function of the conditional mean and the conditional variance. Many fundamental functionals are not covered by their theory, e.g., conditional quantiles or many inequality, welfare, and poverty measures. Second, regret lower bounds for functional targets (beyond the mean) are not derived, and thus the question whether the algorithm they suggest is optimal is left unanswered.

[1] In fact, in early arXiv versions of Kock et al. (2022), the first version dating back to December 2018, a chapter on incorporating covariates is contained, which was later dropped from that paper and builds the basis for the present article.

[2] We kindly thank a referee for pointing out another related strand of literature on "distributional robust policy learning." In this literature, however, the target is the expected outcome, which one intends to maximize in a distributionally robust way (cf. Si et al., 2020b; Zhou et al., 2021b; Si et al., 2020a).

From a technical point of view, it is nontrivial to incorporate covariate information as the setting under investigation in the present article allows the covariates to be continuously distributed. Hence, one cannot simply fully condition on the covariates by treating each of the (infinitely many) values of the covariate vector separately. In the terminology of Stoye (2009), this would result in "no-data rules" since each value of the covariate vector would be observed at most once with probability 1. Thus, more care needs to be taken in the construction of good polices, and one needs to clarify which assumptions are necessary for informative policies to even exist for any given functional of interest.

The goal of the present article is to develop a minimax expected regret optimality theory for sequential treatment problems with functional targets and covariates. The regret function we work with is cumulative and, thus, has the following properties that are relevant in many situations:

- Every subject not assigned to the best treatment contributes to the regret.
- A loss incurred for one subject cannot be compensated by future assignments.

Although there are interesting regret notions that do not satisfy these criteria, there are certainly many situations where every individual matters such that a suboptimal assignment of one individual cannot be nullified by an improved assignment for the next individual.

Our contributions are as follows: To fix the setup, we begin by investigating which assumptions are necessary for informative policies to even exist. It turns out that even when the conditional potential outcome distributions depend uniformly equicontinuously on the covariates, i.e., under considerable regularity, no policy with sublinear maximal expected regret in the number of assignments exists. Here, we note that our cumulative regret notion implies that no policy will have a worse than linear dependence of regret in the number of assignments. This insight motivates us to impose a minimally stronger Hölder-equicontinuity assumption. As a consequence, even a slight relaxation of this assumption would imply that every policy incurs the worst-case linear maximal expected regret. We also show that if a policy does not incorporate covariate information, then its regret grows linearly even without considering the worst-case regret. We then introduce the functional upper-confidence-bound (F-UCB) policy in the presence of covariates. This is a binned version of the F-UCB policy introduced in Kock et al. (2022). Binning continuous covariates builds on the UCBogram of Rigollet and Zeevi (2010) and the work of Perchet and Rigollet (2013) who, however, focused exclusively on the conditional mean. We then establish regret upper bounds for the F-UCB policy and obtain lower bounds, proving its near minimax expected regret optimality when targeting general functionals. A challenge that arises when considering general functional targets compared with the setting of targeting the conditional mean in Rigollet and Zeevi (2010) is that we cannot rely on the specific linearity properties of the conditional mean and the ensuing concentration inequalities for sample averages. Instead, we work under a Lipschitz-type continuity assumption, which is satisfied by many functionals.

From a technical point of view, obtaining sharp lower bounds is the greatest challenge: In contrast to establishing lower bounds for policies targeting the conditional mean, it does not suffice to simply study regret over Bernoulli distributions. First, many functionals do not show sufficient variation over these. Second, lower bounds based on Bernoulli distributions are not informative in settings where one considers continuous outcome distributions. Hence, the standard arguments for providing lower bounds do not apply and we provide a novel construction that for any given functional exhibits distributions for which all policies must incur a high regret. This is challenging since the construction must obey several smoothness conditions when combining families of conditional distributions into a joint distribution. We stress that the lower bounds are established under an assumption that essentially only requires the functionals not to be constant over the set of potential outcome distributions considered. This requirement is very weak, and thus guarantees that the lower bounds hold even under quite stringent restrictions on the conditional outcome distributions.

2. THE SETUP AND TWO IMPOSSIBILITY RESULTS

The observational structure in this paper is the one of a multi-armed bandit problem with covariates. That is, the subjects to be treated $t = 1, \dots, n$ arrive sequentially, and have to be assigned to one out of $K \geq 2$ treatments. The assignment decision can incorporate previously observed outcomes, covariates, and randomization. We denote the potential outcome of assigning subject t to treatment i by $Y_{i,t}$, and assume throughout that $a \leq Y_{i,t} \leq b$, where $a < b$ are real numbers. Let $D_{cdf}([a,b])$ be the set of all cdfs F such that $F(a-) = 0$ and $F(b) = 1$. The vector of potential outcomes is defined as $Y_t = (Y_{1,t}, \dots, Y_{K,t})$; note that per subject only one coordinate of this vector can be observed. The covariate vector that comes with subject t is denoted by X_t, and we assume throughout that $X_t \in [0,1]^d$. Furthermore, for every t, we let G_t be a random variable, which can be used for randomization in assigning the tth subject. *Throughout this article, we assume that* $(Y_t, X_t) = (Y_{1,t}, \dots, Y_{K,t}, X_t)$, *for* $t \in \mathbb{N}$, *are i.i.d., and we assume that the sequence of randomizations* G_t *is i.i.d., and is independent of the sequence* (Y_t, X_t). The distribution of G_t will be referred to as the randomization measure, which we think of as being fixed, e.g., the uniform distribution on $[0,1]$. Note that the dependence structure within each Y_t is not restricted. We denote the distribution of (Y_t, X_t) by $\mathbb{P}_{Y,X}$, and let $\mathbb{P}_X$ be the marginal distribution of X_t. The conditional cumulative distribution function (cdf) of $Y_{i,t}$ given $X_t = x$ is defined as $F^i(y,x) = \mathsf{K}^i((-\infty, y], x)$, where $\mathsf{K}^i : \mathcal{B}(\mathbb{R}) \times [0,1]^d \to [0,1]$ denotes a regular conditional distribution of $Y_{i,t}$ given X_t, where $\mathcal{B}(\mathbb{R})$ are the Borel sets of $\mathbb{R}$. We shall often impose the following condition (cf. Remark 3.5 for a discussion of discrete covariates).

Assumption 2.1. The distribution $\mathbb{P}_X$ is absolutely continuous w.r.t. Lebesgue measure on $[0,1]^d$, with a density that is bounded from below and above by $\underline{c} > 0$ and $\overline{c}$, respectively.

A policy π is a triangular array $\{\pi_{n,t} : n \in \mathbb{N}, 1 \leq t \leq n\}$ of (measurable) functions,[3] where the assignment of the tth subject $\pi_{n,t}$ takes as input the covariates X_t, previously observed outcomes, covariates, and randomizations (i.e., the complete observational history), and a randomization G_t. We therefore have

$$\pi_{n,t} : [0,1]^d \times \Big[[a,b] \times [0,1]^d \times \mathbb{R}\Big]^{t-1} \times \mathbb{R} \to \mathcal{I}. \tag{1}$$

Given a policy π and $n \in \mathbb{N}$, the input to $\pi_{n,t}$ is denoted by (X_t, Z_{t-1}, G_t), where Z_{t-1} is defined recursively: The first treatment $\pi_{n,1}$ is a function of $(X_1, Z_0, G_1) = (X_1, G_1)$. The second treatment is a function of X_2, of $Z_1 := (Y_{\pi_{n,1}(X_1,Z_0,G_1),1}, X_1, G_1)$, and of G_2. For $t \geq 3$, we have

$$Z_{t-1} := (Y_{\pi_{n,t-1}(X_{t-1},Z_{t-2},G_{t-1}),t-1}, X_{t-1}, G_{t-1}, Z_{t-2}).$$

The $(t-1)(d+2)$-dimensional random vector Z_{t-1} can be interpreted as the information available after the $(t-1)$th treatment outcome has been observed.

Remark 2.2. In this article, we do not explicitly study the case of subjects arriving in batches, i.e., when the decision-maker cannot update the policy after every single observation, but only after the outcomes of all subjects in the current batch have been observed. If all subjects in the same batch need to be assigned to the same treatment and all batches are of a comparable size, the variable $Y_{i,t}$ may be interpreted as a summary statistic of the outcomes of batch t when all of its subjects were assigned to treatment i, e.g., the average outcome. Likewise, the covariate vector X_t can be interpreted as a summary statistic of the covariates in batch t, e.g., the average of the individual covariate vectors. With this reinterpretation, our results then go through as they are (the assumptions now being imposed at the level of the summary statistics).

The treatments are evaluated according to a functional $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ of the conditional potential outcome distribution, where the conditioning is on the covariates. As discussed after Assumption 2.3, our theory allows the policymaker to use a diverse set of functionals including many popular inequality, welfare, and poverty measures. The specific functional chosen by the policymaker will depend on the application, and encodes the particular distributional characteristics the policymaker is interested in. The best assignment for a subject with covariate vector $x \in [0,1]^d$ is defined as

$$\pi^\star(x) = \min \arg\max_{i \in \mathcal{I}} \mathsf{T}(F^i(\cdot, x)),$$

where the minimum has been taken as a concrete choice of breaking ties. Thus, the best assignment for a subject with covariate vector $x \in [0,1]^d$ is one whose *conditional* outcome distribution maximizes the functional of interest T.

[3]We allow a policy to incorporate n, because a decision-maker who knows the number of subjects to be assigned might want to incorporate this into the assignment mechanism and thus choose different sequences of assignments for different n.

We denote the "parameter"-space of all potential conditional cdfs $F^i(\cdot,x)$ by $\mathscr{D}$. More precisely, we assume that

$$\{F^i(\cdot,x) : i = 1,\dots,K \text{ and } x \in [0,1]^d\} \subseteq \mathscr{D}, \tag{2}$$

where $\mathscr{D}$ is a potentially large and nonparametric subset of $D_{cdf}([a,b])$. The set $\mathscr{D}$ encodes the assumptions one is willing to impose on the conditional outcome distributions. For example, it is sometimes convenient to restrict attention to conditional cdfs that are sufficiently smooth.

The main assumption on T we work with in the present paper is a Lipschitz-type condition first introduced in Kock et al. (2022) in a setting without covariates. The assumption takes the following form, where $(F,G) \mapsto \|F-G\|_\infty$ denotes the supremum metric on the set of cdfs on $\mathbb{R}$.

Assumption 2.3. The functional $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ and the nonempty set $\mathscr{D} \subseteq D_{cdf}([a,b])$ satisfy

$$|\mathsf{T}(F) - \mathsf{T}(G)| \le C\|F-G\|_\infty \quad \text{for every} \quad F \in \mathscr{D} \text{ and every } G \in D_{cdf}([a,b]), \tag{3}$$

for some $C > 0$.

As discussed at length in Appendices E and G of Kock et al. (2022), under suitable assumptions on $\mathscr{D}$, Assumption 2.3 is satisfied, e.g., for quantiles, moment-based functionals such as the variance, (trimmed) U-functionals, and generalized L-functionals (cf. Serfling, 1984), and many inequality, poverty, and welfare measures important for socioeconomic decision-making. The results in Appendix G of Kock et al. (2022) apply more generally, and can be used to show that Assumption 2.3 is also satisfied for other functionals, e.g., the semivariance. In the present paper, we keep the functional abstract and refer the interested reader to the just-mentioned appendices for examples and detailed discussions. Finally, we point to Lambert (2001), Chakravarty (2009), or Cowell (2011) for textbook treatments of policy relevant functionals. Apart from Assumption 2.3, we shall also impose the following measurability condition, which does not impose any practical restrictions.

Assumption 2.4. For every $m \in \mathbb{N}$, the function on $[a,b]^m$ that is defined via $x \mapsto \mathsf{T}(m^{-1}\sum_{j=1}^m \mathbb{1}\{x_j \le \cdot\})$, i.e., T evaluated at the empirical cdf corresponding to $x_1,\dots,x_m$, is Borel measurable.

We now introduce the regret function used in the present paper to compare different policies. Given a policy π, we define its (cumulative) regret as

$$\begin{aligned} R_n(\pi) &= R_n(\pi; F^1,\dots,F^K, X_n, Z_{n-1}, G_n) \\ &= \sum_{t=1}^n \left[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big) - \mathsf{T}\big(F^{\pi_{n,t}(X_t,Z_{t-1},G_t)}(\cdot,X_t)\big)\right]. \end{aligned}$$

This regret function incorporates the notion that every individual matters in the sense that a suboptimal allocation made for an individual cannot be nullified by later assignments. In the absence of covariates, this type of regret was also used in Kock et al. (2022).[4] In other settings going beyond the mean, cumulative regret functions have previously been used by Maillard (2013), Zimin et al. (2014), and Kock and Thyrsgaard (2018). Furthermore, Vakili et al. (2019) have stressed the practical importance of cumulative regret in the context of clinical trials where the loss to each individual must be controlled.

Remark 2.5. If we interpret $\mathsf{T}(F^i(\cdot, X_t))$ as subject t's subjective utility under uncertainty of treatment i, one may interpret the regret function we work with as that of a utilitarian planner, who intends to take each subject's utilities under uncertainty equally into account while learning what is best for each individual. We owe this interpretation to a referee.

We evaluate policies based on their worst-case behavior, i.e., we shall study minimax expected regret properties of policies. Here, the maximum will be taken over sets of possible joint distributions $\mathbb{P}_{Y,X}$.

When establishing lower bounds on maximal expected regret, we shall impose the following rather weak condition. It guarantees that there is a minimal amount of variation in the functional over a small subset of $\mathscr{D}$ (the set of all potential conditional outcome distributions).

Assumption 2.6. The functional $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ satisfies Assumption 2.3, and $\mathscr{D}$ contains two elements H_1 and H_2, such that

$$J_\tau := \tau H_1 + (1-\tau)H_2 \in \mathscr{D} \quad \text{for every } \tau \in [0,1],$$

and such that, for some $c_- > 0$, we have

$$\mathsf{T}(J_{\tau_2}) - \mathsf{T}(J_{\tau_1}) \geq c_-(\tau_2 - \tau_1) \quad \text{for every } \tau_1 \leq \tau_2 \text{ in } [0,1]. \tag{4}$$

We emphasize that equation (4) in Assumption 2.6 is satisfied if, e.g., $\tau \mapsto \mathsf{T}(J_\tau)$ is continuously differentiable on $[0,1]$ with an everywhere positive derivative.

Up to this point, *no* assumption has been imposed on the dependence of the conditional cdfs $F^i(\cdot,x)$ on $x \in [0,1]^d$. Keeping this dependence unrestricted would allow two subjects with similar covariates to have completely different conditional outcome distributions. We now prove that the maximal expected regret of *any* policy increases linearly in n if the dependence of $F^i(\cdot,x)$ on x is not further restricted. It even turns out that this statement continues to hold if one imposes the restriction that subjects with similar covariates have similar outcome distributions

[4] In the decision problem considered in Kock et al. (2022), one could alternatively try to develop policies that maximize the functional evaluated at the empirical cdf of all outcomes observed. If the functional is quasi-convex, this essentially results in the objective of assigning as many subjects to the best treatment as possible, which is strongly related to minimizing cumulative regret.

in the sense that[5]

$$\{F^i(y,\cdot) : i = 1,\dots,K \text{ and } y \in \mathbb{R}\} \quad \text{is uniformly equicontinuous.} \qquad \textbf{(5)}$$

The theorem we give next is obtained as an application of the lower bound developed in Theorem 3.9 of Section 3.2. Recall that Assumption 2.3 (which is a part of Assumption 2.6) implies that T is bounded, from which it follows that *no* policy will have a maximal expected regret increasing faster than linearly in n.

THEOREM 2.7. *Suppose $K = 2$ and that Assumption 2.6 is satisfied. Then there exists a constant $c_l > 0$ such that for every policy π and any randomization measure, we have*

$$\sup \mathbb{E}[R_n(\pi)] \geq c_l n \quad \textit{for every } n \in \mathbb{N},$$

where the supremum is taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1,\dots,n$, where $\mathbb{P}_{Y,X}$ satisfies equations (2) and (5), and where $\mathbb{P}_X$ is the uniform distribution on $[0,1]^d$.

Theorem 2.7 thus shows that without imposing further restrictions beyond equations (2) and (5), every policy incurs the worst-case linear maximal expected regret.

Theorem 2.7 shows that further assumptions beyond the uniform equicontinuity in (5) are needed in order for policies with nontrivial regret properties to exist. We shall from now on impose a Hölder-equicontinuity condition instead. This condition is only slightly stronger than uniform equicontinuity, but will turn out to be enough to ensure existence of (near) minimax optimal policies with nontrivial maximal expected regret. Throughout the article, $||\cdot||$ denotes the Euclidean norm.

Assumption 2.8. There exist a $\gamma \in (0,1]$ and an $L > 0$ such that, for every $i = 1,\dots,K$ and every $y \in \mathbb{R}$, we have

$$|F^i(y,x_1) - F^i(y,x_2)| \leq L||x_1 - x_2||^\gamma \text{ for every } x_1, x_2 \in [0,1]^d.$$

Before studying policies that incorporate covariate information, one may wonder (e.g., as a sanity check of the framework considered) what happens if one uses a policy that ignores covariates. Our next result shows that—unless the underlying distribution $\mathbb{P}_{Y,X}$ happens to be such that the covariates are completely irrelevant for the assignment problem—any policy that *ignores covariates* must incur a linear expected regret. Formally, a policy π is said to ignore covariates, if there exists another double array $\tilde{\pi}_{n,t} : \left[[a,b] \times \mathbb{R}\right]^{t-1} \times \mathbb{R} \to \mathcal{I}$ of measurable functions, such that, for every n and every $t = 1,\dots,n$, we have $\pi_{n,t} = \tilde{\pi}_{n,t} \circ \Pi_t$, where the function Π_t projects every $w = (x,z,g)$ in the domain of $\pi_{n,t}$ to $(\tilde{z},g)$, $\tilde{z}$ being obtained from $z \in \left[[a,b] \times [0,1]^d \times \mathbb{R}\right]^{t-1}$ by dropping the $(t-1)$ coordinates taking values in $[0,1]^d$. Note that, then, $\pi_{n,t}(Z_{t-1}, G_t) = \tilde{\pi}_{n,t}(\tilde{Z}_{t-1}, G_t)$, where, for $t \geq 2$, we have $\tilde{Z}_{t-1} = (Y_{\tilde{\pi}_{n,t-1}(\tilde{Z}_{t-2},G_{t-1})}, G_{t-1}, \dots, Y_{\tilde{\pi}_{n,1}(\tilde{Z}_0,G_1)}, G_1)$ and $(\tilde{Z}_0, G_1) = G_1$.

[5]The assumption in equation (5) imposes that, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|x_1 - x_2\| \leq \delta$, for $\|\cdot\|$ the Euclidean norm, implies $|F^i(y,x_1) - F^i(y,x_2)| \leq \varepsilon$ for every $i = 1,\dots,K$ and every $y \in \mathbb{R}$.

THEOREM 2.9. *Let $K = 2$, suppose that $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ satisfies Assumption 2.3, and let $\mathbb{P}_{Y,X}$ satisfy equation (2) and Assumption 2.8. Define the sets*

$$A_1 := \{x \in [0,1]^d : \mathsf{T}(F^1(\cdot,x)) > \mathsf{T}(F^2(\cdot,x))\},$$
$$A_2 := \{x \in [0,1]^d : \mathsf{T}(F^1(\cdot,x)) < \mathsf{T}(F^2(\cdot,x))\}.$$

Then there exists a $c_l > 0$ such that for every policy π ignoring covariates, and any randomization measure, we have

$$\mathbb{E}[R_n(\pi)] \geq c_l \min(\mathbb{P}_X(A_1), \mathbb{P}_X(A_2))n \quad \textit{for every } n \in \mathbb{N}. \tag{6}$$

Thus, the expected regret of any policy ignoring covariates must increase at the worst-case linear rate in n, for *any* distribution $\mathbb{P}_{Y,X}$ for which the identity of the best treatment depends on the covariates in the sense that

$$\min\left(\mathbb{P}_X(A_1), \mathbb{P}_X(A_2)\right) > 0.$$

The proof of Theorem 2.9 relies on a direct construction rather than the classic technique for establishing lower bounds by a reduction to a suitable testing problem and an application of the Bretagnolle–Huber inequality (as exposited in, e.g., Chapter 2 of Tsybakov, 2009). As a result of this direct construction, and contrary to all other lower bounds established in this article, the lower bound in Theorem 2.9 is valid even pointwise, since it makes a statement about any fixed distribution $\mathbb{P}_{Y,X}$. That is, the linear increase in expected regret in (6) is not a result of considering a worst-case scenario.

3. THE F-UCB POLICY IN THE PRESENCE OF COVARIATES

We now introduce a version of the F-UCB policy that incorporates covariate information. This policy generalizes the UCBogram in Rigollet and Zeevi (2010) from the conditional mean setting to the general functional setup. It also generalizes the F-UCB policy of Kock et al. (2022) by allowing the policymaker to incorporate covariate information in a setting where one targets a general functional of the outcome distributions. Since we are studying a setting with continuously distributed covariates, it is not possible to construct policies with low maximal expected regret by fully conditioning on the infinitely many values of x based on a fixed number of observations: Such a full conditioning would result in "no-data rules" as in Stoye (2009) since each value of $x \in [0,1]^d$ would be observed at most once with probability 1 irrespective of $n \in \mathbb{N}$. Thus, incorporating covariate information requires more care. The underlying idea that we use is to categorize subjects into groups according to the similarity of their covariate vector and to run, separately within each group, a policy targeting the treatment that is best for the "average" subject in each group. This will be justified by Assumption 2.8.

Two covariate vectors x_1 and x_2 are considered similar, if they fall into the same element of a given partition $B_{n,1}, \ldots, B_{n,M(n)}$ of $[0,1]^d$, where every $B_{n,j}$ is a nonempty Borel set. Targeting the "on average"-best treatment for each group here means that for $B_{n,j}$ with $\mathbb{P}_X(B_{n,j}) > 0$ our policy targets a treatment that attains $\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j})$, where $F^i_{n,j}$ is the conditional cdf of $Y_{i,t}$ given $X_t \in B_{n,j}$, i.e.,

$$F^i_{n,j}(y) := \frac{1}{\mathbb{P}_X(B_{n,j})} \int_{B_{n,j}} F^i(y,x) d\mathbb{P}_X(x). \tag{7}$$

Note that in general $\arg\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j}) \neq \arg\max_{i\in\mathcal{I}} \mathsf{T}(F^i(\cdot,x))$ even though $x \in B_{n,j}$. Targeting $\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j})$, hence, results in a bias. The choice of the partition $B_{n,1},\ldots,B_{n,M(n)}$ needs to balance this bias against an increase in variance due to having fewer subjects in each group. This is akin to choosing a bandwidth to balance variance and bias terms in nonparametric estimation problems.

In order to describe the F-UCB policy in the presence of covariates, we need to introduce the following notation. For any policy π and $B_{n,1},\ldots,B_{n,M(n)}$ as above, let

$$S^i_{n,j}(t) = \sum_{s=1}^{t} \mathbb{1}_{\{X_s\in B_{n,j},\ \pi_{n,s}(X_s,Z_{s-1},G_s)=i\}}$$

be the number of times that it has assigned treatment i to individuals with covariates in $B_{n,j}$ up to time t. On the event $\{S^i_{n,j}(t) > 0\}$, define the empirical cdf based on the outcomes of all subjects in $\{1,\ldots,t\}$ with covariates in $B_{n,j}$ that have been assigned to treatment i as

$$\hat{F}^i_{n,t,j}(z) = \frac{1}{S^i_{n,j}(t)} \sum_{s=1}^{t} \mathbb{1}_{\{Y_{i,s}\le z\}} \mathbb{1}_{\{X_s\in B_{n,j},\ \pi_{n,s}(X_s,Z_{s-1},G_s)=i\}}.$$

The F-UCB policy with covariates, $\bar{\pi}$, is described in Policy 1 (where C is the constant from Assumption 2.3). We note that it amounts to using the F-UCB policy $\hat{\pi}$, say, of Kock et al. (2022) locally on each $B_{n,j}$. Their policy was defined in a setting without covariates and external randomization. Furthermore, as discussed after Theorem 3.1, we must carefully deal with the fact that the number of observations falling in each group is random as it is a function of the covariates. Choosing the tuning parameter $\beta = 2+\sqrt{2}$ minimizes the constant in the uniform upper bounds on expected regret (cf. Theorem 3.1).[6]

3.1. Upper Bounds on the Maximal Expected Regret of $\bar{\pi}$ and a First Lower Bound

The following theorem gives an upper bound on the maximal expected regret of the F-UCB Policy 1 in the presence of covariates, and for *any* choice of partition. This flexibility may be useful since the policymaker is often constrained in the way groups can be formed. The result quantifies how the partitioning affects the regret guarantees. We denote $\overline{\log}(x) := \max(1, \log(x))$, for $x > 0$.

[6]Note that the policy $\bar{\pi}$ actually does not incorporate randomization, which we do not suppress notationally, however.

Policy 1: F-UCB policy with covariates $\bar{\pi}$

Inputs: $\beta > 2$, Partition $B_{n,1},\dots,B_{n,M(n)}$ of $[0,1]^d$ into non-empty Borel sets
Set: $N_j = 1$ for $j = 1,\dots,M(n)$
for $t = 1,\dots,n$ **do**
  for $j = 1,\dots,M(n)$ **do**
    if $X_t \in B_{n,j}$ *and* $N_j \le K$ **then**
      assign $\bar{\pi}_t(X_t, Z_{t-1}, G_t) = N_j$
      $N_j \leftarrow N_j + 1$
    end
    if $X_t \in B_{n,j}$ *and* $N_j > K$ **then**
      assign $\bar{\pi}_t(X_t, Z_{t-1}, G_t) =$
       $\min \arg\max_{i \in \mathcal{I}} \left\{ \mathsf{T}(\hat{F}^i_{n,t-1,j}) + C\sqrt{\beta \log(N_j)/(2S^i_{n,j}(t-1))} \right\}$
      $N_j \leftarrow N_j + 1$
    end
  end
end

THEOREM 3.1. *Suppose Assumptions 2.3 and 2.4 hold. Assume further that $\mathscr{D}$ is convex. Consider the F-UCB policy with covariates $\bar{\pi}$, and let $V_{n,j} = \sup_{x_1,x_2 \in B_{n,j}} \|x_1 - x_2\|$ be the diameter of $B_{n,j}$. Then, for $c = c(\beta, C) = C\sqrt{2\beta + (\beta+2)/(\beta-2)}$, it holds that*

$$\sup \mathbb{E}[R_n(\bar{\pi})] \le \sum_{j=1}^{M(n)} \left[c\sqrt{Kn\mathbb{P}_X(B_{n,j})\overline{\log}(n\mathbb{P}_X(B_{n,j}))} + 2CLV^{\gamma}_{n,j} n\mathbb{P}_X(B_{n,j}) \right], \ \forall n \in \mathbb{N}, \tag{8}$$

where the supremum is taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1,\dots,n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2) and Assumption 2.8 with L and γ, and where $\mathbb{P}_X$, the marginal distribution of X_t, is fixed.[7]

Each of the summands $j = 1,\dots,M(n)$ in the upper bound on the maximal expected regret in equation (8) consists of two parts: The first part is structurally very similar to the upper bound of Theorem 4.1 in Kock et al. (2022), which the proof of Theorem 3.1 draws on. The difference is that the total number of subjects to be treated, n, has now been replaced by $n\mathbb{P}_X(B_{n,j})$, the number of subjects expected to fall into $B_{n,j}$. Inspection of the proof shows that the first part of the upper bound in (8) is the regret we expect to accumulate on $B_{n,j}$, compared to always assigning the treatment that is best for the "average subject" in $B_{n,j}$,

[7] Here, $\mathbb{P}_X(B_{n,j})\overline{\log}(n\mathbb{P}_X(B_{n,j}))$ is to be interpreted as 0 in case $\mathbb{P}_X(B_{n,j}) = 0$.

i.e., compared to always assigning an element of $\arg\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j})$, where we recall the definition of $F^i_{n,j}$ from equation (7). The second part in each summand in the upper bound in (8) is a bias term: It is the approximation error incurred due to $\bar{\pi}$ effectively targeting $\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j})$ instead of $\mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big)$ for every $x \in B_{n,j}$.

The proof of Theorem 3.1 introduces the following challenges compared to that for conditional mean only in Rigollet and Zeevi (2010) and Perchet and Rigollet (2013). First, we cannot rely on the special linear structure of the conditional mean and the related concentration inequalities. Second, since Policy 1 relies on using the F-UCB policy separately for each group, it is important that the $F^i_{n,j}$ in (7) belong to (the closure of) $\mathscr{D}$ as the $F^i_{n,j}$ become the groupwise treatment outcome distributions of interest. Otherwise, one would not be able to invoke Assumption 2.3 separately for each group. That $F^i_{n,j}$ belongs to the closure of $\mathscr{D}$ for each j with $\mathbb{P}_X(B_{n,j}) > 0$ is established in Lemma A.3 by an approximation argument relying on the convexity of $\mathscr{D}$. Finally, we overcome the fact that the number of subjects falling in each group (that is, the sample size for each group) is random by means of careful conditioning arguments. These arguments are different from those used for the conditional mean in Rigollet and Zeevi (2010) and Perchet and Rigollet (2013) as we do not impose a margin condition and cannot make use of the special linearity structure of the conditional mean (cf. the arguments beginning after (C.4) in the proof of Theorem 3.1).

A frequently used class of partitions of $[0,1]^d$ are hypercubes, which are obtained by hard thresholding each coordinate of X_t. The so-created groups may not only result in low regret, but are also relevant due to their simplicity and resemblance to ways of grouping subjects in practice. More precisely, fix $P \in \mathbb{N}$ and define, for every $k = (k_1,\ldots,k_d) \in \{1,\ldots,P\}^d$, the hypercube

$$\left\{x \in [0,1]^d : \frac{k_l - 1}{P} \le x_l \preceq \frac{k_l}{P},\ l = 1,\ldots,d\right\}, \tag{9}$$

where $\preceq$ is to be interpreted as $\le$ for $k_l = P$, and as $<$ otherwise. This defines a partition of $[0,1]^d$ into P^d hypercubes with side length $\frac{1}{P}$ each. We now order these hypercubes lexicographically according to their index vector k, to obtain the corresponding *cubic partition* $B^P_1,\ldots,B^P_{pd}$. The following result specializes Theorem 3.1 to this specific partition and for a choice of P that will be shown to be optimal below.

COROLLARY 3.2. *Suppose Assumptions 2.3 and 2.4 hold. Assume further that $\mathscr{D}$ is convex. Let $\gamma \in (0,1]$. Consider the F-UCB policy with covariates $\bar{\pi}$, based on a cubic partition $B_{n,j} = B^P_j$, for $j = 1,\ldots,M(n) = P^d$, as defined in equation (9), and with $P = \lceil n^{1/(2\gamma+d)} \rceil$. Then there exists a constant $c = c(d,L,\gamma,\bar{c},C,\beta) > 0$ such that*

$$\sup \mathbb{E}\left[R_n(\bar{\pi})\right] \le c\sqrt{K\overline{\log}(n)}\; n^{1-\frac{\gamma}{2\gamma+d}} \quad \textit{for every } n \in \mathbb{N}, \tag{10}$$

where the supremum is taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$ *for* $t = 1, \ldots, n$*, where* $\mathbb{P}_{Y,X}$ *satisfies equation (2), Assumption 2.1 with* $\overline{c}$ *(and any* $\underline{c}$*), and Assumption 2.8 with* L *and* γ*.*

Corollary 3.2 reveals that it is possible to achieve sublinear (in n) maximal expected regret under the Hölder-equicontinuity condition imposed through Assumption 2.8. This is interesting also in light of Theorem 2.7, which showed that under the slightly weaker assumption of uniform equicontinuity, every policy has linearly increasing maximal expected regret. Hence, there is little room for weakening Assumption 2.8. Note that a "curse of dimensionality" is present, in the sense that the upper bound in Corollary 3.2 gets close to linear in n, as the number of covariates d increases. This is due to the fact that as a part of the regret minimization, one *sequentially* estimates the conditional distributions $F^i(y, \cdot)$ of the treatment outcomes, where each cdf is a function of d variables. Finally, we observe that the upper bound is increasing in the number of available treatments K. Intuitively, this is because more observations must be used for experimentation when more treatments are available.

The partitioning used in Corollary 3.2 results in a near minimax optimal policy, as we show in the following theorem, which establishes a lower bound on maximal expected regret. The statement follows from Theorem 3.9 in Section 3.2.

THEOREM 3.3. *Suppose* $K = 2$ *and that Assumption 2.6 is satisfied. Let* $\gamma \in (0,1]$*. Then, for every* $\varepsilon \in (0, \gamma/(2\gamma + d))$*, every policy* π*, and any randomization measure, we have*

$$\sup \mathbb{E}[R_n(\pi)] \geq n^{1-\frac{\gamma}{2\gamma+d}}\, n^{-\varepsilon} c_l(\varepsilon) \quad \textit{for every } n \in \mathbb{N},$$

where the supremum is taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$*, for* $t = 1, \ldots, n$*, where* $\mathbb{P}_{Y,X}$ *satisfies equation (2), Assumption 2.8 with parameters* γ *and* $L = 1/\sqrt{17}$*,* $\mathbb{P}_X$ *is the uniform distribution on* $[0,1]^d$*, and where*

$$c_l^{-1}(\varepsilon) = 64^{1+1/\alpha(\varepsilon)}(8d(c\_2L)^{-\alpha(\varepsilon)} + 1)^{1/\alpha(\varepsilon)} \quad \textit{with} \quad \alpha(\varepsilon) = (2\gamma + d)\varepsilon/\gamma.$$

Comparing the lower bound on maximal regret in Theorem 3.3 to the upper bound on maximal expected regret established in Corollary 3.2 reveals that the F-UCB policy with a cubic partition and with $P = \lceil n^{1/(2\gamma+d)} \rceil$ is near-optimal: If a policy with strictly smaller maximal expected regret exists, the order of improvement must be $o(n^{\varepsilon})$ for all $\varepsilon \in (0, \gamma/(2\gamma + d))$, e.g., logarithmic. In particular, this also means that if nothing prohibits cubic partitioning, not much can be gained from a maximal expected regret point of view in searching for "better" partitions under the given set of assumptions.

Remark 3.4 (Unknown horizon and the doubling trick). The policy $\bar{\pi}$ with cubic partitioning $P = \lceil n^{1/(2\gamma+d)} \rceil$, as considered in Corollary 3.2, can be used in practice only if one knows n, i.e., the policy is not "anytime." If n is unknown, however, one can use the "doubling trick" to construct a policy with an upper bound on the maximal expected regret that is of the same order as in Corollary 3.2, but with higher multiplicative constants. In essence, the doubling trick works by "restarting" the policy at times 2^m, $m \in \mathbb{N}$. We refer to Shalev-Shwartz (2012) and the recent work by Besson and Kaufmann (2018) for more details on the doubling trick.

Remark 3.5 (Discrete covariates). We mostly focus on the case of continuous covariates (although this is not formally required in Theorem 3.1). A natural, and also near minimax rate-optimal, solution to incorporate discrete covariates would be to fully condition on these, i.e., to apply the F-UCB policy of Kock et al. (2022) separately for each combination of discrete covariates. In the present article, we omit formal statements concerning discrete covariates, but we emphasize that corresponding results can be obtained by conditioning arguments.

3.2. Optimality Properties under the Margin Condition

Besides mild conditions on $\mathbb{P}_X$, our results so far have only assumed that the conditional distributions of the treatment outcomes are Hölder-equicontinuous. In particular, the sets of distributions over which the F-UCB policy has been shown to be optimal do not restrict the (unknown) similarity of the best and second-best treatment. In the present section, we shall see that in classes of distributions where the best and second-best treatment are "well separated," the upper bound on maximal expected regret of the F-UCB policy can be lowered (without changing the policy), and that the F-UCB policy optimally adapts to the degree of similarity of the best and the remaining treatments.

Besides being of interest in their own right, the results in the present section are instrumental in proving our impossibility result Theorem 2.7 and to establishing the expected regret lower bound in Theorem 3.3.

To formally define the well-separateness condition we shall work with, we need to define, for every $x \in [0,1]^d$, the second-best treatment $\pi^\sharp(x)$; note that in principle there can be multiple treatments that are as good as the best treatment $\pi^*(x)$. For $x \in [0,1]^d$, if $\min_{i \in \mathcal{I}} \mathsf{T}(F^i(\cdot,x)) < \mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big)$, we define the second-best treatment as

$$\pi^\sharp(x) := \min \arg\max_{i \in \mathcal{I}} \left\{ \mathsf{T}(F^i(\cdot,x)) : \mathsf{T}(F^i(\cdot,x)) < \mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big) \right\},$$

and we set $\pi^\sharp(x) = 1$ otherwise, i.e., if all treatments are equally good. We can now introduce the *margin condition*.

Assumption 3.6. There exist an $\alpha \in (0,1)$ and a $C_0 > 0$ such that[8]

$$\mathbb{P}_X\Big(x \in [0,1]^d : 0 < \mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big) - \mathsf{T}\big(F^{\pi^\sharp(x)}(\cdot,x)\big) \le \delta\Big) \le C_0 \delta^\alpha \text{ for all } \delta \in [0,1].$$

[8] We note that the events in the displayed equation of Assumption 3.6 are not necessarily Borel measurable. Therefore, Assumption 3.6 implicitly imposes measurability on all events considered. Note, however, that in case Assumptions 2.3 and 2.8 as well as the inclusion in equation (2) are assumed, this measurability condition is easily seen to be satisfied.

The margin condition restricts how likely it is that the best and second-best treatment are close to each other. In particular, it limits the probability of these two treatments being almost equally good, i.e., being within a δ-margin. Assumptions of this type have previously been used in the works of Mammen and Tsybakov (1999), Tsybakov (2004), and Audibert and Tsybakov (2007) in the statistics literature. In the context of statistical treatment rules, the margin condition has recently been used in the work of Kitagawa and Tetenov (2018), who considered empirical welfare maximization in a static treatment allocation problem. Finally, the margin condition was used by Rigollet and Zeevi (2010) and Perchet and Rigollet (2013) in the context of a multi-armed bandit problem targeting the conditional mean. The proofs of the results in the present section draw on their ideas.

Adding the margin condition, the maximal expected regret of the F-UCB policy based on cubic partitions can be bounded as follows.

THEOREM 3.7. *Suppose Assumptions 2.3 and 2.4 hold. Assume further that $\mathscr{D}$ is convex. Let $\gamma \in (0,1]$. Consider the F-UCB policy with covariates $\bar{\pi}$, based on a cubic partition $B_{n,j} = B_j^P$, for $j = 1,\ldots,M(n) = P^d$, as defined in equation (9), and with $P = \lceil n^{1/(2\gamma+d)} \rceil$. Then there exists a constant $c = c(d,L,\gamma,\underline{c},\bar{c},C,C_0,\alpha,\beta) > 0$ such that*

$$\sup \mathbb{E}\left[R_n(\bar{\pi})\right] \leq cK\overline{\log}(n) n^{1-\frac{\gamma(1+\alpha)}{2\gamma+d}} \quad \textit{for every } n \in \mathbb{N}, \tag{11}$$

where the supremum is taken over all $(Y_t,X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1,\ldots,n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2), Assumption 2.1 with $\underline{c}$ and $\bar{c}$, Assumption 2.8 with L and γ, and Assumption 3.6 with $\alpha \in (0,1)$ and $C_0 > 0$.

Compared with Corollary 3.2, the exponent on n in the upper bound on regret is smaller, the difference depending on α. Thus, in the presence of Assumption 3.6, the regret guarantee of the F-UCB policy is stronger, even without incorporating α into the policy. We shall see in Theorem 3.9 that the upper bound on maximal expected regret in Theorem 3.7 is optimal in n up to logarithmic factors. Furthermore, the order of the upper bound on maximal expected regret in Theorem 3.7 is, up to logarithmic factors, the same as the rate that Rigollet and Zeevi (2010) and Perchet and Rigollet (2013) obtained for the case of targeting the mean functional. The margin condition also allows us to prove an upper bound on the expected number of suboptimal assignments made by the F-UCB policy. We shall define the number of suboptimal assignments for a policy π over the course of a total of n assignments as

$$\begin{aligned} S_n(\pi) &= S_n(\pi; F^1,\ldots,F^K,X_n,Z_{n-1},G_n) \\ &= \sum_{t=1}^{n} \mathbb{1}\left\{\pi_{n,t}(X_t,Z_{t-1},G_t) \notin \arg\max\{\mathsf{T}(F^i(\cdot,X_t)) : i = 1,\ldots,K\}\right\}. \end{aligned}$$

We now establish a uniform upper bound on $\mathbb{E}[S_n(\bar{\pi})]$ for the F-UCB policy $\bar{\pi}$ based on cubic partitions.

THEOREM 3.8. *Suppose Assumptions 2.3 and 2.4 hold. Assume further that $\mathscr{D}$ is convex. Let $\gamma \in (0,1]$. Consider the F-UCB policy with covariates $\bar{\pi}$, based on a cubic partition $B_{n,j} = B_j^P$, for $j = 1,\dots,M(n) = P^d$, as defined in equation (9), and with $P = \lceil n^{1/(2\gamma+d)} \rceil$. Then there exists a constant $c = c(d, L, \gamma, \underline{c}, \bar{c}, C, C_0, \alpha, \beta) > 0$ such that*

$$\sup \mathbb{E}\left[S_n(\bar{\pi})\right] \le c[K\overline{\log}(n)]^{\frac{\alpha}{1+\alpha}} n^{1-\frac{\alpha\gamma}{2\gamma+d}} \quad \textit{for every } n \in \mathbb{N}, \tag{12}$$

where the supremum is taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1,\dots,n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2), Assumption 2.1 with $\underline{c}$ and $\bar{c}$, Assumption 2.8 with L and γ, and Assumption 3.6 with $\alpha \in (0,1)$ and $C_0 > 0$.

The upper bound in Theorem 3.8 is a useful theoretical guarantee because it limits the number of subjects who receive suboptimal treatments. As the last result in this section, we prove that the upper bounds in Theorems 3.7 and 3.8 are near-minimax optimal. This ensures, in particular, that the good behavior of the maximal expected regret of the F-UCB policy does not come at the price of excessive experimentation, leading to unnecessarily many suboptimal assignments.

THEOREM 3.9. *Suppose $K = 2$ and that Assumption 2.6 is satisfied. Let $\gamma \in (0,1]$. Then, for every policy π and any randomization measure, we have*

$$\sup \mathbb{E}[R_n(\pi)] \ge n^{1-\frac{\gamma(1+\alpha)}{2\gamma+d}} \Big/ \left[64^{1+1/\alpha}(C_0+1)^{1/\alpha}\right] \quad \textit{for every } n \in \mathbb{N}, \tag{13}$$

and

$$\sup \mathbb{E}[S_n(\pi)] \ge n^{1-\frac{\alpha\gamma}{d+2\gamma}} \Big/ 32 \quad \textit{for every } n \in \mathbb{N}, \tag{14}$$

where both suprema are taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1,\dots,n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2), Assumption 2.8 with parameters γ and $L = 17^{-1/2}$, Assumption 3.6 with $\alpha \in (0,1)$ and $C_0 = 8d(c\_2L)^{-\alpha}$, and where $\mathbb{P}_X$ is the uniform distribution on $[0,1]^d$.

Together with Theorem 3.7, the statement in equation (13) shows that the F-UCB policy is near minimax optimal in terms of maximal expected regret. Similarly, together with Theorem 3.8, the lower bound in equation (14) proves that the F-UCB policy assigns the minimal number of suboptimal treatments.

To prove (13), we first use the margin condition to lower-bound the expected regret by the expected number of false assignments (cf. Lemma C.1). This strategy is similar to the one used by Rigollet and Zeevi (2010), who target the conditional mean functional. Their proof of the lower bound on maximal expected regret when targeting the conditional mean is based on considering joint distributions of (Y_t, X_t) for which the conditional distribution of Y_t given X_t is Bernoulli distributed with the success probability being a function of X_t. Thus, they obtain a useful lower bound over all sets $\mathscr{D}$ containing such distributions. Depending on the functional considered, however, Bernoulli distributions may not create sufficient variation in the functional to get good lower bounds. Furthermore, Bernoulli distributions may not be contained in $\mathscr{D}$, if, e.g., the latter does not contain discrete distributions, in which case a lower bound derived for Bernoulli distributions is not informative. For these two reasons, we have tailored the lower bounds in Theorem 3.9 toward the functional and parameter space $\mathscr{D}$ under consideration. In particular, Step 2 of the proof of Theorem 3.9 establishes, for *any* functional T satisfying Assumption 2.6, the existence of a family of joint distributions of (Y_t, X_t) over which any policy must incur high maximal expected regret. This family is, of course, a subset of those over which the supremum is taken in Theorem 3.9. We stress that a family of joint distributions that results in useful lower bounds for one functional may not give informative lower bounds for other functionals. This motivates our general construction which is applicable to a large set of functionals.

4. SIMULATION STUDY

In this section, we investigate the performance of Policy 1 by simulations. The functional of interest is the Gini welfare

$$\mathsf{T}(F) = \int x dF(x) - \frac{1}{2} \int \int |x_1 - x_2| \, dF(x_1) dF(x_2) \qquad \textbf{(15)}$$

(cf. Sen, 1974). When targeting this functional, a treatment is favorable if it has a high mean and low dispersion. It follows from Appendix D of Kock et al. (2022) that Assumption 2.3 is satisfied with $\mathscr{D} = D_{cdf}([a, b])$, which is trivially convex, and with constant $C = 2(b - a)$. We study the setting of a single covariate $x \in [0, 1]$ (i.e., $d = 1$) and $K = 2$ treatments. The covariate is uniformly distributed on $[0, 1]$.

Treatment 1 follows a Beta distribution with shape parameters $x + 1$ and 1, whereas Treatment 2 is Beta distributed with shape parameters $2 - x$ and 1.[9] Figure 1 shows the Gini welfare of the two treatments as a function of the covariate x: for small values of x, Treatment 2 is best, whereas Treatment 1 yields the highest welfare for large values of x. By construction, the two treatments have identical outcome distributions for $x = 0.5$. A routine calculation shows that Assumption 2.8 is satisfied with $\gamma = 1$.

We use $n = 50,000$ and approximate the expected regret by an average over $1,000$ Monte Carlo replications. Policy 1 is implemented as in Corollary 3.2, resulting in a partition of $[0, 1]$ into 37 intervals of equal length (and near minimax optimal regret). Specifically,

$$B_{n,j} = \left[\frac{j-1}{37}, \frac{j}{37}\right) \quad \text{for } j = 1, \ldots, 36, \qquad \text{and} \qquad B_{n,37} = \left[\frac{36}{37}, 1\right]$$

[9]The Beta distribution is a natural choice in the context of the Gini welfare measure as it has a long history in modeling income distributions (see, for example, Thurow, 1970; McDonald, 1984; McDonald and Ransom, 2008).

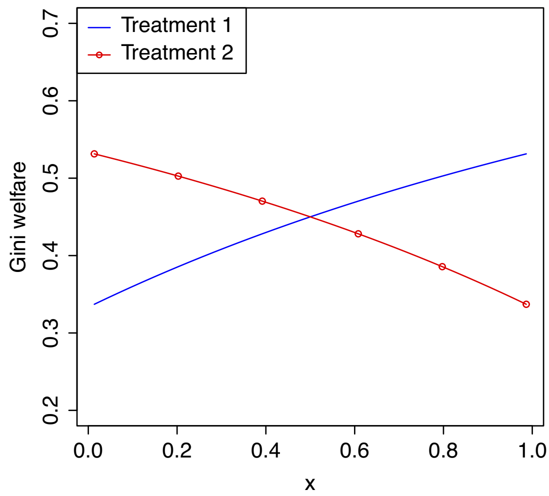

FIGURE 1. Gini welfare of Treatments 1 and 2 as a function of the covariate x.

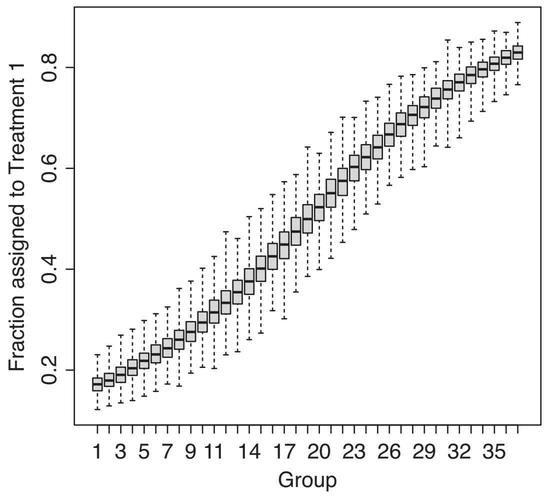

FIGURE 2. Boxplots indicating the fraction of subjects assigned to Treatment 1 in each group over all 1,000 replications.

such that larger values of x correspond to larger group numbers. The maximization step in Policy 1 causes no complications as it merely locates the maximum of $K = 2$ real numbers.

Figure 2 indicates the fraction of subjects that were assigned to Treatment 1 in each of the 37 groups, i.e., the fraction of subjects assigned to Treatment 1 as a

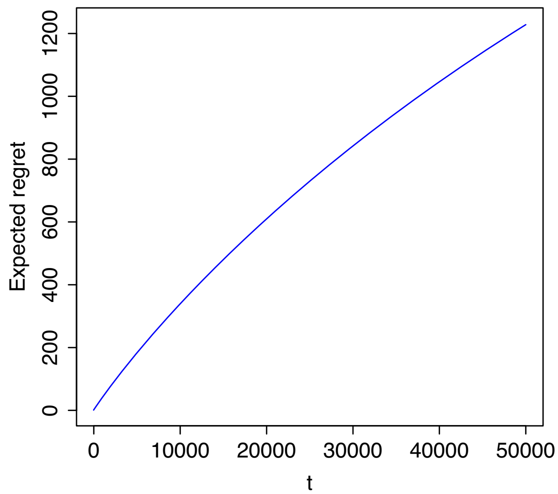

FIGURE 3. Expected regret as a function of $t \in \{1, \ldots, n\}$.

function of x. Among the subjects with large x, a much higher share is assigned to Treatment 1. This is desirable in light of Figure 1 as Treatment 1 yields the highest welfare for such subjects and vice versa for subjects with small values of x.

Finally, Figure 3 shows the expected regret as a function of t. As anticipated by our theoretical results, the expected regret accumulates at a decreasing rate.

5. CONCLUSION

In the present paper, we have established lower and upper bounds on maximal expected regret in a functional sequential assignment problem with covariates. Targeting distributional characteristics beyond the mean is important for policy-makers who are concerned about inequality, welfare, or poverty implications of their decisions.

In practice, the environment and outcome distributions of treatments may change over time as the treatment is rolled out. Therefore, fruitful avenues to generalize our results include the development of policies allowing for nonstationary environments, i.e., relaxing the i.i.d. assumption.

Appendix A. Auxiliary Results

We shall use similar notational conventions as discussed in Appendix A of Kock et al. (2022). We repeat them here for the convenience of the reader: The (unique) probability measure on the Borel sets of $\mathbb{R}$ corresponding to a cdf F will be denoted by μ_F (cf., e.g., Folland, 1999, p. 35). We employ standard notation and terminology concerning stochastic kernels and their semidirect products as discussed, e.g., in Appendix A.3 of Liese and Miescke (2008) (cf., in particular, their equation (A.3)). The random variables and vectors appearing in the proofs are defined on an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with corresponding expectation $\mathbb{E}$. This underlying probability space is assumed to be rich enough to support all random variables we work with. A generic element of Ω shall be denoted by ω. For a definition and proofs of elementary properties of the Kullback–Leibler divergence $\mathsf{KL}(P,Q)$ between two probability measures P and Q, we refer to Tsybakov (2009). We use the following general version of a chain rule for Kullback–Leibler divergences. A proof can be found in Appendix B of Kock et al. (2022).

LEMMA A.1 ("Chain rule" for Kullback–Leibler divergence). *Let $(\mathcal{X}, \mathfrak{A})$ and $(\mathcal{Y}, \mathfrak{B})$ be measurable spaces. Suppose that $\mathfrak{B}$ is countably generated. Let $\mathsf{A}, \mathsf{B} : \mathcal{B} \times \mathcal{X} \to [0,1]$ be stochastic kernels, and let P and Q be probability measures on $(\mathcal{X}, \mathfrak{A})$. Then,*

$$\begin{aligned}\mathsf{KL}(\mathsf{A} \otimes P, \mathsf{B} \otimes Q) &= \int_{\mathcal{X}} \mathsf{KL}(\mathsf{A}(\cdot, x), \mathsf{B}(\cdot, x)) dP(x) + \mathsf{KL}(P, Q) \\ &= \mathsf{KL}(\mathsf{A} \otimes P, \mathsf{B} \otimes P) + \mathsf{KL}(P, Q). \end{aligned} \tag{A.1}$$

We begin by establishing two auxiliary results that will be useful in the proofs of Theorems 3.1 and 3.7. For $n \in \mathbb{N}$, let $B_{n,1}, \ldots, B_{n,M}$ be a partition of $[0,1]^d$, where every $B_{n,j}$ is Borel measurable. Given such a partition, for every j such that $\mathbb{P}_X(B_{n,j}) > 0$, we shall denote by $F^*_{n,j}$ an element of $\{F^i_{n,j} : i = 1, \ldots, K\}$ (see equation (7) for a definition of $F^i_{n,j}$) such that $\mathsf{T}(F^*_{n,j}) = \max_{i \in \mathcal{I}} \mathsf{T}(F^i_{n,j})$. Furthermore, we often write $\pi_{n,t}(X_t)$ instead of $\pi_{n,t}(X_t, Z_{t-1}, G_t)$ in many places throughout the appendix.

LEMMA A.2. *Suppose that Assumptions 2.3 and 2.8 are satisfied (the latter with $\gamma \in (0,1]$ and $L > 0$), and assume that the inclusion in equation (2) holds. Let $B_{n,1}, \ldots, B_{n,M}$ be a partition of $[0,1]^d$, where every $B_{n,j}$ is Borel measurable. As in the statement of Theorem 3.1, we let $V_{n,j} = \sup_{x_1, x_2 \in B_{n,j}} \|x_1 - x_2\|$. Then, for every $i \in \{1, \ldots, K\}$, every $j \in \{1, \ldots, M\}$, and every pair x and $\tilde{x} \in B_{n,j}$, we have*

$$|\mathsf{T}(F^i(\cdot, x)) - \mathsf{T}(F^i(\cdot, \tilde{x}))| \le CLV^\gamma_{n,j} \quad \textit{and} \quad |\mathsf{T}\big(F^{\pi^\star(x)}(\cdot, x)\big) - \mathsf{T}\big(F^{\pi^\star(\tilde{x})}(\cdot, \tilde{x})\big)| \le CLV^\gamma_{n,j}; \tag{A.2}$$

furthermore, if $\mathbb{P}_X(B_{n,j}) > 0$ holds, then

$$|\mathsf{T}(F^i_{n,j}) - \mathsf{T}(F^i(\cdot, x))| \le CLV^\gamma_{n,j} \quad \textit{and} \quad |\mathsf{T}(F^{\pi^\star(x)}(\cdot, x)) - \mathsf{T}(F^*_{n,j})| \le CLV^\gamma_{n,j}. \tag{A.3}$$

Proof. Fix i, j, x, and $\tilde{x}$ as in the statement of the lemma. By Assumption 2.8,

$$||F^i(\cdot, x) - F^i(\cdot, \tilde{x})||_\infty \le L||x - \tilde{x}||^\gamma \le LV^\gamma_{n,j}. \tag{A.4}$$

Assumption 2.3 and (2) thus imply the first inequality in (A.2), and the second follows from

$$\begin{aligned}|\mathsf{T}\big(F^{\pi^\star(x)}(\cdot, x)\big) - \mathsf{T}\big(F^{\pi^\star(\tilde{x})}(\cdot, \tilde{x})\big)| &= |\max_{i \in \mathcal{I}} \mathsf{T}(F^i(\cdot, x)) - \max_{i \in \mathcal{I}} \mathsf{T}(F^i(\cdot, \tilde{x}))| \\ &\le \max_{i \in \mathcal{I}} |\mathsf{T}(F^i(\cdot, x)) - \mathsf{T}(F^i(\cdot, \tilde{x}))| \le CLV^\gamma_{n,j}.\end{aligned}$$

Next, assume that $\mathbb{P}_X(B_{n,j}) > 0$. For every $y \in \mathbb{R}$, from equation (A.4), we obtain

$$|F^i_{n,j}(y) - F^i(y,x)| \leq \frac{1}{\mathbb{P}_X(B_{n,j})} \int_{B_{n,j}} |F^i(y,s) - F^i(y,x)| d\mathbb{P}_X(s) \leq LV^{\gamma}_{n,j}.$$

The first inequality in (A.3) is now a direct consequence of Assumption 2.3 and (2) (noting that $F^i_{n,j} \in D_{cdf}([a,b])$), and the second inequality follows via

$$\begin{aligned} |\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^*_{n,j})| &= \left| \max_{i\in\mathcal{I}} \mathsf{T}(F^i(\cdot,x)) - \max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j}) \right| \\ &\leq \max_{i\in\mathcal{I}} |\mathsf{T}(F^i(\cdot,x)) - \mathsf{T}(F^i_{n,j})|. \end{aligned} \tag{A.5}$$

□

LEMMA A.3. *Suppose that Assumption 2.3 is satisfied and that $\mathscr{D}$ is convex. Suppose further that $\mathbb{P}_{Y,X}$ is such that equation (2) holds, and that Assumption 2.8 is satisfied. Then, for every Borel set $B \subseteq [0,1]^d$ that satisfies $\mathbb{P}_X(B) > 0$ and every $i = 1,\dots,K$, the cdf*

$$G_i := \mathbb{P}_X(B)^{-1} \int_B F^i(\cdot,x) d\mathbb{P}_X(x) \tag{A.6}$$

is an element of the closure of $\mathscr{D} \subseteq D_{cdf}([a,b])$ w.r.t. $\|\cdot\|_\infty$.

Proof. Let $i \in \{1,\dots,K\}$. We construct a sequence of convex combinations of (finitely many) elements of $\mathscr{D}$ that converges to G_i in $\|\cdot\|_\infty$-distance: To this end, let $B_{m,1},\dots,B_{m,l_m}$, for $m \in \mathbb{N}$, be a triangular array of partitions of $[0,1]^d$ into nonempty Borel subsets, such that the maximal diameter $v_m := \sup_{i=1,\dots,l_m} \sup_{x_1,x_2 \in B_{m,i}} \|x_1 - x_2\| \to 0$ as $m \to \infty$. For simplicity, define the probability measure $\mathbb{P}^*$ on the Borel sets of $\mathbb{R}^d$ by $\mathbb{P}^*(A) = \mathbb{P}_X(A \cap B)/\mathbb{P}_X(B)$. Write

$$G_i = \int F^i(\cdot,x) d\mathbb{P}^*(x) = \sum_{j=1}^{l_m} \int_{B_{m,j}} F^i(\cdot,x) d\mathbb{P}^*(x). \tag{A.7}$$

For every m and every j, pick an $x_{m,j} \in B_{m,j}$. Note that $F^i(\cdot,x_{m,j}) \in \mathscr{D}$ by equation (2). From Assumption 2.8, we know that, for any $x \in B_{m,j}$ we have $\|F^i(\cdot,x_{m,j}) - F^i(\cdot,x)\|_\infty, \leq L\|x_{m,j} - x\|^\gamma \leq Lv_m^\gamma$. Thus,

$$\|G_i - \sum_{j=1}^{l_m} \mathbb{P}^*(B_{m,j}) F^i(\cdot,x_{m,j})\|_\infty \leq \sum_{j=1}^{l_m} \int_{B_{m,j}} \|F^i(\cdot,x) - F^i(\cdot,x_{m,j})\|_\infty d\mathbb{P}^*(x) \leq Lv_m^\gamma \to 0. \tag{A.8}$$

□

Appendix B. Proofs of Results in Section 2

B.1. Proof of Theorem 2.7

Because Assumption 2.8 (for any $\gamma \in (0,1]$ and any $L > 0$) implies the assumption in equation (5), the statement follows immediately from the lower bound in equation (13) in Theorem 3.9 upon letting $\gamma \to 0$.

B.2. Proof of Theorem 2.9

If $\min(\mathbb{P}_X(A_1), \mathbb{P}_X(A_2)) = 0$, then the statement in the theorem trivially holds. Hence, assume that $p := \min(\mathbb{P}_X(A_1), \mathbb{P}_X(A_2)) > 0$. Let $n \in \mathbb{N}$, and let π be a policy that ignores covariates, i.e., as described before Theorem 2.9. We write $\pi_{n,t} = \pi_t$. Fix a randomization measure $\mathbb{P}_G$.

As a preparation, for every $m \in \mathbb{N}$, define

$$A_{1,m} := \{x \in [0,1]^d : \mathsf{T}(F^1(\cdot,x)) > m^{-1} + \mathsf{T}(F^2(\cdot,x))\},$$
$$A_{2,m} := \{x \in [0,1]^d : \mathsf{T}(F^1(\cdot,x)) + m^{-1} < \mathsf{T}(F^2(\cdot,x))\}.$$

The sets A_1, A_2 and $A_{1,m}, A_{2,m}$ for $m \in \mathbb{N}$ are Borel measurable, because Assumptions 2.3 and 2.8 together with equation (2) imply the continuity of $x \mapsto \mathsf{T}(F^i(\cdot,x))$ for $i = 1,2$. Note that $A_{i,m} \subseteq A_{i,m+1}$ and $\bigcup_{m\in\mathbb{N}} A_{i,m} = A_i$ hold for $i = 1,2$. Hence, as $m \to \infty$, $\mathbb{P}_X(A_{i,m}) \to \mathbb{P}_X(A_i)$ for $i = 1,2$. Because of $p > 0$, we can conclude the existence of an $\bar{m} \in \mathbb{N}$ such that $p_{\bar{m}} := \min(\mathbb{P}_X(A_{1,\bar{m}}), \mathbb{P}_X(A_{2,\bar{m}})) > p/2$. To prove the inequality in equation (6), note that by definition, and since π is a policy that does not depend on covariates, i.e., the tth assignment only depends on the previously observed outcomes and randomizations, $\tilde{Z}_{t-1}$ and a novel randomization G_t, we have (cf. the discussion and notation discussed right before the statement of Theorem 2.9) that

$$R_n(\pi) = \sum_{t=1}^{n} \left|\mathsf{T}\big(F^1(\cdot,X_t)\big) - \mathsf{T}\big(F^2(\cdot,X_t)\big)\right| \mathbb{1}_{\{\pi^\star(X_t) \neq \tilde{\pi}_t(\tilde{Z}_{t-1}, G_t)\}}.$$

Note furthermore that

$$\Big[\{X_t \in A_{1,\bar{m}}\} \cap \{\tilde{\pi}_t(\tilde{Z}_{t-1}, G_t) \neq 1\}\Big] \cup \Big[\{X_t \in A_{2,\bar{m}}\} \cap \{\tilde{\pi}_t(\tilde{Z}_{t-1}, G_t) \neq 2\}\Big] \tag{B.1}$$

$$\subseteq \{\pi^\star(X_t) \neq \tilde{\pi}_t(\tilde{Z}_{t-1}, G_t)\}, \tag{B.2}$$

where the union in the first line is a disjoint union. Hence,

$$R_n(\pi) \geq \bar{m}^{-1} \sum_{t=1}^{n} \Big(\mathbb{1}_{A_{1,\bar{m}}}(X_t)\mathbb{1}_{\{\tilde{\pi}_t(\tilde{Z}_{t-1}, G_t) \neq 1\}} + \mathbb{1}_{A_{2,\bar{m}}}(X_t)\mathbb{1}_{\{\tilde{\pi}_t(\tilde{Z}_{t-1}, G_t) \neq 2\}}\Big).$$

Since X_t is independent of $\tilde{Z}_{t-1}$ and G_t, the law of iterated expectations implies that $\mathbb{E}(R_n(\pi)) \geq np/(2\bar{m})$.

Appendix C. Proofs of Results in Section 3

C.1. Proof of Theorem 3.1

Fix $n \in \mathbb{N}$, and let $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1, \ldots, n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2), and Assumption 2.8 with L and γ. Because n is fixed, we abbreviate $B_{n,j} = B_j$, $V_{n,j} = V_j$, $M(n) = M$, and denote $\bar{\pi}_{n,t} = \bar{\pi}_t$. First, we decompose $R_n(\bar{\pi}) = \sum_{j=1}^{M} \tilde{R}_j(\bar{\pi})$, where

$$\tilde{R}_j(\bar{\pi}) := \sum_{t=1}^{n} \left[\mathsf{T}\left(F^{\pi^\star(X_t)}(\cdot, X_t)\right) - \mathsf{T}\left(F^{\bar{\pi}_t(X_t)}(\cdot, X_t)\right)\right] \mathbb{1}_{\{X_t \in B_j\}}, \tag{C.1}$$

where, as often done in the present section, we dropped the argument Z_{t-1} from $\bar{\pi}_t$. Note furthermore that the policy does not rely on an external randomization G_t, which is therefore suppressed in the notation as well.

Note first that the boundedness of T on $\mathscr{D}$ (cf. Assumption 2.3) implies $\mathbb{E}(\tilde{R}_j(\bar{\pi})) = 0$ for every j such that $\mathbb{P}_X(B_j) = 0$. Hence, we now fix an index $j \in \{1, \ldots, M\}$ such that $\mathbb{P}_X(B_j) > 0$. Then, recalling the definition of $F^i_{n,j}$ in equation (7), which we here abbreviate as F^i_j, each summand in (C.1) can be written as

$$\left[\mathsf{T}(F^{\pi^\star(X_t)}(\cdot, X_t)) - \mathsf{T}(F^*_j) + \mathsf{T}(F^*_j) - \mathsf{T}(F_j^{\bar{\pi}_t(X_t)}) + \mathsf{T}(F_j^{\bar{\pi}_t(X_t)}) - \mathsf{T}(F^{\bar{\pi}_t(X_t)}(\cdot, X_t))\right] \mathbb{1}_{\{X_t \in B_j\}}, \tag{C.2}$$

which, by Lemma A.2, is not greater than $\mathsf{T}(F^*_j) - \mathsf{T}(F_j^{\bar{\pi}_t(X_t)}) + 2CLV_j^{\gamma}$, and where F^*_j was defined just before Lemma A.2. Therefore, we obtain

$$\tilde{R}_j(\bar{\pi}) \leq \sum_{t=1}^{n} \left[\mathsf{T}\left(F^*_j\right) - \mathsf{T}\left(F_j^{\bar{\pi}_t(X_t)}\right)\right] \mathbb{1}_{\{X_t \in B_j\}} + 2CLV_j^{\gamma} \sum_{t=1}^{n} \mathbb{1}_{\{X_t \in B_j\}}. \tag{C.3}$$

Obviously, $\mathbb{E}(\sum_{t=1}^{n} \mathbb{1}_{\{X_t \in B_j\}}) = n\mathbb{P}_X(B_j)$. Hence, to prove the theorem, it remains to show that, for $c = c(\beta, C)$ as defined in the statement of the theorem, it holds that

$$\mathbb{E}\left(\sum_{t=1}^{n} \left[\mathsf{T}\left(F^*_j\right) - \mathsf{T}\left(F_j^{\bar{\pi}_t(X_t)}\right)\right] \mathbb{1}_{\{X_t \in B_j\}}\right) \leq c\sqrt{Kn\mathbb{P}_X(B_j)\overline{\log}(n\mathbb{P}_X(B_j))}. \tag{C.4}$$

To this end, we will use a conditioning argument in combination with Theorem 4.1 in Kock et al. (2022). Define, for every $v = (v_1, \ldots, v_n) \in \{0, 1\}^n$, the event

$$\Omega(v) := \{\omega : \mathbb{1}_{\{X_t \in B_j\}}(\omega) = v_t \text{ for } t = 1, \ldots, n\}, \tag{C.5}$$

and denote $f := \sum_{t=1}^{n} [\mathsf{T}(F^*_j) - \mathsf{T}(F_j^{\bar{\pi}_t(X_t)})] \mathbb{1}_{\{X_t \in B_j\}}$. Then,

$$\mathbb{E}(f) = \sum_{v \in \{0,1\}^n} \mathbb{E}(\mathbb{1}_{\Omega(v)} f) = \sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v)) \mathbb{E}(f | \Omega(v)), \tag{C.6}$$

where (as usual) we define

$$\mathbb{E}(f | \Omega(v)) := \begin{cases} \mathbb{P}^{-1}(\Omega(v)) \mathbb{E}(\mathbb{1}_{\Omega(v)} f), & \text{if } \mathbb{P}(\Omega(v)) > 0, \\ 0, & \text{else.} \end{cases} \tag{C.7}$$

Fix $v \neq 0$. Denote the elements of $\{s : v_s = 1\}$ by $t_1, \ldots, t_{\bar{m}}$, ordered from smallest to largest. On the event $\Omega(v)$, i.e., for every $\omega \in \Omega(v)$, we can use the definition of $\bar{\pi}$ (cf. the description of the F-UCB policy with covariates of display Policy 1) to rewrite

$$f = \sum_{s=1}^{\bar{m}} \left[\mathsf{T}(F_j^*) - \mathsf{T}\left(F_j^{\hat{\pi}_s(W^{s-1})} \right) \right], \tag{C.8}$$

where $\hat{\pi}$ is the F-UCB policy from Kock et al. (2022), and where W^s is defined recursively via $W^s = (Y_{\hat{\pi}_{s-1}(W^{s-1}), t_s}, W^{s-1})$ with W^0 the empty vector (cf. also the discussion before our Policy 1). Hence, for $\omega \in \Omega(v)$, f is a function of $(Y_{t_1}, \ldots, Y_{t_{\bar{m}}})$, i.e., $f = H(Y_{t_1}, \ldots, Y_{t_{\bar{m}}})$, say. We conclude that

$$\mathbb{E}(f|\Omega(v)) = \mathbb{E}\left(H(Y_{t_1}, \ldots, Y_{t_{\bar{m}}})|\Omega(v) \right) = \mathbb{E}^v(H(Y_{t_1}, \ldots, Y_{t_{\bar{m}}})), \tag{C.9}$$

where the probability measure $\mathbb{P}^v$ corresponding to $\mathbb{E}^v$ is defined as the $\mathbb{P}$-measure with density $\mathbb{P}^{-1}(\Omega(v))\mathbb{1}_{\Omega(v)}$. Note that, for $A_i \in \mathcal{B}(\mathbb{R}^K)$, for $i = 1, \ldots, \bar{m}$, we have that $\mathbb{P}^v(Y_{t_1} \in A_1, \ldots, Y_{t_{\bar{m}}} \in A_{\bar{m}})$ equals

$$\mathbb{P}^{-1}(\Omega(v))\mathbb{P}\left(Y_{t_1} \in A_1, \ldots, Y_{t_{\bar{m}}} \in A_{\bar{m}}, \Omega(v) \right) = \prod_{s=1}^{\bar{m}} \frac{\mathbb{P}(Y_{t_s} \in A_s, X_{t_s} \in B_j)}{\mathbb{P}(X_{t_s} \in B_j)} \tag{C.10}$$

$$= \prod_{s=1}^{\bar{m}} \mathbb{P}(Y_{t_s} \in A_s | \{X_{t_s} \in B_j\}). \tag{C.11}$$

Hence, the image measure $\mathbb{P}^v \circ (Y_{t_1}, \ldots, Y_{t_{\bar{m}}})$ is the $\bar{m}$-fold product of $\mathbb{Q}(\cdot) := \mathbb{P}(Y_1 \in .|\{X_1 \in B_j\})$. For i.i.d. random K-vectors $Y_1^*, \ldots, Y_{\bar{m}}^*$, say, each with distribution $\mathbb{Q}$, it hence follows from the definition of H that

$$\mathbb{E}(H(Y_{t_1}, \ldots, Y_{t_m})|\Omega(v)) = \mathbb{E}(H(Y_1^*, \ldots, Y_{\bar{m}}^*)) = \mathbb{E}\left(\sum_{s=1}^{\bar{m}} \left[\mathsf{T}(F_j^*) - \mathsf{T}\left(F_j^{\hat{\pi}_s(Z_{s-1}^*)} \right) \right] \right), \tag{C.12}$$

where $Z_s^* = (Y_{\hat{\pi}_s(Z_{s-1}^*), s}^*, \ldots, Z_{s-1}^*)$ (and where Z_0^* is the empty vector). The rth marginal of $\mathbb{Q}$ has cdf F_j^r, which by Lemma A.3 is an element of the closure of $\mathscr{D} \subseteq D_{cdf}([a,b])$ w.r.t. $\|\cdot\|_\infty$, which we here denote as $\mathrm{cl}(\mathscr{D})$. Therefore, it now follows from Theorem 4.1 in Kock et al. (2022), applied with $\mathrm{cl}(\mathscr{D})$ (cf. their Remark 2.3) and with "$n = \bar{m}$," that the quantity in the previous display, and thus $\mathbb{E}(f|\Omega(v))$, is not greater than $c\sqrt{K\bar{m}\overline{\log}(\bar{m})}$. From (C.6) (noting that f vanishes on $\Omega(0)$), we see that

$$\mathbb{E}(f) \leq c \sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v))\sqrt{K\bar{m}\overline{\log}(\bar{m})}. \tag{C.13}$$

Recall that $\bar{m} = \sum_{s=1}^n v_s$. Hence, we can interpret $\bar{m}$ as a random variable on the set $\{0,1\}^n$, equipped with the probability mass function $p(v) = \mathbb{P}(\Omega(v))$. Obviously, this random variable is Bernoulli-distributed with success probability $\mathbb{P}_X(B_j)$ and "sample size" n. Thus, its expectation is $n\mathbb{P}_X(B_j)$. It remains to observe that the function h defined via $x \mapsto (Kx\overline{\log}(x))^{0.5}$ is concave on $[0,\infty)$, allowing us to apply Jensen's inequality to upper-bound the right-hand side in the previous display by $ch(n\mathbb{P}_X(B_j))$, which establishes the statement in equation (C.4).

C.2. Proof of Corollary 3.2

Fix $n \in \mathbb{N}$, and let $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1, \ldots, n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2), Assumption 2.1 with $\underline{c}$ and $\overline{c}$, and Assumption 2.8 with L and γ. We shall apply Theorem 3.1 to get an upper bound on $\mathbb{E}[R_n(\bar{\pi})]$. The specific partition results in $M(n) = P^d$ and $V_{n,j} = \sqrt{d}P^{-1}$, where $P = \lceil n^{1/(2\gamma+d)} \rceil$. Furthermore, from Assumption 2.1, we obtain $\mathbb{P}_X(B_{n,j}) \leq \overline{c}P^{-d}$. Therefore, equation (8) implies the upper bound

$$\mathbb{E}[R_n(\bar{\pi})] \leq c(\beta, C)\sqrt{Kn\bar{c}P^d\overline{\log}(n\bar{c}P^{-d})} + 2CL(\sqrt{d}P^{-1})^\gamma n\bar{c}, \tag{C.14}$$

which (using monotonicity of $\overline{\log}$, and $\overline{\log}(xy) \leq \overline{\log}(x) + \overline{\log}(y)$ for positive x and y) is bounded from above by

$$c(\beta, C)\sqrt{K\bar{c}(1+\overline{\log}(\bar{c}))\overline{\log}(n)nP^d} + 2CLd^{\gamma/2}\bar{c}nP^{-\gamma} \leq c^*\left(\sqrt{K\overline{\log}(n)nP^d} + nP^{-\gamma}\right) \tag{C.15}$$

$$\leq c^*\sqrt{K\overline{\log}(n)}\left(\sqrt{nP^d} + nP^{-\gamma}\right), \tag{C.16}$$

where $c^* := \max[c(\beta, C)(\bar{c}(1 + \overline{\log}(\bar{c})))^{1/2}, 2CLd^{\gamma/2}\bar{c}]$. From $P^{-\gamma} \leq n^{-\gamma/(2\gamma+d)}$ and $P^d \leq 2^d n^{d/(2\gamma+d)}$, we obtain the bound

$$\mathbb{E}[R_n(\bar{\pi})] \leq (2^{d/2} + 1)c^*\sqrt{K\overline{\log}(n)}n^{1-\frac{\gamma}{2\gamma+d}}, \tag{C.17}$$

which proves the theorem.

C.3. Proof of Theorem 3.3

The statement follows from the first lower bound established in Theorem 3.9, upon setting $\alpha = \alpha(\varepsilon) = (2\gamma + d)\varepsilon/\gamma$ there; note that $\alpha(\varepsilon)$ is an element of $(0, 1)$ because $\varepsilon \in (0, \gamma/(2\gamma + d))$ holds by construction.

C.4. Proof of Theorem 3.7

Define $c_1 := 4CLd^{\gamma/2} + 1$. Recall that $P = \lceil n^{1/(2\gamma+d)} \rceil$. Note first that it suffices to establish the inequality in equation (11) for all n large enough ($n \geq n_0$, say), such that $c_1P^{-\gamma} \leq 1$ holds (this will allow us to apply Assumption 3.6 with $\delta = c_1P^{-\gamma}$ in the arguments below). To see this, note that, by Assumption 2.3, for all $n < n_0$, it holds (for all random vectors as in the statement of the theorem) that $\mathbb{E}[R_n(\pi)] \leq Cn_0$. Hence, once the claimed inequality in the theorem has been established for all $n \geq n_0$, the constant c in the statement of Theorem 3.7 can be chosen large enough to deal with the initial terms smaller than n_0. Hence, fix $n \geq n_0$. Because n is fixed, we abbreviate $B_{n,j} = B_j$, $V_{n,j} = V_j = \sqrt{d}P^{-1}$, and denote $\bar{\pi}_{n,t} = \bar{\pi}_t$.

Let $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, for $t = 1, \ldots, n$, where $\mathbb{P}_{Y,X}$ satisfies equation (2), Assumption 2.1 with $\underline{c}$ and $\bar{c}$, Assumption 2.8 with L and γ, and Assumption 3.6 with $\alpha \in (0, 1)$ and $C_0 > 0$. We establish $\mathbb{E}[R_n(\bar{\pi})] \leq cK\overline{\log}(n)n^{1-\frac{\gamma(1+\alpha)}{2\gamma+2}}$ for a constant that depends on the quantities indicated in the statement of the theorem in five steps.

Step 1: Decomposition of bins into different types. To obtain the desired upper bound, we shall treat three types of bins separately. An analogous division of bins was also used in Perchet and Rigollet (2013) to establish the properties of their successive elimination algorithm in a classic bandit problem targeting the distribution with the highest (conditional) mean. The bins are split into

$$\begin{aligned}\mathcal{J} &:= \left\{j \in \{1,\ldots,P^d\} : \exists\, \bar{x} \in B_j, \mathsf{T}(F^{\pi^\star(\bar{x})}(\cdot,\bar{x})) - \mathsf{T}(F^{\pi^\sharp(\bar{x})}(\cdot,\bar{x})) > c_1 P^{-\gamma}\right\},\\ \mathcal{J}_s &:= \left\{j \in \{1,\ldots,P^d\} : \exists\, \bar{x} \in B_j, \mathsf{T}(F^{\pi^\star(\bar{x})}(\cdot,\bar{x})) = \mathsf{T}(F^{\pi^\sharp(\bar{x})}(\cdot,\bar{x}))\right\},\\ \mathcal{J}_w &:= \left\{j \in \{1,\ldots,P^d\} : 0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x)) \le c_1 P^{-\gamma} \text{ for all } x \in B_j\right\}.\end{aligned} \tag{C.18}$$

The bins corresponding to indices in $\mathcal{J}$, $\mathcal{J}_s$, and $\mathcal{J}_w$ will be referred to as "well-behaved," "strongly ill-behaved," and "weakly ill-behaved" bins, respectively. Note that $\mathcal{J}_w$ and $\mathcal{J} \cup \mathcal{J}_s$ are clearly disjoint. That $\mathcal{J}$ and $\mathcal{J}_s$ are disjoint is shown in Step 2 below. Hence, the sets of bins corresponding to indices in $\mathcal{J}$, $\mathcal{J}_s$, and $\mathcal{J}_w$ constitute a partition of the set of all P^d bins B_j, and we can thus write

$$\mathbb{E}(R_n(\bar{\pi})) = \sum_{j\in\mathcal{J}_s} \mathbb{E}(\tilde{R}_j(\bar{\pi})) + \sum_{j\in\mathcal{J}_w} \mathbb{E}(\tilde{R}_j(\bar{\pi})) + \sum_{j\in\mathcal{J}} \mathbb{E}(\tilde{R}_j(\bar{\pi})), \tag{C.19}$$

where, as in equation (C.1), we define

$$\tilde{R}_j(\bar{\pi}) := \sum_{t=1}^{n} \left[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big) - \mathsf{T}\big(F^{\bar{\pi}_t(X_t)}(\cdot,X_t)\big)\right] \mathbb{1}_{\{X_t\in B_j\}}. \tag{C.20}$$

Step 2: Strongly ill-behaved bins. For every $j \in \mathcal{J}_s$, by definition, there exists a $\bar{x} \in B_j$ such that $\mathsf{T}\big(F^{\pi^\star(\bar{x})}(\cdot,\bar{x})\big) = \mathsf{T}\big(F^{\pi^\sharp(\bar{x})}(\cdot,\bar{x})\big)$. From the definition of $\pi^\sharp$, it thus follows that $\mathsf{T}\big(F^{\pi^\star(\bar{x})}(\cdot,\bar{x})\big) = \mathsf{T}\big(F^{i}(\cdot,\bar{x})\big)$ for every $i \in \mathcal{I}$. Therefore, for every $x \in B_j$ and every $i \in \mathcal{I}$, Lemma A.2 yields

$$\begin{aligned}&\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{i}(\cdot,x))\\ &= \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{i}(\cdot,x)) - [\mathsf{T}(F^{\pi^\star(\bar{x})}(\cdot,\bar{x})) - \mathsf{T}(F^{i}(\cdot,\bar{x}))]\\ &\le 2CLd^{\gamma/2}P^{-\gamma} \le c_1 P^{-\gamma}.\end{aligned} \tag{C.21}$$

First of all, this shows that $\mathcal{J}$ and $\mathcal{J}_s$ are disjoint. Furthermore, from equations (C.20) and (C.21), we obtain

$$\sum_{j\in\mathcal{J}_s} \tilde{R}_j(\bar{\pi}) \le c_1 P^{-\gamma} \sum_{j\in\mathcal{J}_s}\sum_{t=1}^{n} \mathbb{1}_{\{X_t\in B_j\}} \mathbb{1}_{\{0<\mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t))\}} \tag{C.22}$$

$$\le c_1 P^{-\gamma} \sum_{t=1}^{n} \mathbb{1}_{\{0<\mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t)) \le c_1 P^{-\gamma}\}}. \tag{C.23}$$

From Condition 3.6, we, hence, obtain

$$\sum_{j\in\mathcal{J}_s}\mathbb{E}[\tilde{R}_j(\bar{\pi})] \le c_1 nP^{-\gamma}\mathbb{P}_X\big(0<\mathsf{T}\big(F^{\pi^\star(X)}(\cdot,X)\big)-\mathsf{T}\big(F^{\pi^\sharp(X)}(\cdot,X)\le c_1P^{-\gamma}\big) \le C_0c_1^{1+\alpha}nP^{-\gamma(1+\alpha)}. \tag{C.24}$$

Step 3: Weakly ill-behaved bins. Since $\{X_t\in B_j\}$ for $j\in\mathcal{J}_w$ are disjoint subsets of

$$\{0<\mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t))-\mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t))\le c_1P^{-\gamma}\},$$

we obtain from Condition 3.6, recall that $\mathbb{P}(X_t\in B_j)\ge\frac{\underline{c}}{P^d}$, that

$$|\mathcal{J}_w|\frac{\underline{c}}{P^d}\le\sum_{j\in\mathcal{J}_w}\mathbb{P}(X_t\in B_j)\le\mathbb{P}\big(0<\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big)-\mathsf{T}\big(F^{\pi^\sharp(X_t)}(\cdot,X_t)\big)\le c_1P^{-\gamma}\big) \tag{C.25}$$

$$\le C_0c_1^{\alpha}P^{-\gamma\alpha}, \tag{C.26}$$

which yields $|\mathcal{J}_w|\le(C_0c_1^\alpha/\underline{c})P^{d-\gamma\alpha}$. Using (C.3) and (C.4) with $V_j=\sqrt{d}P^{-1}$ and $\mathbb{P}_X(B_j)\le\bar{c}P^{-d}$, we obtain (by similar arguments as in Section C.2)

$$\mathbb{E}[\tilde{R}_j(\bar{\pi})]\le c'\left(\sqrt{Kn\overline{\log}(n)}P^{-d/2}+nP^{-\gamma-d}\right), \tag{C.27}$$

where c' depends on $d,L,\gamma,\bar{c},C,\beta$, but *not* on n. Combining (C.27) with $|\mathcal{J}_w|\le(C_0c_1^\alpha/\underline{c})P^{d-\gamma\alpha}$ leads to

$$\sum_{j\in\mathcal{J}_w}\mathbb{E}[\tilde{R}_j(\bar{\pi})]\le c''\big(\sqrt{Kn\overline{\log}(n)}P^{d/2-\gamma\alpha}+nP^{-\gamma(1+\alpha)}\big), \tag{C.28}$$

where c'' depends on $d,L,\gamma,\underline{c},\bar{c},C,C_0,\alpha,\beta$, but *not* on n.

Step 4: Well-behaved bins. For every $j\in\mathcal{J}$, let $x_j\in B_j$ be such that

$$\mathsf{T}(F^{\pi^\star(x_j)}(\cdot,x_j))-\mathsf{T}(F^{\pi^\sharp(x_j)}(\cdot,x_j))>c_1P^{-\gamma}. \tag{C.29}$$

Next, define the following sets of indices ("corresponding to the optimal and suboptimal treatments given x_j"):

$$I_j^\star:=\{i\in\mathcal{I}:\mathsf{T}\big(F^{\pi^\star(x_j)}(\cdot,x_j)\big)=\mathsf{T}(F^i(\cdot,x_j))\},$$
$$I_j^0:=\{i\in\mathcal{I}:\mathsf{T}\big(F^{\pi^\star(x_j)}(\cdot,x_j)\big)-\mathsf{T}(F^i(\cdot,x_j))>c_1P^{-\gamma}\}.$$

Clearly, $\pi^\star(x_j)\in I_j^\star$ and $\pi^\sharp(x_j)\in I_j^0$ (cf. (C.29)). Hence, $I_j^\star$ and I_j^0 define a nontrivial partition of $\mathcal{I}$. For every $j\in\mathcal{J}$, we can thus decompose $\tilde{R}_j(\bar{\pi})$ defined in equation (C.20) as the sum of

$$\tilde{R}_{j,I_j^\star}(\bar{\pi}) := \sum_{i\in I_j^\star}\sum_{t=1}^{n}\Big[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big)-\mathsf{T}\big(F^i(\cdot,X_t)\big)\Big]\mathbb{1}_{\{X_t\in B_j\}}\mathbb{1}_{\{\bar{\pi}_t(X_t)=i\}},$$
$$\tilde{R}_{j,I_j^0}(\bar{\pi}) := \sum_{i\in I_j^0}\sum_{t=1}^{n}\Big[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big)-\mathsf{T}\big(F^i(\cdot,X_t)\big)\Big]\mathbb{1}_{\{X_t\in B_j\}}\mathbb{1}_{\{\bar{\pi}_t(X_t)=i\}}. \tag{C.30}$$

Step 4a: A bound for $\mathbb{E}(\tilde{R}_{j,I_j^\star}(\bar{\pi}))$. For any $i\in I_j^\star$ and every $x\in B_j$ satisfying $\mathsf{T}(F^{\pi^\star(x)}(\cdot,x))\neq\mathsf{T}(F^i(\cdot,x))$, the triangle inequality, the definition of $\pi^\sharp$, and Lemma A.2 yield

$$\begin{aligned}
0 &< \mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x))\\
&\le \mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^i(\cdot,x))\\
&= \mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^{\pi^\star(x_j)}(\cdot,x_j))+\mathsf{T}(F^i(\cdot,x_j))-\mathsf{T}(F^i(\cdot,x))\\
&\le 2CLd^{\gamma/2}P^{-\gamma}\le c_1P^{-\gamma},
\end{aligned}$$

the last inequality following from $c_1=4CLd^{\gamma/2}+1$. But this means (applying the inequality chain in the previous display twice) that, for any $i\in I_j^\star$ and every $x\in B_j$,

$$\mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^i(\cdot,x))\le c_1P^{-\gamma}\mathbb{1}_{\{v:0<\mathsf{T}(F^{\pi^\star(v)}(\cdot,v))-\mathsf{T}(F^{\pi^\sharp(v)}(\cdot,v))\le c_1P^{-\gamma}\}}(x). \tag{C.31}$$

We deduce that

$$\begin{aligned}
\mathbb{E}[\tilde{R}_{j,I_j^\star}(\bar{\pi})] &\le \mathbb{E}\sum_{t=1}^{n}c_1P^{-\gamma}\mathbb{1}_{\{0<\mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t))-\mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t))\le c_1P^{-\gamma}\}}\mathbb{1}_{\{X_t\in B_j\}}\\
&\le nc_1P^{-\gamma}q_j,
\end{aligned} \tag{C.32}$$

where $q_j:=\mathbb{P}(0<\mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t))-\mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t))\le c_1P^{-\gamma},X_t\in B_j)$, which is independent of t due to the X_t being identically distributed.

Step 4b: A bound for $\mathbb{E}(\tilde{R}_{j,I_j^0}(\bar{\pi}))$. By Lemma A.2, noting that $\mathbb{P}_X(B_j)>\underline{c}P^{-d}>0$, for every $x\in B_j$ and every $i\in I_j^0$, we have (abbreviating $F^i_{n,j}$ by F^i_j)

$$\mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^i(\cdot,x))\le\Big[\mathsf{T}(F_j^*)-\mathsf{T}(F_j^i)\Big]+c_1P^{-\gamma}, \tag{C.33}$$

from which it follows that

$$\mathbb{E}[\tilde{R}_{j,I_j^0}(\bar{\pi})]\le\sum_{i\in I_j^0}\Delta_j^i\mathbb{E}S(i,n,j)+c_1P^{-\gamma}\sum_{i\in I_j^0}\mathbb{E}S(i,n,j), \tag{C.34}$$

where, for every $i \in I_j^0$, we let $S(i,n,j) := \sum_{t=1}^n \mathbb{1}_{\{X_t \in B_j\}} \mathbb{1}_{\{\bar{\pi}_t(X_t)=i\}}$ and $\Delta_j^i := \mathsf{T}(F_j^*) - \mathsf{T}(F_j^i)$. We now claim that (this claim will be verified before moving to Step 4c below)

$$\mathbb{E}S(i,n,j) \le \frac{2C^2\beta \log(\bar{c}nP^{-d})}{[\Delta_j^i]^2} + \frac{\beta+2}{\beta-2}. \tag{C.35}$$

Define $\underline{\Delta}_j := \min_{i \in I_j^0} \Delta_j^i$. We note that $\underline{\Delta}_j > 0$ follows from inserting $x = x_j$ in equation (C.33), and from using the definition of I_j^0. Next, noting that $\max_{i \in I_j^0} \Delta_j^i \le 2C$ by Assumption 2.3, and combining equations (C.34) and (C.35), we obtain the bound

$$\mathbb{E}[\tilde{R}_{j,I_j^0}(\bar{\pi})] \le K\frac{2C^2\beta \log(\bar{c}nP^{-d})}{\underline{\Delta}_j}\left(1 + \frac{c_1P^{-\gamma}}{\underline{\Delta}_j}\right) + (c_1+2C)K\frac{\beta+2}{\beta-2}. \tag{C.36}$$

It remains to prove the claim in equation (C.35). To this end, we apply a conditioning argument as in the proof of Theorem 3.1. We shall now use some quantities (in particular, the sets $\Omega(v)$) that were defined in that proof: Note that

$$\mathbb{E}S(i,n,j) = \sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v))\mathbb{E}(S(i,n,j)|\Omega(v)). \tag{C.37}$$

Arguing as in the proof of Theorem 3.1, it is now easy to see that $\mathbb{E}(S(i,n,j)|\Omega(v))$ can be written as the expected number of times treatment i is selected in running the F-UCB policy $\hat{\pi}$ (without covariates) in a problem with $\bar{m} = \sum_{s=1}^n v_s$ (fixed) i.i.d. inputs with distribution $\mathbb{Q}$ (the marginals of which have a cdf that lies in the closure of $\mathscr{D}$ w.r.t. $\|\cdot\|_\infty$ as a consequence of Lemma A.3). We can, hence, (cf. Remark 2.3 in Kock et al. (2022)) apply the bound established in equation (C.19) of Appendix C of Kock et al. (2022), to the just mentioned problem, to obtain

$$\mathbb{E}(S(i,n,j)|\Omega(v)) \le \frac{2C^2\beta \log(\bar{m})}{[\Delta_j^i]^2} + \frac{\beta+2}{\beta-2}. \tag{C.38}$$

We can now combine the obtained inequality with equation (C.37) to see that

$$\mathbb{E}S(i,n,j) \le \sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v))\frac{2C^2\beta \log(\bar{m})}{[\Delta_j^i]^2} + \frac{\beta+2}{\beta-2}. \tag{C.39}$$

The claim in (C.35) now follows from Jensen's inequality, and (cf. the end of the proof of Theorem 3.1) $\sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v))\bar{m} \le \bar{c}nP^{-d}$.

Step 4c: A bound for $\mathbb{E}(\tilde{R}_j(\bar{\pi}))$ with $j \in \mathcal{J}$. For all $i \in I_j^0$ and all $x \in B_j$, the triangle inequality and Lemma A.2 with $V_j = \sqrt{d}P^{-1}$ show that $c_1P^{-\gamma}$ is smaller than

$$\begin{aligned}
&|\mathsf{T}(F^{\pi^\star(x_j)}(\cdot,x_j)) - \mathsf{T}(F^i(\cdot,x_j))| \\
&\quad \le |\mathsf{T}(F^{\pi^\star(x_j)}(\cdot,x_j)) - \mathsf{T}(F^{\pi^\star(x)}(\cdot,x))| + |\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x))| \\
&\qquad + |\mathsf{T}(F^i(\cdot,x)) - \mathsf{T}(F^i(\cdot,x_j))| \\
&\quad \le 2CLd^{\gamma/2}P^{-\gamma} + |\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x))|.
\end{aligned}$$

Recalling that $c_1 = 4CLd^{\gamma/2}+1$, we obtain

$$\mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big) - \mathsf{T}(F^i(\cdot,x)) > (1+2CLd^{\gamma/2})P^{-\gamma}. \tag{C.40}$$

(In particular, since $I_j^0 \neq \emptyset$ holds, $0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x))$, for all $x \in B_j$ if $j \in \mathcal{J}$, an observation we shall need later in Step 4d.) For every $i \in I_j^0$ and every $x \in B_j$, (C.40) and Lemma A.2 (recalling that $\mathbb{P}_X(B_j) > \underline{c}P^{-d} > 0$) imply

$$\Delta_j^i = \mathsf{T}(F_j^*) - \mathsf{T}(F_j^i) \geq \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x)) - 2CLd^{\gamma/2}P^{-\gamma} > P^{-\gamma}; \tag{C.41}$$

in particular, for any $j \in \mathcal{J}$, we have $\underline{\Delta}_j = \min_{i\in I_j^0} \Delta_j^i > P^{-\gamma}$. Recalling that $\tilde{R}_j(\bar{\pi}) = \tilde{R}_{j,I_j^*}(\bar{\pi}) + \tilde{R}_{j,I_j^0}(\bar{\pi})$, we combine (C.32) and (C.36) (with the just observed $\underline{\Delta}_j > P^{-\gamma}$) to see that, for any $j \in \mathcal{J}$,

$$\mathbb{E}[\tilde{R}_j(\bar{\pi})] \leq nc_1P^{-\gamma}q_j + \frac{2C^2(c_1+1)K\beta\log(\bar{c}nP^{-d})}{\underline{\Delta}_j} + (c_1+2C)K\frac{\beta+2}{\beta-2}. \tag{C.42}$$

Step 4d: A bound for $\sum_{j\in\mathcal{J}}\mathbb{E}[\tilde{R}_j(\bar{\pi})]$. Using equation (C.42) and $|\mathcal{J}| \leq P^d$, we obtain

$$\sum_{j\in\mathcal{J}}\mathbb{E}[\tilde{R}_j(\bar{\pi})] \leq (c_1+2C)K\frac{\beta+2}{\beta-2}P^d + nc_1P^{-\gamma}\sum_{j\in\mathcal{J}}q_j + \sum_{j\in\mathcal{J}}\frac{2C^2(c_1+1)K\beta\log(\bar{c}nP^{-d})}{\underline{\Delta}_j}. \tag{C.43}$$

Since the B_j are disjoint, we obtain, recalling the definition of q_j after equation (C.32), that

$$\frac{nc_1}{P^\gamma}\sum_{j\in\mathcal{J}}q_j \leq \frac{nc_1}{P^\gamma}\mathbb{P}\big(0 < \mathsf{T}(F^{\pi^\star(X_1)}(\cdot,X_1)) - \mathsf{T}(F^{\pi^\sharp(X_1)}(\cdot,X_1)) < c_1P^{-\gamma}\big) \tag{C.44}$$

$$\leq C_0c_1^{1+\alpha}nP^{-\gamma(1+\alpha)}, \tag{C.45}$$

where we used Assumption 3.6 to obtain the last inequality.

To deal with the last sum in the upper bound in (C.43), we need a better lower bound on the $\underline{\Delta}_j$-s than the already available $P^{-\gamma}$. For notational simplicity, suppose that the well-behaved bins are indexed as $\mathcal{J} = \{1,2,\ldots,j_1\}$ such that $0 < P^{-\gamma} \leq \underline{\Delta}_1 \leq \underline{\Delta}_2 \leq \cdots \leq \underline{\Delta}_{j_1}$. Fix $j \in \mathcal{J}$. Then, for any $k = 1,\ldots,j$, we claim that

$$B_k \subseteq \left\{x : 0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x)) < \underline{\Delta}_j + 2CLd^{\gamma/2}P^{-\gamma}\right\}. \tag{C.46}$$

To see (C.46), note that, by definition, there exists an $i \in \mathcal{I}_k^0$ such that $\underline{\Delta}_k = \mathsf{T}(F_k^*) - \mathsf{T}(F_k^i)$. Given $x \in B_k$, Lemmas A.2 and A.3 and Remark 2.3 in Kock et al. (2022) yield (the first inequality following from the observation after equation (C.40))

$$\begin{aligned}0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x)) &\leq \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x))\\ &\leq \underline{\Delta}_k + 2CLd^{\gamma/2}P^{-\gamma}\\ &\leq \underline{\Delta}_j + 2CLd^{\gamma/2}P^{-\gamma},\end{aligned}$$

and thus x is an element of the set on the right-hand-side of (C.46). Since all bins B_k are disjoint and $\underline{\Delta}_j + 2CLd^{\gamma/2}P^{-\gamma} \le c_1\underline{\Delta}_j$ (obtained by recalling $c_1 = 4CLd^{\gamma/2}+1$, and using $\underline{\Delta}_j > P^{-\gamma}$), the inclusion (C.46) yields that, for any $j \in \mathcal{J}$,

$$\mathbb{P}_X\big(x : 0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x)) < c_1\underline{\Delta}_j\big) \ge \sum_{k=1}^{j} \mathbb{P}_X(B_k) \ge \frac{\underline{c}j}{P^d}. \tag{C.47}$$

Denote $j_2 := \max\{j \in \mathcal{J} : \underline{\Delta}_j \le 1/c_1\}$ (here interpreting the maximum of an empty set as 0). Then, for each $j \in \{1,\dots,j_2\}$ by Assumption 3.6,

$$\mathbb{P}_X\big(0 < \mathsf{T}(F^{\pi^\star(X)}(\cdot,X)) - \mathsf{T}(F^{\pi^\sharp(X)}(\cdot,X)) < c_1\underline{\Delta}_j\big) \le C_0(c_1\underline{\Delta}_j)^\alpha. \tag{C.48}$$

Combining (C.47), (C.48), and $\underline{\Delta}_j > P^{-\gamma}$, for any $j \in \{1,\dots,j_2\}$, we obtain the inequality $\underline{\Delta}_j \ge \max\left(c_*\big(jP^{-d}\big)^{1/\alpha}, P^{-\gamma}\right)$, with constant $c_* := c_1^{-1}\underline{c}^{1/\alpha}C_0^{-1/\alpha}$. Combining this with the identity $\underline{\Delta}_j > 1/c_1$, for $j > j_2$, we obtain that

$$\begin{aligned}\sum_{j\in\mathcal{J}} \frac{1}{\underline{\Delta}_j} &\le \sum_{j=1}^{j_2} \min\left(c_*^{-1}\big(P^d/j\big)^{1/\alpha}, P^\gamma\right) + \sum_{j=j_2+1}^{j_1} c_1\\ &\le \sum_{j=1}^{P^d} \min\left(c_*^{-1}\big(P^d/j\big)^{1/\alpha}, P^\gamma\right) + c_1P^d.\end{aligned}$$

For $\tilde{P} := \lceil P^{d-\alpha\gamma}\rceil$ (in fact, for *any* $\tilde{P} \in \{1,\dots,P^d\}$, and thus in particular for our particular choice), it holds that

$$\sum_{j=1}^{P^d} \min\left(c_*^{-1}\big(P^d/j\big)^{1/\alpha}, P^\gamma\right) \le \sum_{j=1}^{\tilde{P}} P^\gamma + c_*^{-1}P^{d/\alpha} \sum_{j=\tilde{P}+1}^{\infty} j^{-1/\alpha} \le c_{**}P^{d+\gamma(1-\alpha)},$$

for $c_{**} := [2 + c_*^{-1}(\alpha^{-1}-1)^{-1}]$, where we used $\sum_{j=\tilde{P}+1}^{\infty} j^{-1/\alpha} \le (\alpha^{-1}-1)^{-1}\tilde{P}^{1-\alpha^{-1}}$. Hence, equations (C.43) and (C.44) and the bounds in the previous two displays imply

$$\sum_{j\in\mathcal{J}} \mathbb{E}[\tilde{R}_j(\bar{\pi})] \le c'''\left(nP^{-\gamma(1+\alpha)} + K\overline{\log}(nP^{-d})P^d + K\overline{\log}(nP^{-d})P^{d+\gamma(1-\alpha)}\right), \tag{C.49}$$

for a constant c''', say, that depends on $d, L, \gamma, \underline{c}, \overline{c}, C, C_0, \alpha$ and β, but *not* on n.

Step 5: Combining. From equations (C.19), (C.24), (C.28), and (C.49), we obtain that $\mathbb{E}[R_n(\bar{\pi})]$ is bounded from above by

$$\frac{c''''}{4}\left(nP^{-\gamma(1+\alpha)} + \sqrt{Kn\overline{\log}(n)}P^{d/2-\gamma\alpha} + K\overline{\log}(nP^{-d})P^d + K\overline{\log}(nP^{-d})P^{d+\gamma(1-\alpha)}\right), \tag{C.50}$$

for a constant c'''' that depends on $d, L, \gamma, \underline{c}, \bar{c}, C, C_0, \alpha$ and β, but *not* on n. From $P = \lceil n^{1/(2\gamma+d)} \rceil$, we get $n \leq P^{2\gamma+d}$, and obtain

$$\mathbb{E}[R_n(\bar{\pi})] \leq \frac{c''''}{4} K\overline{\log}(n) \left(nP^{-\gamma(1+\alpha)} + n^{1/2} P^{d/2-\gamma\alpha} + 2P^{d+\gamma(1-\alpha)} \right) \tag{C.51}$$

$$\leq c'''' K\overline{\log}(n) P^{d+\gamma(1-\alpha)}, \tag{C.52}$$

from which the conclusion follows.

C.5. Proof of Theorem 3.8

To prove the theorem, we just combine Theorem 3.7 and the following lemma, which allows one to upper-bound the number of suboptimal assignments made by any policy.

LEMMA C.1. *Suppose Assumptions 2.3 and 3.6 hold. Let $D_0 \geq \max(2, C_0^{-1})$, and define $\tilde{C}(\alpha, D_0, C_0) = (1 - 1/D_0)/(C_0 D_0)^{1/\alpha}$. Then, for any policy π, for any randomization measure, and for all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, such that $\mathbb{P}_{Y,X}$ satisfies equation (2) and Assumption 2.8, it holds that*

$$\mathbb{E}[R_n(\pi)] \geq \tilde{C}(\alpha, D_0, C_0) n^{-1/\alpha} \left(\mathbb{E}[S_n(\pi)] \right)^{1+1/\alpha} \quad \text{for every } n \in \mathbb{N}. \tag{C.53}$$

Remark C.2. In Lemma C.1, we impose Assumptions 2.3 and 2.8 and equation (2) to guarantee that $R_n(\pi)$ and $S_n(\pi)$ are random variables, and that $\pi^\star$ and $\pi^\sharp$ are measurable (cf. also the discussion in the footnote of Assumption 3.6).

Proof. The proof idea is quite standard, and we follow Rigollet and Zeevi (2010): Choose $D_0 \geq \max(2, C_0^{-1})$, implying that $1/(C_0 D_0)^{1/\alpha} \leq 1$. Let $n \in \mathbb{N}$, and let π be a policy as defined in Section 2. We write $\pi_{n,t} = \pi_t$. Let $\mathbb{P}_G$ be a randomization measure. We show that

$$\mathbb{E}[R_n(\pi)] \geq \tilde{C} n^{-1/\alpha} \left(\mathbb{E}[S_n(\pi)] \right)^{1+1/\alpha}, \tag{C.54}$$

for $\tilde{C} = \tilde{C}(\alpha, D_0, C_0)$. If $\mathbb{E}[S_n(\pi)] = 0$, (C.54) trivially holds. Thus, suppose that $\mathbb{E}[S_n(\pi)] > 0$. Note that, for any $\delta > 0$,

$$\begin{aligned}
R_n(\pi) &\geq \delta \sum_{t=1}^{n} \mathbb{1}_{\{\mathsf{T}(F^{\pi^\star(X_t)}(\cdot, X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot, X_t)) > \delta\}} \mathbb{1}\left\{\pi_t(X_t, Z_{t-1}, G_t) \notin \arg\max_{i \in \mathcal{I}} \{\mathsf{T}(F^i(\cdot, X_t))\}\right\} \\
&= \delta S_n(\pi) - \delta \sum_{t=1}^{n} \mathbb{1}_{\{\mathsf{T}(F^{\pi^\star(X_t)}(\cdot, X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot, X_t)) \leq \delta\}} \\
&\quad \times \mathbb{1}\left\{\pi_t(X_t, Z_{t-1}, G_t) \notin \arg\max_{i \in \mathcal{I}} \{\mathsf{T}(F^i(\cdot, X_t))\}\right\} \\
&= \delta S_n(\pi) - \delta \sum_{t=1}^{n} \mathbb{1}_{\{0 < \mathsf{T}(F^{\pi^\star(X_t)}(\cdot, X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot, X_t)) \leq \delta\}} \\
&\quad \times \mathbb{1}\left\{\pi_t(X_t, Z_{t-1}, G_t) \notin \arg\max_{i \in \mathcal{I}} \{\mathsf{T}(F^i(\cdot, X_t))\}\right\} \\
&\geq \delta S_n(\pi) - \delta \sum_{t=1}^{n} \mathbb{1}_{\{0 < \mathsf{T}(F^{\pi^\star(X_t)}(\cdot, X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot, X_t)) \leq \delta\}},
\end{aligned}$$

where the second equality used that if $\pi_t(X_t, Z_{t-1}, G_t) \notin \arg\max_{i \in \mathcal{I}} \{\mathsf{T}(F^i(\cdot, X_t))\}$, then $0 < \mathsf{T}(F^{\pi^\star(X_t)}(\cdot, X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot, X_t))$. Choosing $\delta := (\mathbb{E}[S_n(\pi)]/(nC_0D_0))^{1/\alpha} \leq 1/(C_0D_0)^{1/\alpha} \leq 1$ (the first inequality following from $\mathbb{E}[S_n(\pi)] \leq n$), Assumption 3.6 yields

$$\begin{aligned}\mathbb{E}[R_n(\pi)] \geq \delta(\mathbb{E}[S_n(\pi)] - C_0 n\delta^\alpha) &= \delta(1 - 1/D_0)\mathbb{E}[S_n(\pi)] \\ &= \tilde{C} n^{-1/\alpha} \left(\mathbb{E}[S_n(\pi)]\right)^{1+1/\alpha}, \end{aligned} \tag{C.55}$$

which proves (C.54). □

C.6. Proof of Theorem 3.9

Let π be a policy, let $\mathbb{P}_X$ be the uniform distribution on $[0,1]^d$, let $\mathbb{P}_G$ be a randomization measure, and fix an $n \in \mathbb{N}$. To simplify notation, we abbreviate $\pi_{n,t} = \pi_t$. The proof of the inequalities in (13) and (14) now proceeds in five steps.

Step 0: Preliminary observations and some notation. (a) From the maintained assumptions and Assumption 2.3 (imposed through Assumption 2.6), it follows that

$$\begin{aligned} c_-(\tau_2 - \tau_1) \leq \mathsf{T}(J_{\tau_2}) - \mathsf{T}(J_{\tau_1}) &\leq C\|J_{\tau_2} - J_{\tau_1}\|_\infty \\ &\leq C(\tau_2 - \tau_1) \quad \text{for every } \tau_1 \leq \tau_2 \text{ in } [0,1]. \end{aligned} \tag{C.56}$$

Let $\varepsilon := 2/\sqrt{17} < 1/2$, set $H_v := J_{1/2+v}$ for every $v \in [-\varepsilon, \varepsilon]$, and define the map $h : [-\varepsilon, \varepsilon] \to [h(-\varepsilon), h(\varepsilon)]$ via $v \mapsto \mathsf{T}(H_v)$; note that h is strictly increasing because of $c_- > 0$ and the observation in the previous display. (b) The previous display also implies that h is Lipschitz continuous with constant C and that $h(w) - h(v) \geq c_-(w - v)$ for every $v \leq w$ in $[-\varepsilon, \varepsilon]$; implying that h possesses a Lipschitz-continuous inverse function h^{-1} : $[h(-\varepsilon), h(\varepsilon)] \to [-\varepsilon, \varepsilon]$, say, with constant c_-^{-1}. (c) Note that the map $v \mapsto H_v$ (as a map from $[-\varepsilon, \varepsilon]$ to $D_{cdf}([a,b])$ equipped with the supremum metric) is Lipschitz-continuous with constant 1. (d) Finally, we verify that, for $\zeta := c_-^{-1}(0.5^2 - \varepsilon^2)^{-1/2}$, we have (recalling the notational conventions introduced in the first paragraph of Appendix A)

$$\mathsf{KL}^{1/2}(\mu_{H_v}, \mu_{H_w}) \leq \zeta \left(\mathsf{T}(H_w) - \mathsf{T}(H_v)\right) \quad \text{for every } v \leq w \text{ in } [-\varepsilon, \varepsilon]. \tag{C.57}$$

By definition, $\mathsf{T}(H_w) - \mathsf{T}(H_v) = h(w) - h(v)$. Hence, the statement in (C.57) follows from observation (b) once we verify $\mathsf{KL}^{1/2}(\mu_{H_v}, \mu_{H_w}) \leq (w - v)/\sqrt{0.5^2 - \varepsilon^2}$. But the latter is a simple consequence of Lemma B.3 in Kock et al. (2022) (and is established similarly as the last claim in Lemma B.4 in Kock et al., 2022).

Step 1: Construction of a family of functions $\mathcal{C}$. For $P \in \mathbb{N}$ (to be chosen in Step 4), let $B_1^P, \ldots, B_{P^d}^P$ be the hypercubes defined in (9), and sorted lexicographically; we shall drop the superscript P in the following. Let q_i, $i = 1, \ldots, P^d$, denote the center of B_i. Let $m := \lceil P^{d-\gamma\alpha} \rceil$, and observe that $1 \leq m \leq P^d$. Next, let $\Sigma_m := \{-1, 1\}^m$, $|\Sigma_m| = 2^m$, and define $\mathcal{C}_m = \mathcal{C} := \{f_\sigma : \sigma \in \Sigma_m\}$, where for $\sigma \in \Sigma_m$ we construct $f_\sigma : [0,1]^d \to \mathbb{R}$ via

$$f_\sigma(x) := h(0) + c_-\varepsilon \sum_{j=1}^m \sigma_j \varphi_j(x);$$

for every $j \in \{1,\dots,P^d\}$, we denote $\varphi_j(x) := 4^{-1}P^{-\gamma}\phi(2P(x-q_j))\mathbb{1}_{B_j}(x)$, where $\phi(x) := (1-||x||_\infty)^\gamma$, and $\|x\|_\infty := \max_{1\le i\le d}|x_i|$ for $x \in \mathbb{R}^d$. Note that every f_σ is continuous.

We now show that every f_σ is Hölder-continuous. More precisely, we show that, for every $f_\sigma \in \mathcal{C}$,

$$|f_\sigma(x_1) - f_\sigma(x_2)| \le c_-\varepsilon 2^{-1}||x_1 - x_2||^\gamma \quad \text{for every } x_1, x_2 \in [0,1]^d, \tag{C.58}$$

with $\|\cdot\|$ denoting the Euclidean norm. We note that, for any pair $x_1, x_2 \in [0,1]^d$, one has $|\phi(x_1) - \phi(x_2)| \le ||x_1 - x_2||_\infty^\gamma \le ||x_1 - x_2||^\gamma$; the second inequality is obvious, and the first inequality follows from $|p^\gamma - q^\gamma| \le |p-q|^\gamma$ for $p,q \ge 0$ and $0 < \gamma \le 1$, together with the reverse triangle inequality. Now, to show (C.58), we consider two cases: First, if $x_1, x_2 \in B_j$ for $j \in \{1,\dots,P^d\}$, the definition of f_σ and $|\phi(x_1) - \phi(x_2)| \le ||x_1 - x_2||^\gamma$ lead to (note that if $j > m$, the following inequality trivially holds)

$$[c_-\varepsilon]^{-1}|f_\sigma(x_1) - f_\sigma(x_2)| \le |\varphi_j(x_1) - \varphi_j(x_2)| \le \frac{2^\gamma}{4}||x_1 - x_2||^\gamma \le \frac{1}{2}||x_1 - x_2||^\gamma. \tag{C.59}$$

We remark that by continuity of f_σ, equation (C.59) continues to hold if x_1 and x_2 are elements of the closure of B_j, i.e., of $\bar{B}_j$. Second, suppose that $x_1 \in B_j, x_2 \in B_k$, for $j \ne k$. Let $S := \{\theta x_1 + (1-\theta)x_2 : \theta \in [0,1]\}$. Define $y_1 := \operatorname{argmin}_{z \in S \cap \bar{B}_j}||z - x_2||$ and $y_2 := \operatorname{argmin}_{z \in S \cap \bar{B}_k}||z - x_1||$. Clearly, y_1 and y_2 are the elements of the boundary of B_j and B_k, respectively, implying $\varphi_j(y_1) = \varphi_k(y_2) = 0$. Denote $\bar{\sigma}_i = \sigma_i$, for $i = 1,\dots,m$, and $\bar{\sigma}_i = 0$, for $i > m$. We obtain

$$\begin{aligned}
[c_-\varepsilon]^{-1}|f_\sigma(x_1) - f_\sigma(x_2)| = |\bar{\sigma}_j\varphi_j(x_1) - \bar{\sigma}_k\varphi_k(x_2)| &\le |\varphi_j(x_1) - \varphi_j(y_1)| + |\varphi_k(y_2) - \varphi_k(x_2)| \\
&\le \frac{2^\gamma}{4}(||x_1 - y_1||^\gamma + ||y_2 - x_2||^\gamma) \\
&\le 2^{-1}||x_1 - x_2||^\gamma,
\end{aligned}$$

where for the second inequality we made use of the second inequality in (C.59) (cf. also the remark immediately after (C.59)), and for the third inequality, we combined $(a^\gamma + b^\gamma) \le 2^{1-\gamma}(a+b)^\gamma$ for $0 < \gamma \le 1$ and $a,b \ge 0$ with $||x_1 - y_1|| + ||y_2 - x_2|| \le ||x_1 - y_1|| + ||y_1 - y_2|| + ||y_2 - x_2|| = ||x_1 - x_2||$. Since the hypercubes $B_1,\dots,B_{P^d}$ define a partition of $[0,1]^d$, this establishes equation (C.58).

Step 2: Construction of probability measures $\mathbb{P}_f$ indexed by $\mathcal{C}$. Recall from Observation (b) in Step 0 that $h : [-\varepsilon,\varepsilon] \to [-h(\varepsilon), h(\varepsilon)]$ defined via $v \mapsto \mathsf{T}(H_v)$ permits a Lipschitz-continuous inverse $h^{-1} : [h(-\varepsilon), h(\varepsilon)] \to [-\varepsilon,\varepsilon]$, say, with corresponding Lipschitz constant c_-^{-1}. By construction, the range of $f \in \mathcal{C}$ is contained in $[h(-\varepsilon), h(\varepsilon)]$, because $h(\varepsilon) - h(0) \ge c_-\varepsilon$ and similarly $h(0) - h(-\varepsilon) \ge c_-\varepsilon$. Hence, for every $f \in \mathcal{C}$, the composition $A_f := h^{-1} \circ f : [0,1]^d \to [-\varepsilon,\varepsilon]$ is well defined, and equation (C.58) shows that A_f is Hölder-continuous with constant $\varepsilon/2$ and exponent γ. Note that by definition

$$f(x) = h\left(h^{-1} \circ f(x)\right) = h(A_f(x)) = \mathsf{T}\left(H_{A_f(x)}\right) \quad \text{for every } x \in [0,1]^d \text{ and every } f \in \mathcal{C}. \tag{C.60}$$

We next show that $\mu_{H_{A_f(\cdot)}}(\cdot) : \mathcal{B}(\mathbb{R}) \times [0,1]^d \to [0,1]$, defined via $B \times x \mapsto \mu_{H_{A_f(x)}}(B)$, is a stochastic kernel: (i) By definition, $\mu_{H_{A_f(x)}}$ is a probability measure for every $x \in [0,1]^d$. (ii) Recall from Observation (c) in Step 0 that $\|H_v - H_w\|_\infty \le |v - w|$ for every pair

$v,w \in [-\varepsilon,\varepsilon]$. From continuity of A_f, it follows that $x \mapsto H_{A_f(x)}(c) = \mu_{H_{A_f(x)}}((-\infty,c])$ is continuous (and hence measurable) for every $c \in \mathbb{R}$. Since $\{(-\infty,c] : c \in \mathbb{R}\}$ is a "π-system" that generates the Borel σ-algebra on $\mathbb{R}$, Lemma 1.40 of Kallenberg (2001) shows that $\mu_{H_{A_f(\cdot)}}(\cdot) : \mathcal{B}(\mathbb{R}) \times [0,1]^d \to [0,1]$ is a stochastic kernel.

For every $f \in \mathcal{C}$, we define the probability measure

$$\mathbb{P}_f := \mu_{H_0} \otimes [\mu_{H_{A_f(\cdot)}} \otimes \mathbb{P}_X], \tag{C.61}$$

noting that the product in brackets is a semidirect product. For later reference, we note that if $(Y_t,X_t) \sim \mathbb{P}_f$, it holds for every $x \in [0,1]^d$ that $F^1(\cdot,x) = H_0$ and $F^2(\cdot,x) = H_{A_f(x)}$. In particular, equation (2) is satisfied as a consequence of Assumption 2.6. Now, for every $t = 1,\ldots,n$, denote by $\mathbb{P}^t_{\pi,f}$ the probability measure on the Borel sets of $\mathbb{R}^{(d+2)t}$ induced by the (recursively defined) random vector $Z_t = (Y_{\pi_t(X_t,Z_{t-1},G_t),t}, X_t, G_t, \ldots, Y_{\pi_1(X_1,G_1),1}, X_1, G_1)$ with i.i.d. $(Y_t,X_t,G_t) \sim \mathbb{P}_f \otimes \mathbb{P}_G$. In the sequel, for $t = 1,\ldots,n$, the symbol z_t will denote a "generic" element of $\mathbb{R}^{(d+2)t}$ (i.e., a "realization" of the random vector Z_t).

We close this step with an important observation: Note that $\bar{K}_{t,f} : \mathcal{B}(\mathbb{R}) \times [0,1]^d \times \mathbb{R} \times \mathbb{R}^{(t-1)(d+1)}$ defined via

$$B \times x \times g \times z_{t-1} \mapsto \mu_{H_0}(B)\mathbb{1}\{\pi_t(x,z_{t-1},g) = 1\} + \mu_{H_{A_f(x)}}(B)\mathbb{1}\{\pi_t(x,z_{t-1},g) = 2\} \tag{C.62}$$

is a regular conditional distribution of $Y_{\pi_t(X_t,Z_{t-1},G_t),t}$ given (X_t,G_t,Z_{t-1}), and that, for every $t = 1,\ldots,n$, we can therefore write (noting that $Z_t = (Y_{\pi_t(X_t,Z_{t-1},G_t),t}, X_t, G_t, Z_{t-1})$, interpreting Z_0 as the empty vector)

$$\mathbb{P}^t_{\pi,f} = \bar{K}_{t,f} \otimes [\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f}], \tag{C.63}$$

with the convention that in case $t = 1$, one has to drop the factor $\mathbb{P}^{t-1}_{\pi,f}$ in the previous display and the "z_{t-1}" in equation (C.62). Hence, interpreting $\mathsf{KL}(\mathbb{P}^{t-1}_{\pi,f_1}, \mathbb{P}^{t-1}_{\pi,f_2}) = 0$ in case $t = 1$, and with the just mentioned "dropping" convention, the Chain Rule of Lemma A.1 implies that, for $f_1, f_2 \in \mathcal{C}$ and any $t = 1,\ldots,n$, we have

$$\begin{aligned}\mathsf{KL}(\mathbb{P}^t_{\pi,f_1}, \mathbb{P}^t_{\pi,f_2}) &= \mathsf{KL}\left(\bar{K}_{t,f_1} \otimes [\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f_1}], \bar{K}_{t,f_2} \otimes [\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f_2}]\right) \\ &= \mathsf{KL}(\mathbb{P}^{t-1}_{\pi,f_1}, \mathbb{P}^{t-1}_{\pi,f_2}) + \mathsf{KL}\left(\bar{K}_{t,f_1} \otimes [\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f_1}], \bar{K}_{t,f_2} \otimes [\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f_1}]\right),\end{aligned}$$

the right-hand-side being equal to the sum of $\mathsf{KL}(\mathbb{P}^{t-1}_{\pi,f_1}, \mathbb{P}^{t-1}_{\pi,f_2})$ and

$$\int_{[0,1]^d \times \mathbb{R} \times \mathbb{R}^{(t-1)(d+2)}} \mathsf{KL}(\bar{K}_{t,f_1}(\cdot,x,g,z_{t-1}), \bar{K}_{t,f_2}(\cdot,x,g,z_{t-1}))\, d(\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f_1})(x,g,z_{t-1}).$$

Using equation (C.62), this sum further simplifies to

$$\mathsf{KL}(\mathbb{P}^{t-1}_{\pi,f_1}, \mathbb{P}^{t-1}_{\pi,f_2}) + \int_{\{\pi_t = 2\}} \mathsf{KL}(\mu_{H_{A_{f_1}(x)}}, \mu_{H_{A_{f_2}(x)}})\, d(\mathbb{P}_X \otimes \mathbb{P}_G \otimes \mathbb{P}^{t-1}_{\pi,f_1})(x,g,z_{t-1}),$$

which, noting that $\mathbb{P}^{t-1}_{\pi,f_1}$ is obtained by a coordinate projection from $\mathbb{P}^{n}_{\pi,f_1}$, implies

$$\begin{aligned}&\mathsf{KL}\left(\mathbb{P}^t_{\pi,f_1},\mathbb{P}^t_{\pi,f_2}\right)\\ \leq&\mathsf{KL}(\mathbb{P}^{t-1}_{\pi,f_1},\mathbb{P}^{t-1}_{\pi,f_2})+\int_{\{\pi_t=2\}}\mathsf{KL}(\mu_{H_{A_{f_1}(x)}},\mu_{H_{A_{f_2}(x)}})d(\mathbb{P}_X\otimes\mathbb{P}_G\otimes\mathbb{P}^n_{\pi,f_1})(x,g,z_n).\end{aligned}$$

By induction, it now immediately follows that, for every $t=1,\dots,n$,

$$\mathsf{KL}\left(\mathbb{P}^t_{\pi,f_1},\mathbb{P}^t_{\pi,f_2}\right)\leq\int\sum_{i=1}^{t}\mathbb{1}\{\pi_i=2\}\mathsf{KL}(\mu_{H_{A_{f_1}(x)}},\mu_{H_{A_{f_2}(x)}})d(\mathbb{P}_X\otimes\mathbb{P}_G\otimes\mathbb{P}^n_{\pi,f_1})(x,g,z_n). \tag{C.64}$$

Step 3: Verifying Assumptions 2.8 and 3.6 for every $\mathbb{P}_f$. Fix $f=f_\sigma\in\mathcal{C}$. To verify Assumption 2.8 (with γ and $L=\varepsilon/2$ as given in the theorem; cf. Step 0 for the definition of ε) for $\mathbb{P}_f$, which was defined in (C.61), note that

$$\|F^2(\cdot,x_1)-F^2(\cdot,x_2)\|_\infty=\|H_{A_f(x_1)}-H_{A_f(x_2)}\|_\infty\leq|A_f(x_1)-A_f(x_2)|\leq L\|x_1-x_2\|^\gamma,$$

the first inequality following Observation (c) in Step 0, and the second following from A_f being Hölder-continuous with constant $L=\varepsilon/2$ and exponent γ, as observed in Step 2 right before equation (C.60); note further that $F^1(\cdot,x)=H_0$, and that the previous display, hence, trivially holds for F^2 replaced by F^1. Next, to verify Assumption 3.6 (with α and $C_0=8d[c_-\varepsilon]^{-\alpha}$ as given in the theorem), it suffices to show (recall that $K=2$) that

$$\mathbb{P}_X\left(x\in[0,1]^d:0<|\mathsf{T}(H_{A_f(x)})-\mathsf{T}(H_0)|\leq c_-\varepsilon\delta\right)\leq 8d\delta^\alpha\text{ for all }\delta\geq 0. \tag{C.65}$$

The statement in (C.65) is trivial for $\delta=0$. Let $\delta>0$. We use equation (C.60) to write

$$[c_-\varepsilon]^{-1}|\mathsf{T}(H_{A_f(x)})-\mathsf{T}(H_0)|=\sum_{j=1}^{m}\varphi_j(x),$$

where we used that $B_j\cap B_k=\emptyset$ for $j\neq k$. Noting that $\sum_{j=1}^m\varphi_j(x)=0$ for $x\notin\bigcup_{j=1}^m B_j$, we obtain

$$\mathbb{P}_X\left(x\in[0,1]^d:0<|\mathsf{T}(H_{A_f(x)})-\mathsf{T}(H_0)|\leq c_-\varepsilon\delta\right)=\sum_{j=1}^{m}\mathbb{P}_X\left(x\in B_j:0<\varphi_j(x)\leq\delta\right),$$

which we can write as

$$\begin{aligned}m\mathbb{P}_X\left(x\in B_1:\phi(2P(x-q_1))\leq 4P^\gamma\delta\right)&=m(2P)^{-d}\int_{[-1,1]^d}\mathbb{1}_{\{\phi\leq 4P^\gamma\delta\}}dx\\&=mP^{-d}\int_{[0,1]^d}\mathbb{1}_{\{\phi\leq 4P^\gamma\delta\}}dx,\end{aligned}$$

where the first equality follows upon substituting $u=2P(x-q_1)$, and the second equality follows from $\phi(x)$ being invariant to multiplying coordinates of x by -1. To upper-bound the expression to the right in the previous display, we consider two cases: If $4P^\gamma\delta>1$, then

$$mP^{-d}\int_{[0,1]^d}\mathbb{1}_{\{\phi\leq 4P^\gamma\delta\}}dx=mP^{-d}\leq 2P^{-\gamma\alpha}\leq 8\delta^\alpha,$$

where we used $m = \lceil P^{d-\gamma\alpha} \rceil \leq P^{d-\gamma\alpha} + 1 \leq 2P^{d-\gamma\alpha}$ and $\alpha \in (0,1)$. On the other hand, if $4P^\gamma \delta \leq 1$, we write $\mathbb{1}_{\{\phi \leq 4P^\gamma \delta\}} = 1 - \mathbb{1}_{\{4P^\gamma\delta < \phi\}} = 1 - \mathbb{1}_{\{\|\cdot\|_\infty < 1-(4\delta)^{1/\gamma}P\}}$ to obtain

$$mP^{-d} \int_{[0,1]^d} \mathbb{1}_{\{\phi \leq 4P^\gamma \delta\}} dx = mP^{-d}(1 - \int_{[0,1]^d} \mathbb{1}_{\{\|\cdot\|_\infty < 1-(4\delta)^{1/\gamma}P\}} dx)$$
$$= mP^{-d}[1 - (1-(4\delta)^{1/\gamma}P)^d],$$

which, using $(1-(1-s)^d) \leq ds$ for $s \in [0,1]$, $m \leq 2P^{d-\gamma\alpha}$, $P \leq (4\delta)^{-1/\gamma}$, and $\alpha \in (0,1)$, is bounded from above by

$$mP^{1-d} d(4\delta)^{1/\gamma} \leq 2dP^{1-\alpha\gamma}(4\delta)^{1/\gamma} \leq 2d(4\delta)^\alpha \leq 8d\delta^\alpha.$$

Step 4: Lower bounding the suprema in equations (13) and (14). We start with equation (14). We already know that, for every $f \in \mathcal{C}$, the measure $\mathbb{P}_f$ satisfies the inclusion in equation (2) and Assumptions 2.8 and 3.6. It therefore suffices to verify

$$\sup_{f \in \mathcal{C}} \mathbb{E}_{(\mathbb{P}_f \otimes \mathbb{P}_G)^n} \left[S_n(\pi) \right] \geq n^{1-\frac{\alpha\gamma}{d+2\gamma}}/32, \tag{C.66}$$

where $\mathbb{E}_{(\mathbb{P}_f \otimes \mathbb{P}_G)^n}$ denotes the expectation w.r.t. the product measure $\bigotimes_{t=1}^n (\mathbb{P}_f \otimes \mathbb{P}_G)$ (here, we interpret, with some abuse of notation, $S_n(\pi)$ as a function on the range space of (X_t, Y_t, G_t) for $t = 1, \ldots, n$, and we shall denote a generic realization of (X_t, Y_t, G_t) by (x_t, y_t, g_t) to make this convention explicit, where we sometimes drop the subindex t, if no confusion can arise).

We first observe that for $\mathbb{P}_{f_\sigma}$, denoting $\bar{f}_\sigma := [c_-\varepsilon]^{-1}[f_\sigma - h(0)] = \sum_{j=1}^m \sigma_j \varphi_j$, we have

$$S_n(\pi) = \sum_{t=1}^n \mathbb{1}\{\mathsf{T}(F^1(\cdot, x_t)) \neq \mathsf{T}(F^2(\cdot, x_t)),\ \pi^\star(x_t) \neq \pi_t(x_t, z_{t-1}, g_t)\}$$
$$= \sum_{t=1}^n \mathbb{1}\{\bar{f}_\sigma(x_t) \neq 0,\ 2\pi_t(x_t, z_{t-1}, g_t) - 3 \neq \operatorname{sign}(\bar{f}_\sigma(x_t))\},$$

where for the second equality we used that $\pi^\star(x) = 3/2 + \operatorname{sign}(\bar{f}_\sigma(x))/2$ (with the convention that the sign of 0 is -1), and where we recalled from equation (C.60) that $\mathsf{T}(F^1(\cdot, x)) \neq \mathsf{T}(F^2(\cdot, x))$ is equivalent to $\bar{f}_\sigma(x) \neq 0$. Noting that the random vectors X_t, Z_{t-1}, and G_t are independent, it follows that their joint distribution equals $\mathbb{P}_X \otimes \mathbb{P}_{\pi, f_\sigma}^{t-1} \otimes \mathbb{P}_G$. Using Tonelli's theorem, writing $\mathbb{E}_G$ for the expectation w.r.t. $\mathbb{P}_G$, abbreviating $2\pi_t(x, z_{t-1}, g) - 3 := \check{\pi}_t(x, z_{t-1}, g)$, and noting that the tth summand in the previous display depends on z_t only via z_{t-1}, we can write $\sup_{f \in \mathcal{C}} \mathbb{E}_{(\mathbb{P}_f \otimes \mathbb{P}_G)^n}[S_n(\pi)]$ as

$$\sup_{\sigma \in \Sigma_m} \sum_{t=1}^n \mathbb{E}_{\pi, f_\sigma}^{t-1} \mathbb{E}_G \left[\mathbb{P}_X \left(x : \bar{f}_\sigma(x) \neq 0,\ \check{\pi}_t(x, z_{t-1}, g_t) \neq \operatorname{sign}(\bar{f}_\sigma(x)) \right) \right]$$
$$\geq \sup_{\sigma \in \Sigma_m} \sum_{j=1}^m \sum_{t=1}^n \mathbb{E}_{\pi, f_\sigma}^{t-1} \mathbb{E}_G[\mathbb{P}_X(x \in B_j : \check{\pi}_t(x, z_{t-1}, g_t) \neq \sigma_j)]$$
$$\geq \frac{1}{2^m} \sum_{j=1}^m \sum_{t=1}^n \sum_{\sigma \in \Sigma_m} \mathbb{E}_{\pi, f_\sigma}^{t-1} \mathbb{E}_G[\mathbb{P}_X(x \in B_j : \check{\pi}_t(x, z_{t-1}, g_t) \neq \sigma_j)], \tag{C.67}$$

where we used that $m \leq P^d$ and $\mathbb{P}_X(x \in B_j : \bar{f}_\sigma(x) = 0) = 0$ (and where we use a corresponding “dropping” convention for the index $t = 1$ as introduced after equation (C.63)). For every $j \in \{1, \ldots, m\}$ and $t \in \{1, \ldots, n\}$,

$$
\begin{aligned}
Q_t^j &:= \sum_{\sigma \in \Sigma_m} \mathbb{E}^{t-1}_{\pi, f_\sigma} \mathbb{E}_G[\mathbb{P}_X(x \in B_j : \check{\pi}_t(x, z_{t-1}, g) \neq \sigma_j)] \\
&= \sum_{\sigma_{-j} \in \Sigma_{m-1}} \sum_{i \in \{-1,1\}} \mathbb{E}^{t-1}_{\pi, f_{\sigma^i_{-j}}} \mathbb{E}_G[\mathbb{P}_X(x \in B_j : \check{\pi}_t(x, z_{t-1}, g) \neq i)],
\end{aligned}
$$

where $\sigma_{-j} := (\sigma_1, \ldots, \sigma_{j-1}, \sigma_{j+1}, \ldots, \sigma_m)$ and $\sigma^i_{-j} := (\sigma_1, \ldots, \sigma_{j-1}, i, \sigma_{j+1}, \ldots, \sigma_m)$ for $i \in \{-1, 1\}$. Define for every $j \in \{1, \ldots, m\}$ the probability measure $\mathbb{P}^j_X$ via $\mathbb{P}^j_X(A) := \mathbb{P}_X(A \cap B_j)/\mathbb{P}_X(B_j)$ for $A \in \mathcal{B}(\mathbb{R}^d)$, and let $\mathbb{E}^j_X$ be the corresponding expectation operator. Recalling $\mathbb{P}_X(B_j) = P^{-d}$, we obtain for any $z_{t-1} \in \mathbb{R}^{(t-1)(d+1)}$ and any $g \in \mathbb{R}$ that

$$
\mathbb{P}_X(\{x \in B_j : \check{\pi}_t(x, z_{t-1}, g) \neq i\}) = \mathbb{P}^j_X(\{x : \check{\pi}_t(x, z_{t-1}, g) \neq i\})/P^d,
$$

from which we see that the sum over i in the penultimate display coincides, for every $\sigma_{-j} \in \Sigma_{m-1}$, with

$$
\frac{1}{P^d}\left(\mathbb{E}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}} \mathbb{E}_G \mathbb{E}^j_X \mathbb{1}_{\{\check{\pi}_t(x, z_{t-1}, g) = 1\}} + 1 - \mathbb{E}^{t-1}_{\pi, f_{\sigma^{1}_{-j}}} \mathbb{E}_G \mathbb{E}^j_X \mathbb{1}_{\{\check{\pi}_t(x, z_{t-1}, g) = 1\}}\right) =: \frac{1}{P^d} e(\sigma, j, t). \tag{C.68}
$$

Clearly, $e(\sigma, j, t)$ is the sum of the Type 1 and Type 2 error of the test $(x, z_{t-1}, g) \mapsto \mathbb{1}_{\{\check{\pi}_t(x, z_{t-1}, g) = 1\}}$ for

$$
H_0 : \mathbb{P}^j_X \otimes \mathbb{P}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}} \otimes \mathbb{P}_G \quad \text{against} \quad H_1 : \mathbb{P}^j_X \otimes \mathbb{P}^{t-1}_{\pi, f_{\sigma^{1}_{-j}}} \otimes \mathbb{P}_G.
$$

Using Theorem 2.2(iii) of Tsybakov (2009), we obtain

$$
\begin{aligned}
e(\sigma, j, t) &\geq \frac{1}{4} \exp\left[-\mathsf{KL}\left(\mathbb{P}^j_X \otimes \mathbb{P}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}} \otimes \mathbb{P}_G, \mathbb{P}^j_X \otimes \mathbb{P}^{t-1}_{\pi, f_{\sigma^{1}_{-j}}} \otimes \mathbb{P}_G\right)\right] \\
&= \frac{1}{4} \exp\left[-\mathsf{KL}\left(\mathbb{P}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}}, \mathbb{P}^{t-1}_{\pi, f_{\sigma^{1}_{-j}}}\right)\right],
\end{aligned} \tag{C.69}
$$

the equality following, e.g., from the Chain Rule in Lemma A.1.

To upper-bound $\mathsf{KL}\left(\mathbb{P}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}}, \mathbb{P}^{t-1}_{\pi, f_{\sigma^{1}_{-j}}}\right)$, we will now apply (C.64) with $f_1 = f_{\sigma^{-1}_{-j}}$ and $f_2 = f_{\sigma^{1}_{-j}}$. Note first that $f_1(x) = f_2(x)$ for $x \notin B_j$, and $(f_1(x), f_2(x)) = (h(0) - c_- \varepsilon \varphi_j(x), h(0) + c_- \varepsilon \varphi_j(x))$ for $x \in B_j$, from which it follows from equations (C.57) (note that $A_{f_1(x)} \leq A_{f_2(x)}$ follows from strict monotonicity of h^{-1}; cf. Step 0) and (C.60) that

$$
\mathsf{KL}(\mu_{H_{A_{f_1}(x)}}, \mu_{H_{A_{f_2}(x)}}) \leq \begin{cases} [2\zeta c_- \varepsilon \varphi_j(x)]^2, & \text{if } x \in B_j, \\ 0, & \text{if } x \notin B_j. \end{cases} \tag{C.70}
$$

Since $[2\zeta c_-\varepsilon\varphi_j(x)]^2 \le [\zeta c_-\varepsilon 2^{-1}P^{-\gamma}]^2 =: \bar{r}P^{-2\gamma}$ holds for $x \in B_j$, equation (C.64) delivers

$$\mathsf{KL}(\mathbb{P}^{t-1}_{\pi,f_{\sigma^{-1}_{-j}}}, \mathbb{P}^{t-1}_{\pi,f_{\sigma^{1}_{-j}}}) \le \bar{r}P^{-2\gamma}\int\sum_{i=1}^{t-1}\mathbb{1}\{G(i,j)\}d(\mathbb{P}_X\otimes\mathbb{P}_G\otimes\mathbb{P}^n_{\pi,f_{\sigma^{-1}_{-j}}}) \le \bar{r}P^{-2\gamma}N_{j,\sigma_{-j}},$$

with $G(i,j) := \{(x,z_n,g) : x \in B_j, \pi_i(x,z_{i-1},g) = 2\}$, $N_{j,\sigma_{-j}} := \int\sum_{i=1}^n \mathbb{1}\{G(i,j)\}d(\mathbb{P}_X\otimes \mathbb{P}_G\otimes\mathbb{P}^n_{\pi,f_{\sigma^{-1}_{-j}}})$. The dependence of $N_{j,\sigma_{-j}}$ on π has been suppressed. In combination with equations (C.68) and (C.69), we hence obtain

$$\begin{aligned}\sum_{t=1}^n Q_t^j = \sum_{t=1}^n\sum_{\sigma_{-j}\in\Sigma_{m-1}}\frac{1}{P^d}e(\sigma,j,t) &\ge \sum_{t=1}^n\sum_{\sigma_{-j}\in\Sigma_{m-1}}\frac{1}{4P^d}\exp\left[-\bar{r}P^{-2\gamma}N_{j,\sigma_{-j}}\right]\\ &= \frac{n}{4P^d}\sum_{\sigma_{-j}\in\Sigma_{m-1}}\exp\left[-\bar{r}P^{-2\gamma}N_{j,\sigma_{-j}}\right]\\ &\ge 2^{m-1}\frac{n}{4P^d}\exp\left[-\bar{r}P^{-2\gamma}\varrho_j\right],\end{aligned}$$

the last inequality following from Jensen's inequality and $\varrho_j := 2^{1-m}\sum_{\sigma_{-j}\in\Sigma_{m-1}}N_{j,\sigma_{-j}}$. Furthermore, from the definition of Q_t^j, one directly obtains via Tonelli's theorem that

$$\begin{aligned}\sum_{t=1}^n Q_t^j &= \sum_{t=1}^n\sum_{\sigma_{-j}\in\Sigma_{m-1}}\sum_{i\in\{-1,1\}}\mathbb{E}^{t-1}_{\pi,f_{\sigma^{i}_{-j}}}\mathbb{E}_G[\mathbb{P}_X(x\in B_j : \check{\pi}_t(x,z_{t-1},g_t)\neq i)]\\ &\ge \sum_{t=1}^n\sum_{\sigma_{-j}\in\Sigma_{m-1}}\mathbb{E}^n_{\pi,f_{\sigma^{-1}_{-j}}}\mathbb{E}_G[\mathbb{P}_X(x\in B_j : \pi_t(x,z_{t-1},g_t)=2)]\\ &= \sum_{\sigma_{-j}\in\Sigma_{m-1}}\mathbb{E}^n_{\pi,f_{\sigma^{-1}_{-j}}}\mathbb{E}_G\sum_{t=1}^n[\mathbb{P}_X(x\in B_j : \pi_t(x,z_{t-1},g_t)=2)]\\ &= \sum_{\sigma_{-j}\in\Sigma_{m-1}}N_{j,\sigma_{-j}} = 2^{m-1}\varrho_j.\end{aligned}$$

Combining the lower bounds in the previous two displays with (C.67) yields

$$\sup_{f\in\mathcal{C}}\mathbb{E}_{(\mathbb{P}_f\otimes\mathbb{P}_G)^n}[S_n(\pi)] \ge \frac{1}{2^m}\sum_{j=1}^m\sum_{t=1}^n Q_t^j \ge \frac{1}{2}\sum_{j=1}^m\max\left(\frac{n}{4P^d}\exp\left[-\bar{r}P^{-2\gamma}\varrho_j\right],\varrho_j\right),$$

which can further be lower-bounded by

$$\begin{aligned}\frac{1}{4}\sum_{j=1}^m\left(\frac{n}{4P^d}\exp\left[-\bar{r}P^{-2\gamma}\varrho_j\right]+\varrho_j\right) &\ge \frac{m}{4}\inf_{\varrho\ge 0}\left(\frac{n}{4P^d}\exp\left[-\bar{r}P^{-2\gamma}\varrho\right]+\varrho\right)\\ &\ge \frac{m}{4\bar{r}P^{-2\gamma}}\inf_{\varrho\ge 0}\left(\frac{n\bar{r}}{4P^{d+2\gamma}}\exp\left[-\varrho\right]+\varrho\right).\end{aligned}$$

This lower bound holds for any $P\in\mathbb{N}$ and corresponding $m = \lceil P^{d-\gamma\alpha}\rceil$. We now set $P := \lceil (n\bar{r}/4)^{1/(d+2\gamma)}\rceil$, and can thus use $w\exp(-\varrho)+\varrho \ge w$ for every $\varrho \ge 0$ and every $0 < w \le 1$

to lower-bound the quantity in the last line of the previous display by

$$\frac{mn}{16P^d} \geq \frac{P^{d-\gamma\alpha}n}{16P^d} = \frac{n}{16}P^{-\gamma\alpha} \geq \frac{n}{16}[(n\bar{r}/4)^{1/(d+2\gamma)}+1]^{-\gamma\alpha}$$
$$\geq \frac{n^{1-\frac{\alpha\gamma}{d+2\gamma}}}{16}[(\bar{r}/4)^{1/(d+2\gamma)}+1]^{-\gamma\alpha}.$$

By definition, $\bar{r} = [\zeta c_-\varepsilon 2^{-1}]^2 = [(0.5^2-\varepsilon^2)^{-1/2}\varepsilon 2^{-1}]^2$. Recalling $\varepsilon = 2/\sqrt{17}$ implies $\bar{r} = 4$. Thus, the lower bound in the previous display simplifies to

$$\frac{n^{1-\frac{\alpha\gamma}{d+2\gamma}}}{16}2^{-\gamma\alpha} \geq n^{1-\frac{\alpha\gamma}{d+2\gamma}}/32. \quad \textbf{(C.71)}$$

This establishes equation (14). Finally, Lemma C.1 (cf. Step 3, which verifies the assumptions needed) with $D_0 = 2 + C_0^{-1}$ shows that the lower bound established in Lemma C.1 holds for the corresponding constant $(1-(2+C_0^{-1})^{-1})/(2C_0+1)^{1/\alpha} \geq 2^{-1}(2C_0+1)^{-1/\alpha} \geq 2^{-(1+1/\alpha)}(C_0+1)^{-1/\alpha}$. This version of Lemma C.1 and the already established equation (14) prove equation (13).

REFERENCES

Athey, S. & S. Wager (2021) Policy learning with observational data. *Econometrica* 89, 133–161.
Audibert, J.-Y. & A.B. Tsybakov (2007) Fast learning rates for plug-in classifiers. *Annals of Statistics* 35, 608–633.
Besson, L. & E. Kaufmann (2018) What doubling tricks can and can't do for multi-armed bandits. Preprint, arXiv:1803.06971.
Bhattacharya, D. & P. Dupas (2012) Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics* 167, 168–196.
Bitler, M.P., J.B. Gelbach, & H.W. Hoynes (2006) What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review* 96, 988–1012.
Cassel, A., S. Mannor, & A. Zeevi (2018) A general approach to multi-armed bandits under risk criteria. In S. Bubeck, V. Perchet, & P. Rigollet (eds.), *Proceedings of the 31st Conference on Learning Theory*, vol. 75, pp. 1295–1306. PMLR.
Chakravarty, S.R. (2009) *Inequality, Polarization and Poverty*. Springer.
Chamberlain, G. (2000) Econometrics and decision theory. *Journal of Econometrics* 95, 255–283.
Cowell, F. (2011) *Measuring Inequality*. Oxford University Press.
Currie, J.M. & W.B. MacLeod (2020) Understanding doctor decision making: The case of depression treatment. *Econometrica* 88, 847–878.
Dehejia, R.H. (2005) Program evaluation as a decision problem. *Journal of Econometrics* 125, 141–173.
Folland, G.B. (1999) *Real Analysis: Modern Techniques and Their Applications*. Wiley.
Hirano, K. & J.R. Porter (2009) Asymptotics for statistical treatment rules. *Econometrica* 77, 1683–1701.
Hirano, K. & J.R. Porter (2020) Asymptotic analysis of statistical decision rules in econometrics. In S.N. Durlauf, L. Peter Hansen, J.J. Heckman, & R.L. Matzkin (eds.), *Handbook of Econometrics*, vol. 7A, pp. 283–354. Elsevier.
Kallenberg, O. (2001) *Foundations of Modern Probability*, *2nd Edition*. Springer Science & Business Media.
Kasy, M. & A. Sautmann (2021) Adaptive treatment assignment in experiments for policy choice. *Econometrica* 89, 113–132.

Kitagawa, T. & A. Tetenov (2018) Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* 86, 591–616.

Kitagawa, T. & A. Tetenov (2019) Equality-minded treatment choice. *Journal of Business & Economic Statistics* 39, 561–574.

Kock, A.B., D. Preinerstorfer, & B. Veliyev (2022) Functional sequential treatment allocation. *Journal of the American Statistical Association* 117, 1311–1323.

Kock, A.B. & M. Thyrsgaard (2018) Optimal sequential treatment allocation. Preprint, arXiv:1705.09952.

Lambert, P.J. (2001) *The Distribution and Redistribution of Income*. Manchester University Press.

Liese, F. & K.J. Miescke (2008) *Statistical Decision Theory*. Springer.

Ma, X., Q. Zhang, L. Xia, Z. Zhou, J. Yang, & Q. Zhao (2020) Distributional soft actor critic for risk sensitive learning. Preprint, arXiv:2004.14547.

Maillard, O.-A. (2013) Robust risk-averse stochastic multi-armed bandits. In S. Jain, R. Munos, F. Stephan, & T. Zeugmann (eds), *Algorithmic Learning Theory*, pp. 218–233. Springer.

Mammen, E. & A.B. Tsybakov (1999) Smooth discrimination analysis. *Annals of Statistics* 27, 1808–1829.

Manski, C.F. (2004) Statistical treatment rules for heterogeneous populations. *Econometrica* 72, 1221–1246.

Manski, C.F. (2019) Treatment choice with trial data: Statistical decision theory should supplant hypothesis testing. *American Statistician* 73, 296–304.

Manski, C.F. & A. Tetenov (2016) Sufficient trial size to inform clinical practice. *Proceedings of the National Academy of Sciences* 113, 10518–10523.

McDonald, J.B. (1984) Some generalized functions for the size distribution of income. *Econometrica* 52, 647–663.

McDonald, J.B. & M. Ransom (2008) The generalized beta distribution as a model for the distribution of income: Estimation of related measures of inequality. In D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves*, pp. 147–166. Springer.

Perchet, V. & P. Rigollet (2013) The multi-armed bandit problem with covariates. *Annals of Statistics* 41, 693–721.

Rigollet, P. & A. Zeevi (2010) Nonparametric bandits with covariates. In: In A.T. Kalai & M. Mohri (eds.), *Proceedings title: 23rd Annual Conference on Learning Theory*, pp. 54–66. Omnipress.

Rostek, M. (2010) Quantile maximization in decision theory. *Review of Economic Studies* 77, 339–371.

Sani, A., A. Lazaric, & R. Munos (2012) Risk-aversion in multi-armed bandits. In F. Pereira, C.J.C. Burges, L. Bottou, & K.Q. Weinberger (eds), *Advances in Neural Information Processing Systems 25*, pp. 3275–3283. Curran Associates, Inc.

Sen, A. (1974) Informational bases of alternative welfare approaches: Aggregation and income distribution. *Journal of Public Economics* 3, 387–403.

Serfling, R.J. (1984) Generalized *L*-, *M*-, and *R*-statistics. *Annals of Statistics* 12, 76–86.

Shalev-Shwartz, S. (2012) *Online learning and online convex optimization*. *Foundations and Trends® in Machine Learning* 4, 107–194.

Si, N., F. Zhang, Z. Zhou, & J. Blanchet (2020b) Distributionally robust policy evaluation and learning in offline contextual bandits. In H. Daume III & A. Singh (eds.), *International Conference on Machine Learning*, pp. 8884–8894. PMLR.

Si, N., F. Zhang, Z. Zhou, & J. Blanchet (2020a) Distributional robust batch contextual bandits. Preprint, arXiv:2006.05630.

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

Stoye, J. (2012) Minimax regret treatment choice with covariates or with limited validity of experiments. *Journal of Econometrics* 166, 138–156.

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

Thurow, L.C. (1970) Analyzing the American income distribution. *American Economic Review* 60, 261–269.

Tran-Thanh, L. & J.Y. Yu (2014) Functional bandits. Preprint, arXiv:1405.2432.

Tsybakov, A.B. (2004) Optimal aggregation of classifiers in statistical learning. *Annals of Statistics* 32, 135–166.

Tsybakov, A.B. (2009) *Introduction to Nonparametric Estimation*. Springer.

Vakili, S., A. Boukouvalas, & Q. Zhao (2019) Decision variance in online learning. Preprint, arXiv:1807.09089.

Vakili, S. & Q. Zhao (2016) Risk-averse multi-armed bandit problems under mean-variance measure. *IEEE Journal of Selected Topics in Signal Processing* 10, 1093–1111.

Woodroofe, M. (1979) A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association* 74, 799–806.

Yang, Y. & D. Zhu (2002) Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *Annals of Statistics* 30, 100–121.

Zhou, Z., S. Athey, & S. Wager (2022) Offline multi-action policy learning: Generalization and optimization. *Operations Research* 71, 148–183.

Zhou, Z., Z. Zhou, Q. Bai, L. Qiu, J. Blanchet, & P. Glynn (2021b) Finite-sample regret bound for distributionally robust offline tabular reinforcement learning. In A. Banerjee & K. Fukumizu (eds.), *International Conference on Artificial Intelligence and Statistics*, pp. 3331–3339. PMLR.

Zimin, A., R. Ibsen-Jensen, & K. Chatterjee (2014) Generalized risk-aversion in stochastic multi-armed bandits. Preprint, arXiv:1405.0833.