Functional Sequential Treatment Allocation

Anders Bredahl Kock
University of Oxford
CREATES, Aarhus University
10 Manor Rd, Oxford OX1 3UQ
anders.kock@economics.ox.ac.uk

David Preinerstorfer
ECARES, SBS-EM
Université libre de Bruxelles
50 Ave. F.D. Roosevelt CP 114/04, 1050 Brussels
david.preinerstorfer@ulb.ac.be

Bezirgen Veliyev
CREATES
Aarhus University
Fuglesangs Alle 10, 8210 Aarhus V.
bveliyev@econ.au.dk

Abstract

Consider a setting in which a policy maker assigns subjects to treatments, observing each outcome before the next subject arrives. Initially, it is unknown which treatment is best, but the sequential nature of the problem permits learning about the effectiveness of the treatments. While the multi-armed-bandit literature has shed much light on the situation when the policy maker compares the effectiveness of the treatments through their mean, economic decision making often requires targeting purpose specific characteristics of the outcome distribution, such as its inherent degree of inequality, welfare or poverty. In the present paper we introduce and study sequential learning algorithms when the distributional characteristic of interest is a general functional of the outcome distribution. In particular, it turns out that intuitively reasonable approaches, such as first conducting an experiment on an initial group of subjects followed by rolling out the inferred best treatment to the population, are dominated by the policies we develop and of which we show that they are optimal.

JEL Classification: C21, C44.

Keywords: Sequential Treatment Allocation, Distributional Characteristics, Randomized Controlled Trials, Minimax Optimal Expected Regret, Multi-Armed Bandits.

1 Introduction

In sequential decision making, a policy maker who seeks to assign subjects to the treatment with the highest mean outcome, but who initially does not know which treatment is best, can draw on a rich body of literature on "multi-armed bandits" in Machine Learning and Statistics. This literature provides assignment strategies that efficiently balance exploration and exploitation. In economic decision making, however, one often faces the problem that the quality of treatments cannot be reasonably compared according to the mean of the outcome distribution. Rather, a policy maker may want to incorporate other distributional characteristics, such as the inequality, welfare or poverty implications of a treatment, or may want to compare the quality of the treatments according to a robust measure such as the median or another quantile. While much progress has been made concerning inference on distributional policy effects, most of the literature deals with a "static" setting, in which the dataset has already been collected. Nevertheless, the subjects to be treated often arrive sequentially. For example, workers, who become unemployed throughout the year, or patients, who fall ill at different points in time, must be assigned to treatments as they arrive. Thus, the policy maker can actually adjust the composition of the data set as treatment outcomes are observed. This opportunity, which is not present in the static viewpoint, can be seized to design better policies. In this article, we consider a *sequential* treatment assignment problem and a policy maker who seeks to minimize regret compared to always assigning the unknown best treatment, according to a general *functional* of interest. That is, the policy maker wishes to use a policy that is nearly as good as the infeasible policy of always assigning the best treatment. In order to achieve a low regret, the policy maker must sequentially learn the distributional characteristic of interest for the available treatments, yet treat as many subjects as well as possible. In the following paragraphs we discuss our results. Related literature is discussed thereafter.

A common approach to solving the policy maker's sequential regret minimization problem are *explore-then-commit* policies: In an attempt to infer which treatment is best, one first conducts a randomized controlled trial (RCT) on an initial segment of subjects. Based on the outcome, one then assigns all remaining subjects to the *inferred best* treatment. Note that due to estimation uncertainty, the inferred best treatment is typically not guaranteed to be the optimal one. We show in Theorem 2.10 that if the policy maker does not know in advance the total number of treatments to be made, then the maximal expected regret of *any* explore-then-commit policy grows linearly in the number of treatments n. This is remarkable, because in our setting a linearly growing maximal expected regret is the worst possible rate that any policy can incur. Not knowing in advance the total number of treatments to be made is common, because the policy maker indeed often does not know a priori the number of subjects to be treated since, e.g., the number of individuals becoming unemployed in the course of the year is not known in advance. Theorem 2.10 furthermore shows that if the number of assignments to be made is known in advance, such that the policy maker can adapt the length of the RCT to this *horizon*, any explore-then-commit policy leads to a maximal expected regret that increases at least at the rate $n^{2/3}$. We show in Theorem 2.11 that this optimal rate is attained for an empirical-success-type decision rule, if the sample size used for experimentation is chosen of order $n^{2/3}$.

Having studied policies that strictly separate the exploration and exploitation phase, we next study a policy that interweaves these two phases. The policy is inspired by a standard algorithm in the multi-armed-bandit literature. More precisely, we study a Functional Upper Confidence Bound (F-UCB) policy, which is an extension of the Upper Confidence Bound (UCB) policy to the case where functionals other than the mean of the outcome distribution are targeted. In Theorem 2.12 we establish an upper bound on the maximal expected regret of the F-UCB policy. Most importantly, this upper bound is lower than the *lower* bounds on maximal expected regret of any explore-then-commit policy. This indicates that, in terms of maximal expected regret, one should prefer the F-UCB policy to any explore-then-commit policy. Next, by establishing lower bounds on maximal expected regret valid for all policies, we show in Theorem 2.13 that the F-UCB policy is near minimax rate optimal. The F-UCB policy achieves this near optimality even *without* knowing in advance the total number of treatments to be made in the course of the treatment period.

The finding that for large classes of functional targets explore-then-commit policies are dominated by the F-UCB policy in terms of maximal expected regret does, of course, by no means discredit RCTs and subsequent testing for other purposes. For example, RCTs are often used to test for a causal effect of a treatment, cf. Imbens and Wooldridge (2009) for an overview and further references. The goal of the present article is not to test for a causal effect, but to assist the policy maker in minimizing regret, i.e., to keep to a minimum the sum of all losses due to assigning subjects wrongly. This goal, as pointed out in, e.g., Manski (2004), Manski and Tetenov (2016) and Manski (2019b), is only weakly related to testing. For example, the policy maker may care about more than just controlling the probabilities of Type 1 and Type 2 errors. In particular the magnitude of the losses when errors occur are important components of regret.

Up to this point, we have discussed results for a situation where no covariates are observed. This provides important qualitative insights. Oftentimes in practice, however, a policy maker wants to incorporate covariate information into the decision making process. Therefore, we also consider a setting where the policy maker observes a vector of characteristics of the subject to be treated prior to treatment allocation. For each subject the best treatment is now the one that maximizes a functional of the *conditional* treatment outcome distribution given the vector of characteristics. In this setup, we show in Theorem 3.5 that policies that ignore covariates have linearly increasing regret, unless the best treatment can be chosen independently of the covariate outcome. We then introduce a version of the F-UCB policy which incorporates covariate information. In Theorem 3.6 and Corollary 3.7 we study how this generalized F-UCB policy performs in terms of its maximal expected regret. By deriving lower bounds on maximal expected regret in Theorem 3.8, we establish the near minimax rate optimality of the generalized F-UCB policy. We also show in Theorem 3.3 that the assumptions under which the upper bounds on regret are established are essentially minimal. Even a slight relaxation of the assumptions implies that every policy incurs the worst-case linear maximal expected regret.

The results discussed in the preceding paragraph are obtained without restricting the similarity of the best and second best treatment. Intuitively, this similarity crucially influences the difficulty of the decision problem. One may thus ask whether the F-UCB policy automatically adapts in an optimal way to the degree of similarity. To answer this question affirmatively, we first derive an upper bound on maximal expected regret of the F-UCB policy over classes of distributions that restrict the level of similarity (cf. Theorem 3.12), and then establish matching lower bounds in Theorem 3.14. Furthermore, over such restricted classes of distributions, it is now possible to show that the expected number of suboptimal assignments made increases as slowly as possible in the number of assignments (cf. Theorem 3.13). The latter result can be interpreted as an ethical guarantee on the F-UCB policy: only few persons will receive a treatment which is not optimal for them.

1.1 Related literature

Our paper is related to several strands of literature. In econometrics, our work relates to the literature on statistical treatment rules. This literature has focused on developing policies, the goal of which is to ensure near minimax optimal regret. Here Manski (2004) did seminal work in proposing conditional empirical success rules, and in giving sufficient conditions for productive use of covariates in terms of minimax optimality. Stoye (2009) shed further light on this by an exact finite sample analysis of the problem, while Hirano and Porter (2009) studied asymptotically optimal decisions in the limiting experiment. Furthermore, our work is related to the recent paper by Kitagawa and Tetenov (2018), who focus on minimax optimal treatment rules, when these are restricted to belong to certain classes for, e.g., ethical, legislative or political reasons. Athey and Wager (2017) have used concepts from semiparametric efficiency theory to construct asymptotically minimax optimal policies, even when propensity scores are unknown. Other papers on statistical treatment rules in econometrics include Chamberlain (2000); Dehejia (2005); Bhattacharya and Dupas (2012); Stoye (2012); Tetenov (2012). A recent overview of statistical treatment rules in economics is provided by Hirano and Porter (2018).

While we also study minimax optimal expected regret properties of statistical treatment rules, one distinguishing feature of the setting studied in the present article is that the subjects to be treated arrive sequentially. Thus, we do not presuppose that a data set of size n is at our disposal from the outset, based on which the best treatment must be inferred. Instead the dataset is gradually constructed during the learning process. We emphasize that our sequential setting is different from the dynamic one in, e.g., Robins (1997), Lavori et al. (2000), Murphy et al. (2001), Murphy (2003) and Murphy (2005), where the *same* subjects are treated repeatedly.

Another distinguishing feature of the setting studied in the present article is that we focus on the problem of a policy maker who targets a general functional of the outcome distribution of the treatments, as opposed to only focusing on the mean. The importance of considering other characteristics of the outcome distribution than its mean has previously been underscored in a non-sequential setting by many contributions, e.g., Gastwirth (1974), Manski (1988), Thistle (1990), Mills and Zandvakili (1997), Davidson and Duclos (2000), Abadie et al. (2002), Abadie (2002), Chernozhukov and Hansen (2005), Davidson and Flachaire (2007), Barrett and Donald (2009), Hirano and Porter (2009), Schluter and van Garderen (2009), Rostek (2010), Rothe (2010, 2012), Kitagawa and Tetenov (2017) and Manski (2019a). Concerning the functionals we permit, our theory is very general and we verify that it covers many inequality, welfare, and poverty measures, such as the Schutz coefficient, the Atkinson-, Gini- and Kolm-indices (cf. Section 4 for details). We also show that our theory covers quantiles, U-functionals and generalized L-functionals, cf. Appendix D. In contrast to much of the existing theoretical results concerning inference on inequality, welfare, or poverty measures, we do not investigate (first or higher-order) asymptotic approximations, but we establish exact finite sample results with explicit constants. To this end we cannot rely on classical asymptotic techniques, e.g., distributional approximations based on linearization arguments.

The sequential setting adopted in this paper is akin to the one encountered in multi-armed bandit problems studied in machine learning and statistics. In terms of theoretical performance guarantees, Robbins (1952) pioneered this literature by introducing an algorithm whose average reward will converge to that of the best arm. One of the policies studied in the present paper, the F-UCB policy, resembles the Upper Confidence Bound (UCB) strategy of Lai and Robbins (1985) (to which it reduces in case of the mean functional); cf. also Auer et al. (2002) for finite-sample results concerning the regret properties of the UCB strategy. There are relatively few papers on multi-armed-bandit problems in which a functional of the reward distribution is targeted instead of the mean: Cassel et al. (2018) consider bandit problems, where the target can be a general (risk) measure defined on the empirical distribution functions of the path of assignments. Furthermore, Tran-Thanh and Yu (2014) consider a multi-armed bandit problem, where the target is a functional, e.g., some risk or information measures, but where a completely different regret environment is considered. Besides targeting different types of regret than we do, neither of the two articles studies optimality of the algorithms proposed (i.e., minimax lower bounds, which are crucial for our purpose), or how covariate information can be incorporated.

While the F-UCB policy is inspired by the UCB policy, the analysis of it poses new challenges, as witnessed in the proof of Theorem 2.12 and its supporting lemmas. We also stress that the lower bounds on expected maximal regret are throughout established under very weak conditions. In particular, they hold over very small, actually parametric, classes of distributions that depend on the functional under consideration. This settles rather firmly what can and cannot be achieved in a functional sequential treatment assignment problem. Furthermore, we provide a rigorous comparison (in terms of upper and lower bounds on maximal expected regret) of explore-then-commit policies based on an RCT to bandit-type policies. Explore-then-commit policies were recently studied (in a 2-arm Gaussian setting, without covariates, and targeting the mean) by Garivier et al. (2016). Besides not being restricted to a parametric setup and the mean functional as a target, our results contribute to this literature by characterizing optimal explore-then-commit policies in terms of the length of the exploration period. This can be useful in case a policy maker must use a policy of this type.

As we also study a setting in which the distribution of treatment outcomes can depend on covariates, the works of Rigollet and Zeevi (2010) and Perchet and Rigollet (2013) are related to our paper. Both consider a setting where one targets the distribution with the highest conditional mean. Kock and Thyrsgaard (2017) consider a setting in which the policy maker is also interested in how risky a treatment is and takes this into account by targeting a tradeoff between expected outcome and variance of the treatments (a very specific class of functionals, which excludes many measures of relevance for economic decision making).

For a good general overview of multi-armed bandit problems we refer to, e.g., Bubeck and Cesa-Bianchi (2012) and Lattimore and Szepesvári (2019).

2 Functional sequential treatment allocation without covariates

In this section we consider a treatment allocation problem without covariates. This stripped case allows us to introduce and motivate the most important concepts, assumptions, and terminology in a fairly clean environment. Furthermore, the setting without covariates is sufficiently rich to demonstrate certain differences in performance of policies that fully exploit the sequential nature of the treatment problem to policies based on randomized controlled trials and, more generally, to explore-then-commit policies. The results in the present section are also instrumental for establishing some of the results in later sections.

2.1 Setup

We consider a setting, where at each point in time $t = 1, \ldots, n$ a policy maker must assign a subject to one out of K treatments. Each subject is only treated once. Thus, the index t can equivalently be thought of as indexing subjects instead of time. The observational structure is the one of a multi-armed bandit problem: After assigning a treatment, its outcome is observed, but the policy maker does not observe the counterfactuals. Having observed the outcomes of treatments $1, \ldots, t-1$, subject t arrives, and must be assigned to a treatment. The assignment can be based on the information gathered from all *previous* assignments and their outcomes, and, potentially, randomization. Thus, the data set is gradually constructed in the course of the treatment program. Without knowing a priori the identity of the "best" treatment, the policy maker seeks to assign subjects to treatments so as to minimize maximal expected regret (which we introduce in Equation (3) further below).

This setting is a sequential version of the potential outcomes framework with multiple treatments. Note also that restricting attention to problems where only one of the K treatments can be assigned does not exclude that a treatment consists of a combination of several other treatments (for example a combination of several drugs) — one simply defines this combined treatment as a separate treatment at the expense of increasing the set of treatments.

The precise setup is as follows: let the random variable $Y_{i,t}$ denote the outcome of assigning subject $t \in \{1, \ldots, n\}$ to treatment $i \in \mathcal{I} := \{1, \ldots, K\}$.[1] That is, the potential outcomes of subject t are $Y_t := (Y_{1,t}, \ldots, Y_{K,t})$. We assume that $a \leq Y_{i,t} \leq b$, where $a < b$ are real numbers. Furthermore, for every t, let G_t be a random variable, which can be used for randomization in assigning the t-th subject. *Throughout Section 2, we assume that* Y_t for $t \in \mathbb{N}$ are independent and identically distributed (i.i.d.); and we assume that the sequence G_t is i.i.d., and is independent of the sequence Y_t. Note that no assumptions are imposed

[1] We do not explicitly consider the case of individuals arriving in batches. However, in our setup, one may also interpret $Y_{i,t}$ as a summary statistic of the outcomes of batch t, when all of its subjects were assigned to treatment i. For a more sophisticated way of handling batched data in case of targeting the mean treatment outcome, we refer to Perchet et al. (2016).

concerning the dependence between the components of each random vector Y_t. We think of the *randomization measure*, i.e., the distribution of G_t, as being fixed, e.g., the uniform distribution on $[0,1]$. We denote the cumulative distribution function (cdf) of $Y_{i,t}$ by $F^i \in D_{cdf}([a,b])$, where $D_{cdf}([a,b])$ denotes the set of all cdfs F such that $F(a-)=0$ and $F(b)=1$. The cdfs F^i for $i=1,\ldots,K$ are unknown to the policy maker.

A *policy* is a triangular array of (measurable) functions $\pi = \{\pi_{n,t} : n \in \mathbb{N}, 1 \leq t \leq n\}$. Here $\pi_{n,t}$ denotes the assignment of the t-th subject out of n subjects. In each row of the array, i.e., for each $n \in \mathbb{N}$, the assignment $\pi_{n,t}$ can depend only on previously observed treatment outcomes and randomization. Formally,

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

Given a policy π and $n \in \mathbb{N}$, the input to $\pi_{n,t}$ is denoted as (Z_{t-1}, G_t), where Z_{t-1} is defined recursively: The first treatment $\pi_{n,1}$ is a function of G_1 alone, as no treatment outcomes have been observed yet (we may interpret $(Z_0, G_1) = G_1$). The second treatment is a function of the outcome of the first treatment $Z_1 := Y_{\pi_{n,1}(G_1),1}$ and of G_2. For $t \geq 3$ we have

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

The $(t-1)$-dimensional random vector Z_{t-1} can be interpreted as the information available after the $(t-1)$-th treatment outcome was observed. We emphasize that Z_{t-1} depends on the policy π via $\pi_{n,1},\ldots,\pi_{n,t-1}$. In particular, Z_{t-1} also depends on n and on $G_1,\ldots,G_{t-1}$, which we do not show in our notation. For convenience, the dependence of $\pi_{n,t}(Z_{t-1},G_t)$ on Z_{t-1} and G_t is often suppressed, i.e., we often abbreviate $\pi_{n,t}(Z_{t-1},G_t)$ by $\pi_{n,t}$ if it is clear from the context that the actual assignment $\pi_{n,t}(Z_{t-1},G_t)$ is meant, instead of the function defined in Equation (1).

Remark 2.1 (Concerning the dependence of $\pi_{n,t}$ on the horizon n)**.** We have chosen to allow the assignments $\pi_{n,1},\ldots,\pi_{n,n}$ to depend on n, the total number of assignments to be made. Consequently, for $n_1 < n_2$ it may be that $\{\pi_{n_1,t} : 1 \leq t \leq n_1\}$ does not coincide with the first n_1 elements of $\{\pi_{n_2,t} : 1 \leq t \leq n_2\}$. This is crucial, as a policy maker who knows n may choose different sequences of allocations for different n. For example, one may wish to explore the efficacies of the available treatments in more detail if one knows that the total sample size is large, such that there is much opportunity to benefit from this knowledge later on. We emphasize that while our setup allows us to study policies that make use of n, we devote much attention to policies that do not. The latter subclass of policies is important. For example, a policy maker may want to run a treatment program for a year, say, but it is unknown in advance how many subjects will arrive to be treated. In such a situation, one needs a policy that works well irrespective of the unknown horizon. Such policies are called "anytime policies," as $\pi_t := \pi_{n,t}$ does not depend on n.

The ideal solution of the policy maker would be to assign every subject to the "best" treatment. In the present paper, this is understood in the sense that the outcome distribution for the best treatment maximizes a given functional

$$\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}. \tag{2}$$

We do not assume that the maximizer is unique, i.e., $\arg\max_{i\in\mathcal{I}} \mathsf{T}(F^i)$ need not be a singleton. The specific functional chosen by the policy maker will depend on the application, and encodes particular characteristics of the distribution that the policy maker is interested in. For a streamlined presentation of our results it is helpful to keep the functional T abstract at this point (see Section 2.1.1 below for a specific example, and a brief overview of examples we study in detail).

The ideal solution of the policy maker of always choosing the best treatment is infeasible, simply because it is not known in advance which treatment is best. Therefore, every policy will make mistakes. To compare different policies, we define the *regret* of a policy π at horizon n as

$$R_n(\pi) = R_n(\pi; F^1, \ldots, F^K, Z_{n-1}, G_1, \ldots, G_n) = \sum_{t=1}^{n} \left[\max_{i\in\mathcal{I}} \mathsf{T}(F^i) - \mathsf{T}(F^{\pi_{n,t}(Z_{t-1}, G_t)}) \right]. \quad (3)$$

The unknown outcome distributions $F^1, \ldots, F^K$ are assumed to vary in a pre-specified class of cdfs. Following the minimax-paradigm, we evaluate policies according to their worst-case behavior over such classes. We refer to Manski and Tetenov (2016) for further details concerning the minimax point-of-view in the context of treatment assignment problems, and for a comparison with other approaches such as the Bayesian. Formally, we seek a policy π that minimizes maximal expected regret, that is, a policy that minimizes

$$\sup_{\substack{F^i \in \mathscr{D} \\ i=1,\ldots,K}} \mathbb{E}[R_n(\pi)],$$

where $\mathscr{D}$ is a subset of $D_{cdf}([a,b])$. The supremum is taken over all potential outcome vectors Y_t such that the marginals $Y_{i,t}$ for $i = 1, \ldots, K$ have a cdf in $\mathscr{D}$. The set $\mathscr{D}$ will typically be nonparametric, and corresponds to the assumptions one is willing to impose on the cdfs of each treatment outcome, i.e., on $F^1, \ldots, F^K$. Note that the maximal expected regret of a policy π as defined in the previous display depends on the horizon n. We will study this dependence on n. In particular, we will study the *rate* at which the maximal expected regret increases in n for a given policy π; furthermore, we will study the question of which kind of policy is optimal in the sense that the rate is optimal.

The following assumption is the main requirement we impose on the functional T and the set $\mathscr{D}$. We denote the supremum metric on $D_{cdf}([a,b])$ by $\|\cdot\|_\infty$, i.e., for cdfs F and G we let $\|F - G\|_\infty = \sup_{x\in\mathbb{R}} |F(x) - G(x)|$.

Assumption 2.2. *The functional $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ and the non-empty set $\mathscr{D} \subseteq D_{cdf}([a,b])$ satisfy*

$$|\mathsf{T}(F) - \mathsf{T}(G)| \leq C\|F - G\|_\infty \quad \text{for every} \quad F \in \mathscr{D} \text{ and every } G \in D_{cdf}([a,b]) \quad (4)$$

for some $C > 0$.

Remark 2.3 (Restricted-Lipschitz continuity)**.** Assumption 2.2 implies that the functional T is Lipschitz continuous when restricted to $\mathscr{D}$ (the domain being equipped with $\|\cdot\|_\infty$). We emphasize, however, that if $\mathscr{D} \neq D_{cdf}([a,b])$, the functional T is not necessarily required to be Lipschitz-continuous on all of $D_{cdf}([a,b])$. This is due to the asymmetry inherent in the condition imposed in Equation (4), where F varies only in $\mathscr{D}$, but G varies in all of $D_{cdf}([a,b])$.

Remark 2.4. A simple approximation argument[2] shows that if Assumption 2.2 is satisfied with $\mathscr{D}$ and C, then Assumption 2.2 is also satisfied with $\mathscr{D}$ replaced by the closure of $\mathscr{D} \subseteq D_{cdf}([a,b])$ (the ambient space $D_{cdf}([a,b])$ being equipped with the metric $\|\cdot\|_\infty$) and the same constant C.

Remark 2.5. The set $\mathscr{D}$ encodes assumptions imposed on the cdfs of each treatment outcome. In particular, the larger $\mathscr{D}$, the less restrictive is $F^i \in \mathscr{D}$ for $i \in \mathcal{I}$. Ideally, one would thus like $\mathscr{D} = D_{cdf}([a,b])$, which, however, is too much to ask for some functionals. Furthermore, there is a trade-off between the sizes of C and $\mathscr{D}$, in the sense that a larger class $\mathscr{D}$ typically requires a larger constant C. The reader who wants to get an impression of some of the classes of cdfs we consider may want to consult Section 4.1, where several important classes of cdfs are defined.

2.1.1 Functionals that satisfy Assumption 2.2: A summary of results in Section 4 and Appendix D

In the present paper, we do not contribute to the construction of functionals for specific questions. Rather, we take the functional as given. To choose an appropriate functional, the policy maker can already draw on a very rich and still expanding body of literature; cf. Lambert (2001), Chakravarty (2009) or Cowell (2011) for textbook-treatments. To equip the reader with a specific and important example of a functional T, one may think of the Gini-welfare measure (cf. Sen (1974))

$$\mathsf{T}(F) = \int x dF(x) - \frac{1}{2} \int \int |x_1 - x_2| dF(x_1) dF(x_2). \tag{5}$$

Because all of our results impose Assumption 2.2, a natural question concerns its generality. To convince the reader and practitioner that Assumption 2.2 is often satisfied, and to make the policies studied implementable (as they require knowledge of C), we show in Section 4 that Assumption 2.2 is satisfied for many important inequality, welfare, and poverty measures (together with formal results concerning the sets $\mathscr{D}$ along with corresponding constants C). For example, it is shown that for the above Gini-welfare measure, Assumption 2.2 is satisfied with $\mathscr{D} = D_{cdf}([a,b])$, i.e., without any restriction on the treatment cdfs $F^1, \ldots, F^K$ (apart from having support $[a,b]$), and with constant $C = 2(b-a)$. At this point we highlight some further functionals that satisfy Assumption 2.2:

1. The *inequality measures* we discuss in Section 4.2 include the Schutz-coefficient (Schutz (1951), Rosenbluth (1951)), the Gini-index, the class of linear inequality measures of Mehran (1976), the generalized entropy family of inequality indices including Theil's index, the Atkinson family of inequality indices (Atkinson (1970)), and the family of Kolm-indices (Kolm (1976a)). In many cases, we discuss both relative and absolute versions of these measures.

2. In Section 4.3 we provide results for *welfare measures* based on inequality measures.

[2] Let $\bar{F} \in D_{cdf}([a,b])$ be such that $\|F_m - \bar{F}\|_\infty \to 0$ as $m \to \infty$ for a sequence $F_m \in \mathscr{D}$, and let $G \in D_{cdf}([a,b])$. Then, $|\mathsf{T}(\bar{F}) - \mathsf{T}(G)| \leq |\mathsf{T}(\bar{F}) - \mathsf{T}(F_m)| + |\mathsf{T}(F_m) - \mathsf{T}(G)|$, which, by Assumption 2.2, is not greater than $2C\|\bar{F} - F_m\|_\infty + C\|\bar{F} - G\|_\infty \to C\|\bar{F} - G\|_\infty$ as $m \to \infty$.

3. The *poverty measures* we discuss in Section 4.4 are the headcount ratio, the family of poverty measures of Sen (1976) in the generalized form of Kakwani (1980), and the family of poverty measures suggested by Foster et al. (1984).

The results in Sections 4.2, 4.3, and 4.4 mentioned above are obtained from and supplemented by a series of general results that we develop in Appendix D. These result verify Assumption 2.2 for *U-functionals* defined in Equation (107) (i.e., population versions of U-statistics), *quantiles*, generalized *L-functionals* due to Serfling (1984) defined in Equation (112), and *trimmed U-functionals* defined in Equation (117). These techniques are of particular interest in case one wants to apply our results to functionals T that we do not explicitly discuss in Section 4.

The results in Section 4 and Appendix D could also be of independent interest, because they immediately allow the construction of uniformly valid (over $\mathscr{D}$) confidence intervals and tests in finite samples. To see this, observe that Assumption 2.2 together with the measurability Assumption 2.6 given below and the Dvoretzky-Kiefer-Wolfowitz-Massart inequality in Massart (1990) implies that, uniformly over $F \in \mathscr{D}$, the confidence interval $\mathsf{T}(\hat{F}_n) \pm C\sqrt{\log(2/\alpha)/(2n)}$ covers $\mathsf{T}(F)$ with probability not smaller than $1-\alpha$; here $\hat{F}_n$ denotes the empirical cdf based on an i.i.d. sample of size n from F.

2.1.2 Further notation and an additional assumption

Before we consider maximal expected regret properties of certain classes of policies, we need to introduce some more notation: Given a policy π and $n \in \mathbb{N}$, we denote the number of times treatment i has been assigned up to time t by

$$S_{i,n}(t) := \sum_{s=1}^{t} \mathbb{1}\{\pi_{n,s}(Z_{s-1}, G_s) = i\}, \tag{6}$$

and we abbreviate $S_{i,n}(n) = S_i(n)$. Defining the loss incurred due to assigning treatment i instead of an optimal one by $\Delta_i := \max_{k \in \mathcal{I}} \mathsf{T}(F^k) - \mathsf{T}(F^i)$, the regret $R_n(\pi)$, which was defined in Equation (3), can equivalently be written as

$$R_n(\pi) = \sum_{i:\Delta_i>0} \Delta_i \sum_{t=1}^{n} \mathbb{1}\{\pi_{n,t}(Z_{t-1}, G_t) = i\} = \sum_{i:\Delta_i>0} \Delta_i S_i(n). \tag{7}$$

On the event $\{S_{i,n}(t) > 0\}$ we define the empirical cdf based on the outcomes of all subjects in $\{1, \ldots, t\}$ that have been assigned to treatment i

$$\hat{F}_{i,t,n}(z) := S_{i,n}^{-1}(t) \sum_{\substack{1 \le s \le t \\ \pi_{n,s}(Z_{s-1}, G_s) = i}} \mathbb{1}\{Y_{i,s} \le z\}, \quad \text{for every } z \in \mathbb{R}. \tag{8}$$

Note that the random sampling times s such that $\pi_{n,s}(Z_{s-1}, G_s) = i$ depend on previously observed treatment outcomes.

We shall frequently need an assumption that guarantees that the functional T evaluated at empirical cdfs, such as $\hat{F}_{i,t,n}$ just defined in Equation (8), is measurable.

Assumption 2.6. *For every $m \in \mathbb{N}$, the function on $[a,b]^m$ that is defined via $x \mapsto \mathsf{T}(m^{-1}\sum_{j=1}^m \mathbb{1}\{x_j \leq \cdot\})$, i.e., T evaluated at the empirical cdf corresponding to $x_1, \ldots, x_m$, is Borel measurable.*

Assumption 2.6 is typically satisfied and imposes no practical restrictions.

Finally, and following up on the discussion in Remark 2.1, we shall introduce some notational simplifications in case a policy π is such that $\pi_{n,t}$ is independent of n, i.e., is an anytime policy. It is then easily seen that the random quantities $S_{i,n}(t)$ and $\hat{F}_{i,t,n}$ do not depend on n, as long as t and n are such that the quantities are well defined, i.e., as long as $n \geq t$. Therefore, for such policies, we shall drop the index n in these quantities.

2.2 Explore-then-commit policies

A natural approach to assigning subjects to treatments in our sequential setup would be to first conduct a randomized controlled trial (RCT) to study which treatment is best, and then to use the acquired knowledge to assign the inferred best treatment to all remaining subjects. Such policies are special cases of *explore-then-commit* policies, which we study next. Informally, an explore-then-commit policy deserves its name as it (i) uses the first n_1 subjects to *explore*, in the sense that every treatment is assigned, in expectation, at least proportionally to n_1; and (ii) then *commits* to a single (inferred best) treatment after the first n_1 treatments have been used for exploration. Here, n_1 may depend on the horizon n.

Formally, an explore-then-commit policy is defined as follows.

Definition 2.7 (Explore-then-commit policy)**.** *A policy π is an* explore-then-commit policy*, if there exists a function $n_1 : \mathbb{N} \to \mathbb{N}$ and an $\eta \in (0,1)$, such that for every $n \in \mathbb{N}$ we have that $n_1(n) \leq n$, and such that the following conditions hold for every $n \geq K$:*

1. ***Exploration Condition**: We have that*

$$\inf_{\substack{F^i \in \mathscr{D} \\ i=1,\ldots,K}} \inf_{j \in \mathcal{I}} \mathbb{E}[S_{j,n}(n_1(n))] \geq \eta n_1(n).$$

   Here, the first infimum is taken over all potential outcome vectors Y_t such that the marginals $Y_{i,t}$ for $i = 1, \ldots, K$ have a cdf in $\mathscr{D}$.

   [That is, regardless of the (unknown) underlying marginal distributions of the potential outcomes, each arm is assigned, in expectation, at least $\eta n_1(n)$ times among the first $n_1(n)$ subjects.]

2. ***Commitment Condition**: There exists a function $\pi_n^c : [a,b]^{n_1(n)} \to \mathcal{I}$ such that, for every $t = n_1(n)+1, \ldots, n$, we have*

$$\pi_{n,t}(z_{t-1}, g) = \pi_n^c(z_{n_1(n)}) \quad \text{for every } z_{t-1} \in [a,b]^{t-1} \text{ and every } g \in \mathbb{R},$$

   where $z_{n_1(n)}$ is the vector of the last $n_1(n)$ coordinates of z_{t-1}.

   [That is, the subjects $t = n_1(n)+1, \ldots, n$ are all assigned to the same treatment, which is selected based on the $n_1(n)$ outcomes observed during the exploration period.]

It would easily be possible to let the commitment rule π_n^c depend on further external randomization. For simplicity, we omit formalizing such a generalization. We shall now discuss some important examples of explore-then-commit policies.

Example 2.8. A policy that first conducts an RCT based on a sample of $n_1(n) \leq n$ subjects, followed by *any* assignment rule for subjects $n_1(n)+1, \ldots, n$ that satisfies the commitment condition in Definition 2.7, is an explore-then-commit policy, provided the concrete randomization scheme used encompasses sufficient exploration. In particular, $\pi_{n,t}(Z_{t-1}, G_t) = G_t$ with $\mathbb{P}(G_t = i) := \frac{1}{K}$ for every $1 \leq t \leq n_1(n)$ and every $i \in \mathcal{I}$ satisfies the exploration condition in Definition 2.7 with $\eta = \frac{1}{K}$; more generally, if

$$\inf_{i \in \mathcal{I}} \mathbb{P}(G_t = i) > 0,$$

then Definition 2.7 is satisfied with η the infimum in the previous display. Alternatively, a policy that enforces balancedness in the exploration phase through assigning subjects $t = 1, \ldots, n_1(n)$ to treatments "cyclically," i.e., $\pi_{n,t}(Z_{t-1}, G_t) = (t \mod K) + 1$, satisfies the exploration condition in Definition 2.7 with $\eta = 1/(2K)$ if $n_1(n) \geq K$ for every $n \geq K$. Concrete choices for commitment rules for subjects $n_1(n)+1, \ldots, n$ include:

1. In case $K = 2$, a typical approach is to assign the fall-back treatment if, according to some test, the alternative treatment is not significantly better, and to assign the alternative treatment if it is significantly better. The sample size $n_1(n)$ used in the RCT is typically chosen to ensure that the specific test used achieves a desired power against a certain effect size. We refer to the description of the ETC-T policy in Section 2.4.1 for a specific example of a test and a corresponding rule for choosing n_1.

2. As an alternative to test-based commitment rules, one can use an *empirical success rule* as in Manski (2004), which in our general context amounts to assigning an element of $\arg\max_{i \in \mathcal{I}} \mathsf{T}(\hat{F}_{i,n_1(n),n})$ to subjects $n_1(n)+1, \ldots, n$. A specific example of such a policy, together with a concrete way of choosing n_1, is discussed in the description of the ETC-ES policy in Section 2.4.1.

We now consider the case $K = 2$ and establish regret lower bounds for the class of explore-then-commit policies. To exclude trivial cases, we assume that $\mathscr{D}$ (which is typically convex) contains a line segment on which the functional T is not everywhere constant.

Assumption 2.9. *The functional* $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ *satisfies Assumption 2.2, and* $\mathscr{D}$ *contains two elements* H_1 *and* H_2*, such that*

$$J_\tau := \tau H_1 + (1-\tau) H_2 \in \mathscr{D} \quad \text{for every } \tau \in [0,1], \tag{9}$$

and such that $\mathsf{T}(H_1) \neq \mathsf{T}(H_2)$.

Since there only have to exist *two* cdfs H_1 and H_2 as in Assumption 2.9, this is a condition that is practically always satisfied.

The next theorem considers general explore-then-commit policies, as well as the subclass of policies where $n_1(n) \leq n^*$ holds for every $n \in \mathbb{N}$ for some $n^* \in \mathbb{N}$. This subclass models situations, where the horizon n is unknown or ignored in planning the experiment, and the envisioned number of subjects used for exploration n^* is fixed in advance (here $n_1(n) = n^*$ for every $n \geq n^*$, and $n_1(n) = n$, else); the subclass also models situations where the sample size that can be used for experimentation is limited due to budget constraints.

Theorem 2.10. *Suppose $K = 2$ and that Assumption 2.9 holds. Then the following statements hold:*[3]

1. *There exists a constant $c_l > 0$, such that, for every explore-then-commit policy π that satisfies the exploration condition with $\eta \in (0, 1)$, and for any randomization measure, it holds that*

$$\sup_{\substack{F^i \in \{J_\tau : \tau \in [0,1]\} \\ i=1,2}} \mathbb{E}[R_n(\pi)] \geq \eta c_l n^{2/3} \quad \textit{for every } n \geq 2.$$

2. *For every $n^* \in \mathbb{N}$ there exists a constant $c_l = c_l(n^*)$, such that, for every explore-then-commit policy π that satisfies (i) the exploration condition with $\eta \in (0, 1)$ and (ii) $n_1(\cdot) \leq n^*$, and for any randomization measure, it holds that*

$$\sup_{\substack{F^i \in \{J_\tau : \tau \in [0,1]\} \\ i=1,2}} \mathbb{E}[R_n(\pi)] \geq \eta c_l n \quad \textit{for every } n \geq 2.$$

The first part of Theorem 2.10 shows that, under the minimal assumption of $\mathscr{D}$ containing a line segment on which T is not constant, *any* explore-then-commit policy must incur maximal expected regret that increases at least of order $n^{2/3}$ in the horizon n. The second part implies in particular that when n is unknown, such that the exploration period n_1 can not depend on it, any explore-then-commit policy must incur linear maximal expected regret. We note that this is the worst possible rate of regret, since by Assumption 2.2 *no* policy can have larger than linear maximal expected regret.

The lower bounds on maximal expected regret are obtained by taking the maximum only over all potential outcome vectors with marginal distributions in the line segment in Equation (9), i.e., a one-parametric subset of $\mathscr{D}$.

We now prove that the maximal expected regret of rate $n^{2/3}$ is attainable in the class of explore-then-commit policies. In particular, employing an empirical success type commitment rule after the exploration phase as discussed in Example 2.8 yields a maximal expected regret of this order. To be precise, we consider the following policy, which in contrast to test-based commitment rules (which require the choice of a suitable test and taking into account multiple-comparison issues in case $K > 2$) can be implemented seamlessly for any number of treatments:

[3] The constants c_l depend on properties of the function $\tau \mapsto \mathsf{T}(J_\tau)$ for $\tau \in [0, 1]$. More specifically, the constants depend on the quantities ε and c_- from Lemma A.4. The precise dependence is made explicit in the proof.

Policy 1: Explore-then-commit empirical-success policy $\tilde{\pi}$

for $t = 1, \ldots, n_1(n) := \min(K\lceil n^{2/3}\rceil, n)$ **do**
  assign $\tilde{\pi}_{n,t}(Z_{t-1}, G_t) = G_t$, with G_t uniformly distributed on $\mathcal{I}$
end
for $t = n_1(n) + 1, \ldots, n$ **do**
  assign $\tilde{\pi}_{n,t}(Z_{t-1}, G_t) = \min \underset{i: S_{i,n}(n_1(n))>0}{\arg\max} \mathsf{T}(\hat{F}_{i,n_1(n),n})$
end

Note that the policy $\tilde{\pi}$ is an explore-then-commit policy that requires knowledge of the horizon n, which by Theorem 2.10 is necessary for obtaining a rate slower than n. The outer minimum in the second for loop in the policy is just taken to break ties (if necessary). Our result concerning $\tilde{\pi}$ is as follows (an identical statement can be established for a version of $\tilde{\pi}$ with cyclical assignment during the exploration phase as discussed in Remark 2.8; the proof follows along the same lines, and we skip the details).

Theorem 2.11. *Under Assumptions 2.2 and 2.6, the explore-then-commit empirical-success policy $\tilde{\pi}$ satisfies*

$$\sup_{\substack{F^i \in \mathscr{D} \\ i=1,\ldots,K}} \mathbb{E}[R_n(\tilde{\pi})] \leq 6CKn^{2/3} \quad \textit{for every } n \in \mathbb{N}. \tag{10}$$

Theorems 2.10 and 2.11 together prove that within the class of explore-then-commit policies, the policy $\tilde{\pi}$ is rate optimal in n. We shall next show that policies which do not separate the exploration and commitment phase can obtain lower maximal expected regret. In this sense, the natural idea of separating exploration and commitment phases turns out to be suboptimal from a decision-theoretic point-of-view in functional sequential treatment assignment problems.

2.3 Functional UCB policy and regret bounds

We now introduce the *Functional Upper Confidence Bound* (F-UCB) policy and study its properties. It is inspired by the UCB strategy of Lai and Robbins (1985) for multi-armed bandit problems. While the UCB policy was designed for targeting the mean of a distribution, the F-UCB policy can target any functional (and reduces to the UCB policy in case one targets the mean). The F-UCB policy has the practical advantage of not needing to know the horizon n, cf. Remark 2.1 (recall also the notation introduced in Section 2.1.2). Furthermore, no external randomization is required, which will therefore be notationally suppressed as an argument to the policy. The policy is defined as follows, where C is the constant from Assumption 2.2.

Policy 2: F-UCB policy $\hat{\pi}$

Input: $\beta > 2$
for $t = 1, \ldots, K$ **do**
  assign $\hat{\pi}_t(Z_{t-1}) = t$
end
for $t \geq K + 1$ **do**
  assign $\hat{\pi}_t(Z_{t-1}) = \min \arg\max_{i \in \mathcal{I}} \left\{ \mathsf{T}(\hat{F}_{i,t-1}) + C\sqrt{\beta \log(t)/(2S_i(t-1))} \right\}$
end

After the K initialization rounds, the F-UCB policy assigns a treatment that i) is promising, in the sense that $\mathsf{T}(\hat{F}_{i,t-1})$ is large, or ii) has not been well explored, in the sense that $S_i(t-1)$ is small. The parameter β is chosen by the researcher and indicates the weight put on assigning scarcely explored treatments, i.e., treatments with low $S_i(t-1)$. An optimal choice of β, minimizing the upper bound on maximal expected regret, is given after Theorem 2.12 below. We use the notation $\overline{\log}(x) := \max(\log(x), 1)$ for $x > 0$.

Theorem 2.12. *Under Assumptions 2.2 and 2.6, the F-UCB policy $\hat{\pi}$ satisfies*

$$\sup_{\substack{F^i \in \mathscr{D} \\ i=1,\ldots,K}} \mathbb{E}[R_n(\hat{\pi})] \leq c\sqrt{Kn\overline{\log}(n)} \quad \text{for every } n \in \mathbb{N},$$

where $c = c(\beta, C) = C\sqrt{2\beta + (\beta+2)/(\beta-2)}$.

The upper bound on maximal expected regret just obtained is increasing in the number of available treatments K. This is due to the fact that it becomes harder to find the best treatment as the number of available treatments increases. Note also that the choice $\beta = 2 + \sqrt{2}$ minimizes $c(\beta, C)$ and implies $c \leq \sqrt{11}C$.

The proof of Theorem 2.12 is inspired by the proof of Theorem 2.1 in Bubeck and Cesa-Bianchi (2012). However, in contrast to their argumentation, we cannot exploit the specific structure of the mean functional and related concentration inequalities. Instead we rely on the high-level condition of Assumption 2.2 and the Dvoretzky-Kiefer-Wolfowitz-Massart inequality as established by Massart (1990) to obtain suitable concentration inequalities, cf. Equation (106) in Appendix D. To use this reasoning, we also need to show that the empirical cdfs defined in (8) are based on i.i.d. random variables, which is done via the optional skipping theorem of Doob (1936).

The lower bound in Theorem 2.10 combined with the upper bound in Theorem 2.12 shows that the maximal expected regret incurred by *any* explore-then-commit policy grows much faster in n than that of the F-UCB policy. What is more, the F-UCB policy achieves this without making use of the horizon n. Thus, in particular when n is unknown, a large improvement is obtained over *any* explore-then-commit policy, as the order of the regret decreases from n to $\sqrt{n\log(n)}$. Thus, in terms of maximal expected regret, the policy maker is not recommended to separate the exploration and commitment phases.

Theorem 2.12 leaves open the possibility that one can construct policies with even slower growth rates of maximal expected regret. We now turn to establishing a lower bound on maximal expected regret within the class of all policies. In particular, this also includes policies that incorporate the horizon n.

Theorem 2.13. *Suppose $K = 2$ and that Assumption 2.9 holds. Then there exists a constant $c_l > 0$, such that for any policy π and any randomization measure, it holds that*

$$\sup_{\substack{F^i \in \{J_\tau : \tau \in [0,1]\} \\ i=1,2}} \mathbb{E}[R_n(\pi)] \geq c_l n^{1/2} \quad \textit{for every } n \in \mathbb{N}. \tag{11}$$

Under the same assumptions used to establish the lower bound on maximal expected regret in the class of explore-then-commit policies, Theorem 2.13 shows that *any* policy must incur maximal expected regret of order at least $n^{1/2}$. In combination with Theorem 2.12 this shows that, up to a multiplicative factor of $\sqrt{\log(n)}$, no policy exists that has a better dependence of maximal expected regret on n than the F-UCB policy. In this sense the F-UCB policy is near minimax (rate-) optimal.

2.4 Numerical illustrations

We now illustrate the theoretical results established so far by means of numerical examples. Throughout the section, the treatment outcome distributions F^i will be taken from the Beta family, a parametric subset of $D_{cdf}([0,1])$, which has a long history in modeling income distributions; see, for example, Thurow (1970), McDonald (1984) and McDonald and Ransom (2008). An appealing characteristic of the Beta family is its ability to replicate many "shapes" of distributions. We emphasize that the policies investigated do not exploit that the unknown treatment outcome distributions are elements of the Beta family.

Our numerical results cover different functionals T, with a focus on situations where the policy maker targets the distribution that maximizes welfare, and where we consider the case $a = 0$ and $b = 1$. In all our examples the feasible set for the marginal distributions of the treatment outcomes $\mathscr{D} = D_{cdf}([0,1])$.

The specific welfare measures we consider are as follows (and correspond to the Gini-, Schutz- and Atkinson- inequality measure, respectively, through the transformations detailed in Section 4.3, to which we refer the reader for more background information):

1. *Gini-index-based welfare measure*: $\mathsf{W}(F) = \mu(F) - \frac{1}{2}\int\int |x_1 - x_2| dF(x_1)dF(x_2)$, where $\mu(F) := \int x dF(x)$ denotes the mean of F.

   [Assumption 2.2 is satisfied with $\mathscr{D} = D_{cdf}([0,1])$ and $C = 2$, cf. the discussion after Lemma 4.9.]

2. *Schutz-coefficient-based welfare measure*: $\mathsf{W}(F) = \mu(F) - \frac{1}{2}\int |x - \mu(F)| dF(x)$.

   [Observing that $\mathsf{W}(F) = \mu(F) - \mathsf{S}_{abs}(F)$ with S_{abs} as defined in Equation (22), it follows from Lemmas 4.1 and 4.9 that Assumption 2.2 is satisfied with $\mathscr{D} = D_{cdf}([0,1])$ and $C = 2$.]

3. *Atkinson-index-based welfare measure*: $\mathsf{W}(F) = [\int x^{1-\varepsilon} dF(x)]^{1/(1-\varepsilon)}$ for a parameter $\varepsilon \in (0,1) \cup (1,\infty)$.

[Restricting attention to $\varepsilon \in (0,1)$, the mean value theorem along with Example D.4 in Appendix D yield that Assumption 2.2 is satisfied with $\mathscr{D} = D_{cdf}([0,1])$ and $C = \frac{1}{1-\varepsilon}$. We shall consider $\varepsilon \in \{0.1, 0.5\}$.]

2.4.1 Implementation details

We focus on scenarios where the total number of assignments to be made is not known from the outset. Thus, the policies we study do not make use of the horizon n. Throughout, we consider the case of $K = 2$ treatments. In the following, the symbol W shall denote one of the welfare measures just defined in the above enumeration.

We consider explore-then-commit policies as in Section 2.2, and the F-UCB policy from Section 2.3. A detailed description concerning the implementation of the policies investigated is as follows.

- **Explore-then-commit policies**: In all explore-then-commit policies we consider, Treatments 1 and 2 are assigned cyclically in the exploration period. This ensures that the number of assignments to each treatment differs at most by 1 (cf. also Remark 2.8 in Section 2.2).[4] Given this specification, the policy maker must still choose i) the length of the exploration period n_1, and ii) a commitment rule to be used after the exploration phase. The choice of n_1 depends on the commitment rule, of which we now develop a test-based and an empirical-success-based variant:

  1. ETC-T: This policy is built around a *test-based commitment rule*. That is, one uses a test for the testing problem "equal welfare of treatments," i.e., $\mathsf{W}(F^1) = \mathsf{W}(F^2)$, in deciding which treatment to choose after the exploration phase. Given a test that satisfies a pre-specified size requirement, the length of the exploration phase is chosen such that the power of the test against a certain deviation from the null (effect size) is at least of a desired magnitude. A typical desired amount of power against the deviation from the null of interest is 0.8 or 0.9. The deviation from the null that one wishes to detect is clearly context dependent. We refer to Jacob (1988), Murphy et al. (2014) and Athey and Imbens (2017), as well as references therein, for in-depth treatments of power calculations.

     To make this approach implementable, we need to construct an appropriate test. Given $\alpha \in (0,1)$, and for $n_1 \geq 2$, we shall consider the test that rejects if (and only if) $|\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(\hat{F}_{2,n_1})| \geq c_\alpha$ with $c_\alpha = \sqrt{2\log(4/\alpha)C^2/\lfloor n_1/2 \rfloor}$. Under the null, i.e., for every pair F^1 and F^2 in $D_{cdf}([0,1])$ such that $\mathsf{W}(F^1) = \mathsf{W}(F^2)$, this test has rejection probability at most α (a proof of this statement is provided in Appendix B.3.1). Hence, the size of this test does not exceed α.

     For this test, in order to detect a deviation of $\Delta := |\mathsf{W}(F^1) - \mathsf{W}(F^2)| > 0$ with probability at least $1-\eta$, where $\eta \in (0,1)$, it suffices that $n_1 = 2\lceil \frac{8\log(4/\min(\alpha,\eta))C^2}{\Delta^2} \rceil$ (for a proof of this statement, see Appendix B.3.2).

[4] Investigating policies with randomized assignment in the exploration phase would necessitate running the simulations repeatedly, averaging over different draws for the assignments in the exploration phase. The numerical results are already quite computationally intensive, which is why we only investigate a cyclical assignment scheme. This scheme already reflects to a good extent the average behaviour of a randomized assignment with equal assignment probabilities.

In our numerical studies we set $\eta = \alpha = 0.1$. We consider $\Delta \in \{0.15, 0.30\}$, which amounts to a small and moderate desired detectable effect size, respectively. Note that while choosing Δ small allows one to detect small differences in the functionals by the above test, this comes at the price of a larger n_1. Thus, we shall see that neither $\Delta = 0.15$ nor $\Delta = 0.30$ dominates the other uniformly (over $t \in \mathbb{N}$) in terms of maximal expected regret. The commitment rule applied is to assign $\arg\max_{1 \le i \le 2} \mathsf{W}(\hat{F}_{i,n_1})$ if the above test rejects, and to randomize the treatment assignment with equal probabilities otherwise. Finally, we sometimes make the dependence of ETC-T on Δ explicit by writing ETC-T(Δ).

2. ETC-ES: This policy assigns $\pi_n^c(Z_{n_1}) := \min \arg\max_{1 \le i \le K} \mathsf{W}(\hat{F}_{i,n_1})$ to subjects $t = n_1 + 1, \ldots, n$, which is an *empirical success commitment rule* inspired by Manski (2004) and Manski and Tetenov (2016). Here, given a $\delta > 0$, n_1 is chosen such that the maximal expected regret for every subject to be treated after the exploration phase is at most δ; i.e., n_1 satisfies

$$\sup_{\substack{F^i \in \mathscr{D} \\ i=1,\ldots,K}} \mathbb{E}\left(\max_{i \in \mathcal{I}} \mathsf{W}(F^i) - \mathsf{W}(F^{\pi_n^c(Z_{n_1})})\right) \le \delta.$$

We prove in Appendix B.3.3 that $n_1 = 2\lceil 16C^2/(\delta^2 \exp(1)) \rceil$ suffices.

In our numerical results, we consider $\delta \in \{0.15, 0.30\}$, which should be contrasted to the treatment outcomes taking values in $[0, 1]$. Note that the n_1 required to guarantee a maximal expected regret of at most δ for every subject treated *after* the exploration phase is decreasing in δ. Thus, we shall see that it need not be the case that choosing δ smaller will result in lower overall maximal expected regret. Finally, we sometimes make the dependence of ETC-ES on δ explicit by writing ETC-ES(δ).

- **F-UCB policy**: Implemented as described in Policy 2 in Section 2.3 with $\beta = 2.01$.

The following display summarizes the numerical implementation.

Input: $n = 100{,}000$, $r = 20$ and
$\mathcal{G} = \{0.1, 0.75, 0.85, 0.95, 0.975, 1, 1.025, 1.05, 1.15, 1.25, 5\}$.
for $p_1 \in \mathcal{G}$ *such that* $p_1 < 5$ **do**
  for $p_2 \in \mathcal{G}$, $p_2 > p_1$ **do**
    for $l = 1, \ldots, r$ **do**
      Generate n independent observations from $\mathsf{Beta}(1, p_1) \otimes \mathsf{Beta}(1, p_2)$.
      for $t = 1, \ldots, n$ **do**
        Calculate the regret of each policy over all assignments $s = 1, \ldots, t$.
      end
    end
    Estimate expected regret for each policy and for $t = 1, \ldots, n$ by the arithmetic mean of regret over the r data sets.
  end
end
Estimate, for every $t = 1, \ldots, n$, the maximal expected regret by maximizing the arithmetic means over the $|\mathcal{G}|(|\mathcal{G}| - 1)/2 = 55$ parameter vectors (p_1, p_2).

Since maximizing expected regret over all Beta distributions would be numerically infeasible, we have chosen to maximize expected regret over a subset of all Beta distributions indexed by $\mathcal{G}$ as defined in the previous display. We stress that since none of the three policies above needs to know n, the numerical results also contain the maximal expected regret of the policies for any sample size less than $n = 100{,}000$.

2.4.2 Results

The left panel of Figure 1 illustrates the maximal expected regret for the F-UCB, ETC-T and ETC-ES policies in the case of Gini-welfare. Each point on the five graphs is the maximum of expected regret over the 55 different distributions considered. In accordance with Theorems 2.10 and 2.12, the maximal expected regret of the policies in the explore-then-commit family is generally higher than the one of the F-UCB policy. For $t = 100{,}000$, the maximal expected regret of F-UCB is 498 while the corresponding numbers for ETC-T(0.15), ETC-ES(0.15), ETC-T(0.30) and ETC-ES(0.30) are 3,896, 777, $6,424$ and 778, respectively. Note also that no matter the values of Δ and δ, the maximal expected regret of ETC-ES(δ) is much lower than the one of the ETC-T(Δ) policy.[5] In fact, we shall see for all functionals considered that the F-UCB policy generally incurs the lowest maximal expected regret followed by ETC-ES policies, which in turn perform much better than ETC-T policies.

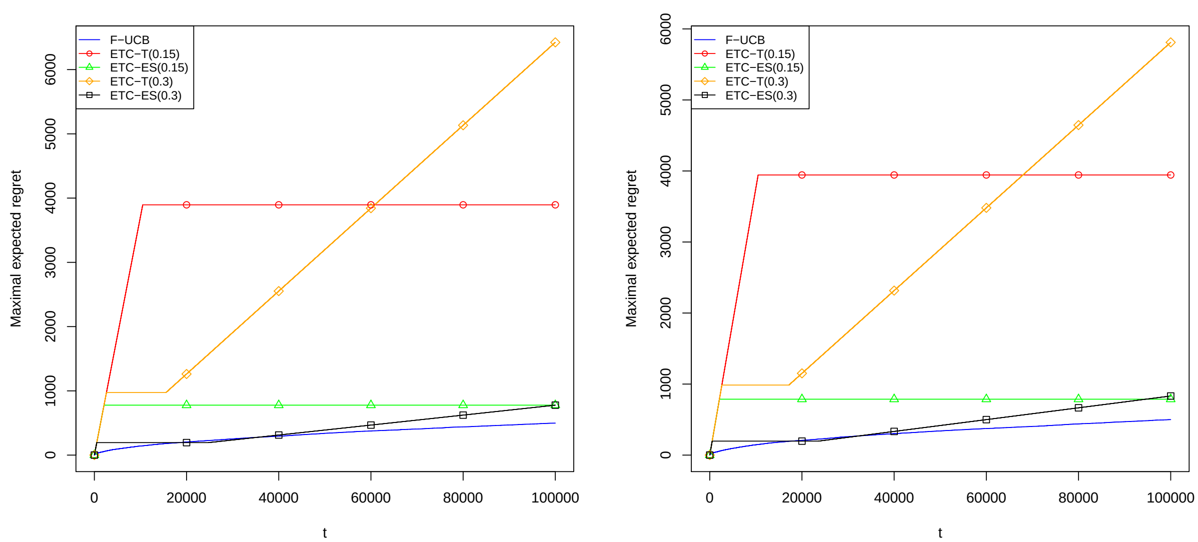

Figure 1: The figure contains the maximal expected regret for F-UCB, ETC-T(Δ) with $\Delta \in \{0.15, 0.30\}$ and ETC-ES(δ) with $\delta \in \{0.15, 0.30\}$. The left panel is for Gini-welfare while the right panel is for Schutz-welfare.

The shape of the graphs of the maximal expected regret of the explore-then-commit policies can be explained as follows: in the exploration phase maximal expected regret is attained by

[5]This result on the ranking of test-based vs. empirical success-based commitment rules is similar to an analogous finding in a non-sequential setting in Manski and Tetenov (2016).

a P_1, say, for which the value of the Gini-welfare differs strongly at the marginals. However, such distributions are also relatively easy to distinguish, such that none of the commitment rules (testing or empirical success) assigns the suboptimal treatment after the exploration phase. This results in no more regret being incurred and thus a horizontal part on the maximal expected regret graph. For t sufficiently large, however, maximal expected regret will be attained by a distribution P_2, say, for which the marginals are sufficiently "close" to imply that the commitment rules occasionally assign the suboptimal treatment. For such a distribution, the expected regret curve will have a positive linear increase even after the commitment time n_1 and this curve will eventually cross the horizontal part of the expected regret curve pertaining to P_1. This implies that maximal expected regret increases again (as seen for ETC-T(0.30) around t = 18,000 and ETC-ES(0.30) around t = 23,000 in the left panel of Figure 1). Eventually, such a kink also occurs for ETC-T(0.15) and ETC-ES(0.15). Thus, the left panel of Figure 1 illustrates the tension between choosing n_1 small in order to avoid incurring high regret in the exploration phase and, on the other hand, choosing n_1 large in order to ensure making the correct decision at the commitment time.

The right panel of Figure 1, which contains the maximal expected regret for the Schutz-welfare, yields results qualitatively similar to the ones for the Gini-welfare. The best explore-then-commit policy again has a terminal maximal expected regret that is more than 50% higher than that of the F-UCB policy.

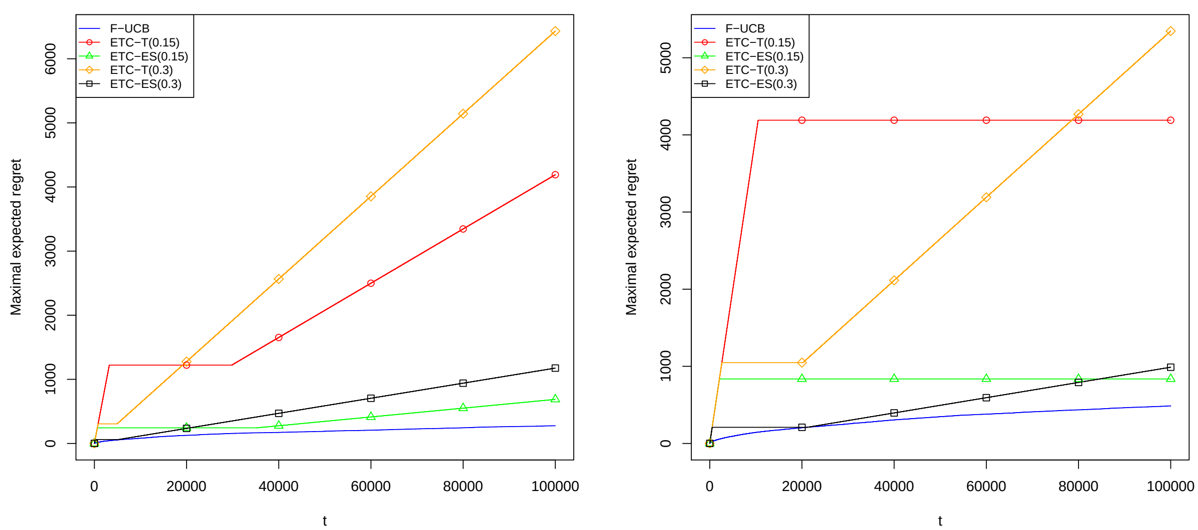

Figure 2: The figure contains the maximal expected regret for F-UCB, ETC-T(Δ) with $\Delta \in \{0.15, 0.30\}$ and ETC-ES(δ) with $\delta \in \{0.15, 0.30\}$ in the case of Atkinson welfare. The left panel is for $\varepsilon = 0.1$, while the right panel is for $\varepsilon = 0.5$.

We next turn to the two welfare measures in the Atkinson family. The left panel of Figure 2 contains the results for the case of $\varepsilon = 0.1$. While F-UCB incurs the lowest maximal expected regret uniformly over $t = 1, \ldots, 100{,}000$, the most remarkable feature of the figure is that maximal expected regret of all *all* explore-then-commit policies is eventually increasing within the sample considered. The reason for this is that $\varepsilon = 0.1$ implies a low value of n_1 such that i) the steep increase in maximal expected regret becomes shorter and ii) more mistakes are made at the commitment time. The ranking of the families of polices is unaltered with F-UCB dominating ETC-ES, which in turns incurs much lower regret than ETC-T.

The right panel of Figure 2 considers the case of Atkinson welfare when $\varepsilon = 0.5$. The findings are qualitatively similar to the ones for the Gini- and Schutz-based welfare measures.

3 Functional sequential treatment allocation with covariates

The results in the previous section are important in highlighting the main theoretical differences between explore-then-commit policies and the F-UCB policy. They are obtained in a simple setting, which does not include covariates. In a situation where covariate information is present, however, it seems reasonable to incorporate this into the policy. For example, a medical treatment may work well for one person, while it may be dangerous to a person who, e.g., is allergic to some of its substances. Similarly, an unemployment program that works well for individuals with low education may not be effective for highly educated individuals.

The goal of the present section is to study how a policy maker can optimally assign subjects to treatments in the presence of covariates. To this end, we now introduce an observational structure that incorporates covariates, generalize the F-UCB policy to this setting, and study its optimality properties. Since we already know that in terms of maximal expected regret it is not recommendable to separate exploration and commitment phases, we do not further consider explore-then-commit policies.

3.1 The setup and two impossibility results

Formally, the setup is similar to the one described in Section 2.1. However, we now suppose that prior to assigning subject t to a treatment, the policy maker observes the realization of a random vector X_t of covariates. For simplicity we assume throughout that $X_t \in [0,1]^d$. As in Section 2.1 the vector of potential outcomes is denoted as Y_t. *Throughout Section 3, 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). We denote the distribution of (Y_t, X_t) as $\mathbb{P}_{Y,X}$, and by $\mathbb{P}_X$ the marginal distribution of X_t. The conditional 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 (as defined in, e.g., Liese and Miescke (2008) Definition A.36) of $Y_{i,t}$ given X_t. We shall often impose the following condition (cf. also Remark 3.10).

Assumption 3.1. *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.*

Similarly as in Section 2.1, a policy π is a triangular array $\{\pi_{n,t} : n \in \mathbb{N}, 1 \le t \le n\}$. However, now the assignment $\pi_{n,t}$ takes as input the covariates X_t, previously observed outcomes and covariates (i.e., the complete observational history), and randomization. We therefore have

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

Given a policy π and $n \in \mathbb{N}$, the input to $\pi_{n,t}$ is denoted as (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 $Z_1 := (Y_{\pi_{n,1}(X_1,Z_0,G_1),1}, X_1)$, and of G_2. For $t \geq 3$ we have

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

The $(t-1)(d+1)$-dimensional random vector Z_{t-1} can be interpreted as the information available after the $(t-1)$-th treatment outcome has been observed and before the t-th subject arrives. We use similar notational simplifications as the ones mentioned in Section 2.1. Remark 2.1 also applies in the present context.

The policy maker observes the covariates before making the assignments. Therefore, treatments will still be evaluated according to a functional as in Equation (2), but now applied to cdfs conditional on the covariates: 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. Correspondingly, in the presence of covariates, the regret of a policy π is defined as

$$\begin{aligned} R_n(\pi) &= R_n(\pi; F^1, \ldots, F^K, X_n, Z_{n-1}, G_1, \ldots, G_n) \\ &= \sum_{t=1}^{n} \left[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot, X_t)\big) - \mathsf{T}\big(F^{\pi_{n,t}(X_t,Z_{t-1},G_t)}(\cdot, X_t)\big)\right]. \end{aligned}$$

In this section, we continue considering a policy maker who seeks a policy π that minimizes the maximal expected regret. But now the maximum will be taken over sets of possible joint distributions $\mathbb{P}_{Y,X}$. Similarly to Section 2, we denote the ("parameter"-) space of all potential conditional cdfs $F^i(\cdot, x)$ by $\mathscr{D}$. More precisely, we assume that

$$\{F^i(\cdot, x) : i = 1, \ldots, K \text{ and } x \in [0,1]^d\} \subseteq \mathscr{D}, \tag{12}$$

where $\mathscr{D}$ is typically a large and nonparametric subset of $D_{cdf}([a,b])$; cf. Section 4.1 for specific examples. When establishing lower bounds on maximal expected regret we shall often impose the following condition, which is slightly stronger than Assumption 2.9, but still very weak. Similarly to Assumption 2.9, it guarantees that there is a minimal amount of variation in the functional over a small subset of $\mathscr{D}$.

Assumption 3.2. *The functional $\mathsf{T} : D_{cdf}([a,b]) \to \mathbb{R}$ satisfies Assumption 2.2, and $\mathscr{D}$ contains two elements H_1 and H_2, such that*

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

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

$$\mathsf{T}(J_{\tau_2}) - \mathsf{T}(J_{\tau_1}) \geq c_-(\tau_2 - \tau_1) \quad \textit{for every } \tau_1 \leq \tau_2 \textit{ in } [0,1]. \tag{13}$$

We emphasize that Equation (13) in Assumption 3.2 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 in the sense that

$$\{F^i(y, \cdot) : i = 1, \dots, K \text{ and } y \in \mathbb{R}\} \quad \text{is uniformly equicontinuous.}^6 \tag{14}$$

The theorem is as follows.

Theorem 3.3. *Suppose $K = 2$ and that Assumption 3.2 is satisfied. Then there exists a constant $c_l > 0$, such that for every policy π and any randomization measure, we have*

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

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 (12) *and* (14)*, and where $\mathbb{P}_X$ is the uniform distribution on $[0, 1]^d$.*

Since Assumption 2.2 (which is a part of Assumption 3.2) implies that T is bounded, Theorem 3.3 shows that every policy incurs the worst case linear maximal expected regret. Therefore, further assumptions beyond the uniform equicontinuity condition in Equation (14) are needed. We shall from now on impose a Hölder equicontinuity condition on $F^i(\cdot, x)$. 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.

Assumption 3.4. *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 further considering policies that incorporate covariate information, one may wonder whether one could not just use the F-UCB policy, i.e., Policy 2 as introduced in Section 2; or another policy that *ignores covariates*, in the sense that it is a policy as defined in Equation (1). Our next result shows that any policy that ignores covariates incurs linear expected regret (unless all covariates happen to be irrelevant).

Theorem 3.5. *Let $K = 2$, suppose $\mathsf{T} : D_{cdf}([a, b]) \to \mathbb{R}$ satisfies Assumption 2.2, and let $\mathbb{P}_{Y,X}$ satisfy Assumption 3.4. Define the sets*

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

[6] The assumption in Equation (14) 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 $y \in \mathbb{R}$.

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

$$\mathbb{E}[R_n(\pi)] \geq c_l \min(\mathbb{P}_X(A_1), \mathbb{P}_X(A_2))n \quad \textit{for every } n \in \mathbb{N}. \tag{15}$$

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

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

Note that the lower bound in the previous theorem is even valid pointwise, as it makes a statement about any fixed distribution $\mathbb{P}_{Y,X}$.

3.2 The F-UCB policy in the presence of covariates

We now introduce a version of the F-UCB policy that incorporates covariate information. The idea is to group subjects with similar values of the covariates, and to run the F-UCB policy without covariates $\hat{\pi}$ as defined in Policy 2 of Section 2 for each group separately. Here, 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 $B_{n,i}$ is a non-empty Borel set. Given a partition, the F-UCB policy with covariates is defined as follows.

Policy 3: F-UCB policy with covariates $\bar{\pi}$

Inputs: $\beta > 2$, Partition $B_{n,1}, \ldots, B_{n,M(n)}$
Set: $N_j = 1$ and W^j the empty vector for $j = 1, \ldots, M(n)$
for $t = 1, \ldots, n$ **do**
  for $j = 1, \ldots, M(n)$ **do**
    if $X_t \in B_{n,j}$ **then**
      assign $\hat{\pi}_{N_j}(W^j)$
      $W^j \leftarrow (Y_{\hat{\pi}_{N_j}(W^j),t}, W^j)$
      $N_j \leftarrow N_j + 1$
    end
  end
end

Note that one could use this partitioning method to extend any policy as defined in Equation (1) that does not depend on the horizon to incorporate covariates through partitioning. We focus on studying the F-UCB policy, however, because it has near-optimal maximal expected regret properties in the no-covariates case, cf. Section 2.3.

Partitioning effectively amounts to targeting the treatment that is best "on average" in each group, instead of fully individualizing the treatments. On $B_{n,j}$ with $\mathbb{P}_X(B_{n,j}) > 0$ this means that the 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.,

$$F^i_{n,j}(y) := \frac{1}{\mathbb{P}_X(B_{n,j})} \int_{B_{n,j}} F^i(y,x) d\mathbb{P}_X(x). \tag{16}$$

In general $\arg\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j}) \neq \arg\max_{i\in\mathcal{I}} \mathsf{T}(F^i(\cdot,x))$. Hence, targeting $\max_{i\in\mathcal{I}} \mathsf{T}(F^i_{n,j})$ 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.

3.3 An upper bound on the maximal expected regret of the F-UCB policy with covariates

The following theorem gives an upper bound on the maximal expected regret of the F-UCB policy in the presence of covariates, and for *any* choice of partition. This flexibility may be useful since the policy maker is often constrained in the way groups can be formed. The result quantifies how the partitioning affects the regret guarantees.

Theorem 3.6. *Suppose Assumptions 2.2 and 2.6 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 $c = c(\beta, C)$ as in Theorem 2.12, it holds that*

$$\sup \mathbb{E}[R_n(\bar{\pi})] \leq \sum_{j=1}^{M(n)} \left[c\sqrt{Kn\mathbb{P}_X(B_{n,j})\overline{\log}(n\mathbb{P}_X(B_{n,j}))} + 2CLV^{\gamma}_{n,j}n\mathbb{P}_X(B_{n,j}) \right] \text{ for every } n\in\mathbb{N}, \tag{17}$$

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 (12)*, and Assumption 3.4 with L and γ.*[7]

Each of the summands $j = 1,\ldots,M(n)$ in the upper bound on the maximal expected regret in Equation (17) consists of two parts: The first part is very similar to the upper bound of Theorem 2.12. 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 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 $B_{n,j}$, 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 (16). The second part in each summand in the upper bound in (17) 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 $x \in B_{n,j}$.

An important 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

$$\left\{ x \in [0,1]^d : \frac{k_l - 1}{P} \leq x_l \preceq \frac{k_l}{P},\ l = 1,\ldots,d \right\}, \tag{18}$$

where $\preceq$ is to be interpreted as $\leq$ for $k_l = P$, and as $<$ otherwise. This defines a partition of $[0,1]^d$ into P^d hypercubes with side length $1/P$ each. We now order these hypercubes

[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 $\mathbb{P}_X(B_{n,j}) = 0$.

lexicographically according to their index vector k, to obtain the corresponding *cubic partition* $B_1^P, \ldots, B_{P^d}^P$. The following results specializes Theorem 3.6 to this specific partition and for a choice of P that will be shown to be optimal below.

Corollary 3.7. *Suppose Assumptions 2.2 and 2.6 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 (18), 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*

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

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 (12), Assumption 3.1 with $\bar{c}$ (and any $\underline{c}$), and Assumption 3.4 with L and γ.

Corollary 3.7 reveals that it is possible to achieve sublinear (in n) maximal expected regret under the Hölder equicontinuity condition imposed through Assumption 3.4. This is interesting also in light of Theorem 3.3, 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 3.4. Note that a "curse of dimensionality" is present, in the sense that the upper bound in Theorem 3.7 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 3.7 results in a near-minimax optimal policy, as we show in the following theorem.

Theorem 3.8. *Suppose $K = 2$ and that Assumption 3.2 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 (12), Assumption 3.4 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 3.8 to the upper bound on maximal expected regret established in Corollary 3.7, 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 3.9 (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 3.7, can be used in practice only if one knows n, i.e., the policy is not anytime. If, n is unknown, however, one can instead 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 3.7, but with higher multiplicative constants. In essence, the doubling trick works by "restarting" the policy at times 2^m, $m \in \mathbb{N}$. The doubling trick is a standard tool in games of unknown horizon and we refer to Shalev-Shwartz (2012) and the recent work of Besson and Kaufmann (2018) for more details.

Remark 3.10 (Discrete covariates)**.** We mostly focus on the case of continuous covariates (although this is not formally required in Theorem 3.6). A natural, and also minimax rate-optimal, solution to incorporate discrete covariates would be to fully condition on these, i.e., to apply the F-UCB policy separately for each combination of discrete covariates. In this article, we omit formal statements concerning discrete covariates, but we emphasize that corresponding results can be obtained from the results provided by conditioning arguments.

3.4 Stronger regret guarantees and number of suboptimal assignments

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 does not restrict the (unknown) similarity of the best and second best treatment. Therefore, the results so far do not convey information about whether the F-UCB policy optimally incorporates this degree of similarity. It is clear that identifying the best treatment becomes easier, as the difference between the best and the remaining treatments gets more pronounced. 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 its own right, the results in the present section are instrumental to proving our impossibility result Theorem 3.3 and to establishing the regret lower bound in Theorem 3.8. Additionally, the "well-separateness condition" imposed will also allow us to bound the expected number of suboptimal assignments of the F-UCB policy.

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 3.11. *There exists an $\alpha \in (0,1)$ and a $C_0 > 0$, such that*[8]

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

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 Perchet and Rigollet (2013) in the context of a multi-armed bandit problem targeting the mean.

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

Theorem 3.12. *Suppose Assumptions 2.2 and 2.6 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 (18), 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*

$$\sup \mathbb{E}\left[R_n(\bar{\pi})\right] \le cK\overline{\log}(n)n^{1-\frac{\gamma(1+\alpha)}{2\gamma+d}} \qquad \text{for every } n \in \mathbb{N}, \tag{19}$$

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 (12), Assumption 3.1 with $\underline{c}$ and $\bar{c}$, Assumption 3.4 with L and γ, and Assumption 3.11 with $\alpha \in (0,1)$ and $C_0 > 0$.

Compared to Corollary 3.7 the exponent on n in the upper bound on regret is smaller, the difference depending on α. Thus, in the presence of Assumption 3.11, the regret guarantee of the F-UCB policy is stronger, even without incorporating α into the policy. We shall see in Theorem 3.14 below that the upper bound on maximal regret in Theorem 3.12 is optimal in n up to logarithmic factors.

The margin condition does not only allow us to prove a lower upper bounds on maximal expected regret than in Corollary 3.7. It also allows us to prove an upper bound on the number of suboptimal assignments made by the F-UCB policy. We shall define the total number of suboptimal assignments for a policy π over the course of a total of n assignments as

$$\begin{aligned} S_n(\pi) &= S_n(\pi; F^1,\ldots,F^K, X_n, Z_{n-1}, G_1,\ldots,G_n) \\ &= \sum_{t=1}^n \mathbb{1}\left\{\pi_{n,t}(X_t, Z_{t-1}, G_t) \notin \arg\max\left\{\mathsf{T}(F^i(\cdot,X_t)) : i = 1,\ldots,K\right\}\right\}. \end{aligned}$$

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.

[8] We note that the events in the displayed equation of Assumption 3.11 are not necessarily Borel measurable. Therefore, Assumption 3.11 implicitly imposes measurability on all events considered. Note, however, that in case Assumptions 2.2 and 3.4 as well as the inclusion in Equation (12) are assumed, this measurability condition is easily seen to be satisfied.

Theorem 3.13. *Suppose Assumptions 2.2 and 2.6 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* (18)*, 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*

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

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 (12)*, Assumption 3.1 with $\underline{c}$ and $\bar{c}$, Assumption 3.4 with L and γ, and Assumption 3.11 with $\alpha \in (0,1)$ and $C_0 > 0$.*

The upper bound in Theorem 3.13 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 3.12 and 3.13 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 3.14. *Suppose $K = 2$ and that Assumption 3.2 is satisfied. Let $\gamma \in (0,1]$. Then for every policy π and any randomization measure, we have*

$$\sup \mathbb{E}[R_n(\pi)] \geq n^{1-\frac{\gamma(1+\alpha)}{2\gamma+d}} \Big/ \left[64^{1+1/\alpha}(C_0+1)^{1/\alpha}\right] \quad \text{for every } n \in \mathbb{N}, \tag{20}$$

and

$$\sup \mathbb{E}[S_n(\pi)] \geq n^{1-\frac{\alpha\gamma}{d+2\gamma}}/32 \quad \text{for every } n \in \mathbb{N}, \tag{21}$$

where both suprema are taken over all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$ for $t = 1, \ldots, n$, where $\mathbb{P}_{Y,X}$ satisfies Equation (12)*, Assumption 3.4 with parameters γ and $L = 17^{-1/2}$, Assumption 3.11 with $\alpha \in (0,1)$ and $C_0 = 8d(c\_2L)^{-\alpha}$, and where $\mathbb{P}_X$ is the uniform distribution on $[0,1]^d$.*

The statement in Equation (20) in Theorem 3.14 shows that the F-UCB policy is near minimax optimal in terms of maximal expected regret. Similarly, the lower bound in Equation (21) of Theorem 3.14 together with Theorem 3.13 proves that the F-UCB policy assigns the minimal number of suboptimal treatments.

4 Assumption 2.2 for inequality, welfare, and poverty measures

To illustrate the scope of our results, and to facilitate their implementation in practice, we shall now discuss several functionals of interest in applied economics that satisfy Assumption 2.2. We also provide a corresponding set $\mathscr{D}$ and a constant C. Appendix D contains a toolbox of *general* methods for verifying Assumption 2.2. The results in the present section are established using these techniques. Therefore, in addition to their intrinsic importance, the following results, and in particular their proofs, also provide a pattern as to how Assumption 2.2 can be verified for functionals that we do not explicitly discuss. Before we proceed to these results, we introduce some nonparametric classes of cdfs $\mathscr{D}$ which will play a major role. The proofs of all results discussed in the present section can be found in Appendix C.

4.1 Important classes of cdfs $\mathscr{D}$

Recall that $a < b$ are throughout assumed to be real numbers. We shall consider the following classes of cdfs.

1. $\mathscr{D}^s([a,b])$: The subset of all $F \in D_{cdf}([a,b])$ that are continuous when restricted to $[a,b]$, and are right-differentiable on (a,b), with right-sided derivative F^+, say, satisfying $F^+(x) \leq s$ for all $x \in (a,b)$.

2. $\mathscr{D}_r([a,b])$: The subset of all $F \in D_{cdf}([a,b])$ that are continuous when restricted to $[a,b]$, and right-differentiable on (a,b), with $F^+(x) \geq r$ for all $x \in (a,b)$.

3. $\mathscr{D}_r^s([a,b]) := \mathscr{D}^s([a,b]) \cap \mathscr{D}_r([a,b])$.

4. Furthermore, the subset of all $F \in \mathscr{D}^s([a,b])$ that are (everywhere) continuous shall be denoted by $\mathscr{C}^s([a,b])$, and we correspondingly define $\mathscr{C}_r([a,b])$ and $\mathscr{C}_r^s([a,b])$.

Note that if $F \in D_{cdf}([a,b])$ is differentiable with a density $f = F'$ that, on $[a,b]$, is bounded from below by r and from above by s, then $F \in \mathscr{C}_r^s([a,b])$. The set $\mathscr{C}_r^s([a,b])$ is contained in all classes of cdfs defined in 1.-4. above. Hence, one can think of the (strongest) assumptions imposed above as putting a lower or an upper bound on the unknown densities of the (conditional) outcome distributions.

4.2 Inequality measures

In this section we verify Assumption 2.2 for functionals that aim to measure the degree of inequality inherent to an (e.g., income, wealth or productivity) distribution F. Such *inequality measures* are relevant in situations where one intends to select that treatment (e.g., one out of several possible taxation schemes) which leads to the most "equal" outcome distribution. To avoid possible misunderstandings, we emphasize that it is neither our goal to discuss theoretical foundations of inequality measures, nor to point out their relative advantages and disadvantages. The functional must be chosen by the applied economist, who can—in making such a choice—rely on excellent book-length treatments, e.g., Lambert (2001), Chakravarty (2009) or Cowell (2011), as well as the original sources, some of which we shall point out further below. Rather, our goal is to demonstrate that Assumption 2.2 is satisfied for a broad range of practically relevant functionals. We also emphasize that the inequality measures discussed in the present section are important building blocks in constructing welfare measures, which we will be the topic of discussion in Section 4.3.

We first discuss inequality measures that derive from the Lorenz curve (cf. Gastwirth (1971) or Equation (26) below for a formal definition). The first such inequality measure we consider is the *Schutz-coefficient* $\mathsf{S}_{\mathrm{rel}}$ (cf. Schutz (1951), Rosenbluth (1951)), say, which is also known as the *Hoover-index* or the *Robin Hood-index*. Formally,

$$\mathsf{S}_{\mathrm{rel}}(F) = \frac{1}{2\mu(F)} \int |x - \mu(F)| dF(x),$$

provided the mean $\mu(F) := \int x dF(x)$ exists and is nonzero. The subindex "rel" in $\mathsf{S}_{\mathrm{rel}}(F)$ signifies that this index is defined "relative" to the mean. Note that, as a consequence, if one *multiplies* each income by the same (positive) amount this does not result in a change of the inequality index, i.e., the index is *scale independent*. A corresponding "absolute" variant, i.e., a measure which remains unchanged if one *adds* to every income the same amount, is obtained by multiplying the relative measure $\mathsf{S}_{\mathrm{rel}}$ by the mean functional, and is denoted by

$$\mathsf{S}_{\mathrm{abs}}(F) = \frac{1}{2} \int |x - \mu(F)| dF(x). \tag{22}$$

For a discussion of relative and absolute inequality measures we refer to Kolm (1976a,b), who calls them "rightist" and "leftist," respectively. As a general rule, absolute inequality indices require less restrictive assumptions on $\mathscr{D}$ than their relative counterparts in order to satisfy Assumption 2.2. This is due to the fact that division by $\mu(F)$ is highly instable for small values of $\mu(F)$. The following lemma provides conditions under which the relative and absolute Schutz-coefficient satisfy Assumption 2.2.

Lemma 4.1. *Let $a < b$ be real numbers. Then the absolute Schutz-coefficient $\mathsf{T} = \mathsf{S}_{abs}$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a, b])$ and $C = b - a$. Next, assume that $a \geq 0$, and define for every $\delta \in (a, b)$ and every $s > 0$ the set*

$$\mathscr{D}(s, \delta) := \{F \in \mathscr{C}^s([a, b]) : \mu(F) \geq \delta\}.$$

*Then, for every $\delta \in (a, b)$ and every $s > 0$, the relative Schutz-coefficient $\mathsf{T} = \mathsf{S}_{rel}$ (*defined *as 0 for the cdf corresponding to point mass 1 at 0) satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}(s, \delta)$ and $C = (b - a)(2s + \delta^{-1}) + 5$.*

The next inequality measure we consider is the *Gini-index*. Formally, its relative variant is defined as

$$\mathsf{G}_{\mathrm{rel}}(F) = \frac{1}{2\mu(F)} \int \int |x_1 - x_2| dF(x_1) dF(x_2), \tag{23}$$

provided that the expression is well defined. A corresponding absolute inequality measure is

$$\mathsf{G}_{\mathrm{abs}}(F) = \frac{1}{2} \int \int |x_1 - x_2| dF(x_1) dF(x_2). \tag{24}$$

The following lemma provides conditions under which Assumption 2.2 is satisfied for these two Gini-indices.

Lemma 4.2. *Let $a < b$ be real numbers. Then the absolute Gini-index $\mathsf{T} = \mathsf{G}_{abs}$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a, b])$ and $C = b - a$. Next, assume that $a \geq 0$, and define for every $\delta \in (a, b)$ the set*

$$\mathscr{D}(\delta) := \{F \in D_{cdf}([a, b]) : \mu(F) \geq \delta\}. \tag{25}$$

Then, for every $\delta \in (a, b)$, the relative Gini-index $\mathsf{T} = \mathsf{G}_{rel}$ (defined as 0 for the cdf corresponding to point mass 1 at 0) satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}(\delta)$ and $C = 2\delta^{-1}(b - a)$.

The Gini-index belongs to the class of *linear inequality measures* introduced by Mehran (1976) (cf. in particular Equation 3 there). An inequality measure is called *linear*, if it is a functional of the form

$$F \mapsto \int_{[0,1]} (u - L(F, u)) dW(u), \quad \text{where} \quad L(F, u) := \mu(F)^{-1} \int_{[0,u]} q_\alpha(F) d\alpha, \tag{26}$$

and where W is a function on $[0,1]$ that is fixed (i.e., independent of F) with finite total variation. Here $q_\alpha(F) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$ is the usual α-quantile of the cdf F, and $L(F;u)$ is the Lorenz curve corresponding to F evaluated at u (cf. also the discussion around our Equation (114)). The following lemma provides conditions under which a linear inequality measure satisfies Assumption 2.2. The result relies on properties of the Lorenz curve established in Lemma D.14 in Appendix D. The class of linear inequality measures is large, and the lemma thus applies quite generally. However, the generality is bought at the price of adding further regularity conditions on $\mathscr{D}$; in particular $a > 0$ has to be assumed. This trade-off in generality and strength of assumptions becomes apparent by comparing the regularity conditions to the ones in Lemma 4.2. Nevertheless, the result shows that Assumption 2.2 can be expected to be quite generically satisfied.

Lemma 4.3. *Let $a < b$ be positive real numbers and let $r > 0$. Assume that $W : [0,1] \to \mathbb{R}$ has finite total variation κ, say. Then the functional defined in Equation* (26) *satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{C}_r([a,b])$ and $C = \kappa a^{-1}(r^{-1} + (b-a)a^{-1}b)$.*

An absolute version of the linear inequality measure in Equation (26) can be obtained through multiplication by $\mu(F)$, i.e.,

$$F \mapsto \int_{[0,1]} (\mu(F)u - Q(F,u))dW(u), \qquad \text{where} \qquad Q(F,u) := \int_{[0,u]} q_\alpha(F)d\alpha. \tag{27}$$

The following result provides conditions under which such absolute linear inequality measures satisfy Assumption 2.2. As usual, the regularity conditions on $\mathscr{D}$ required are weaker than the ones needed for the relative version. In particular $a > 0$ does *not* need to be assumed.

Lemma 4.4. *Let $a < b$ be real numbers and let $r > 0$. Assume that $W : [0,1] \to \mathbb{R}$ has finite total variation κ, say. Furthermore, denote $|\int_{[0,1]} udW(u)| =: c$. Then the functional defined in Equation* (27) *satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{C}_r([a,b])$ and $C = c(b-a) + r^{-1}\kappa$.*

Another important family of scale-independent inequality measures is the so-called *generalized entropy family*, cf. Cowell (1980): Given a parameter $c \in \mathbb{R}$, an inequality measure is obtained via (if the involved expressions are well defined)

$$\mathsf{E}_c(F) = \begin{cases} \frac{1}{c(c-1)} \int \left[\left(x/\mu(F)\right)^c - 1 \right] dF(x) & \text{if } c \notin \{0,1\} \\ \int \left(x/\mu(F)\right) \log\left(x/\mu(F)\right) dF(x) & \text{if } c = 1 \\ \int \log\left(\mu(F)/x\right) dF(x) & \text{if } c = 0. \end{cases}$$

The inequality measure corresponding to $c = 1$ is known as Theil's entropy index (cf. also Theil (1967)), and the measure corresponding to $c = 0$ is the mean logarithmic deviation (cf. Lambert (2001), p.112). A formal result providing conditions under which a generalized entropy measure satisfies Assumption 2.2 is presented next. The regularity conditions we need to impose depend on c. Note in particular that support assumptions implicit in $\mathscr{D}$ are somewhat weaker for $c \in (0,1)$.

Lemma 4.5. *Let $0 \leq a < b$ be real numbers, and let $c \in \mathbb{R}$.*

1. *If $c \in (0,1)$, then, for every $\delta \in (a,b)$, the functional $\mathsf{T} = \mathsf{E}_c$ (*defined *as* 0 *for the cdf corresponding to point mass* 1 *at* 0*) satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}(\delta)$ (cf. Equation* (25)*) and $C = |c(c-1)|^{-1}\left[\delta^{-c}(b^c - a^c) + \delta^{-1}(b-a)\right]$.*

2. *If $c \notin [0,1]$ and $a > 0$, then the functional $\mathsf{T} = \mathsf{E}_c$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$ and $C = |c(c-1)|^{-1}[a^{-c}|b^c - a^c| + |c| \max\left((a/b)^{2c-1}, (b/a)^{2c-1}\right) a^{-1}(b-a)]$.*

3. *If $c \in \{0,1\}$ and $a > 0$, then the functional $\mathsf{T} = \mathsf{E}_c$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$ and $C = (b-a)/a + \log(b/a)$ if $c = 0$, and with $C = \int_{[a/b,b/a]} |1 + \log(x)|dx + \frac{b(b-a)}{a^2}\left\{\log(b/a) + 1\right\}$ if $c = 1$.*

We continue with a family of relative inequality indices introduced by Atkinson (1970). This family depends on an "inequality aversion" parameter $\varepsilon \in (0,1) \cup (1,\infty)$. For a fixed ε in that range, the index obtained equals (if the involved quantities are well defined)

$$\mathsf{A}_\varepsilon(F) = 1 - \frac{1}{\mu(F)}\left[\int x^{1-\varepsilon} dF(x)\right]^{1/(1-\varepsilon)}.$$

It is well known (cf., e.g., Lambert (2001) p.112) that A_ε can be written as

$$\mathsf{A}_\varepsilon(F) = 1 - [\varepsilon(\varepsilon - 1)\mathsf{E}_{1-\varepsilon}(F) + 1]^{1/(1-\varepsilon)}. \tag{28}$$

Together with Lemma 4.5, this relation can be used to obtain the following result:

Lemma 4.6. *Let $0 \le a < b$ be real numbers, let $\varepsilon \in (0,1) \cup (1,\infty)$ and set $c(\varepsilon) = 1 - \varepsilon$.*

1. *If $\varepsilon \in (0,1)$, then, for every $\delta \in (a,b)$, the functional $\mathsf{T} = \mathsf{A}_\varepsilon$ (*defined *as* 0 *for the cdf corresponding to point mass* 1 *at* 0*) satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}(\delta)$ (cf. Equation* (25)*) and $C = c(\varepsilon)^{-1}\left[\delta^{-c(\varepsilon)}(b^{c(\varepsilon)} - a^{c(\varepsilon)}) + \delta^{-1}(b-a)\right]$.*

2. *If $\varepsilon \in (1,\infty)$ and $a > 0$, then the functional $\mathsf{T} = \mathsf{A}_\varepsilon$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$ and*

$$C = (\varepsilon - 1)^{-1}(b/a)^\varepsilon[a^{-c(\varepsilon)}(a^{c(\varepsilon)} - b^{c(\varepsilon)}) + c(\varepsilon)(a/b)^{2c(\varepsilon)-1}a^{-1}(b-a)].$$

As the last example in this section, we proceed to an important family of absolute inequality indices, the *Kolm-indices* (Kolm (1976a), cf. also the discussion in Section 1.8.1 of Chakravarty (2009)). Given a parameter $\kappa > 0$ the corresponding index is defined as

$$\mathsf{K}_\kappa(F) = \kappa^{-1}\log\left(\int e^{\kappa[\mu(F)-x]}dF(x)\right).$$

The following lemma verifies Assumption 2.2 for this class of inequality indices.

Lemma 4.7. *Let $a < b$ and let $\kappa > 0$. Then the functional $\mathsf{T} = \mathsf{K}_\kappa$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$ and $C = e^{\kappa(b-a)}[b-a] + \kappa^{-1}e^{\kappa b}[e^{-\kappa a} - e^{-\kappa b}]$.*

4.3 Welfare measures

The structurally most elementary welfare measures are of the form

$$F \mapsto \int u(x)dF(x), \tag{29}$$

for a utility function u. Functionals as in Equation (29) are accessible to our theory, but are not our main focus, as they fall into the standard multi-armed bandit framework, because a mean is targeted.

There are many important welfare measures that are not of the simple form (29), but can be obtained as a function of the mean functional and an inequality measure.[9] Many such exceptional measures are related to a relative inequality measure $F \mapsto \mathsf{I}_{\text{rel}}(F)$, say, via the transformation

$$\mathsf{W}(F) = \mu(F)(1 - \mathsf{I}_{\text{rel}}(F)); \tag{30}$$

or are related to an absolute inequality measure $F \mapsto \mathsf{I}_{\text{abs}}(F)$, say, via the transformation

$$\mathsf{W}(F) = \mu(F) - \mathsf{I}_{\text{abs}}(F); \tag{31}$$

we refer to Blackorby and Donaldson (1978), Blackorby and Donaldson (1980) and Dagum (1990) for theoretical background on this relationship between inequality and welfare measures.

While many important welfare measures are of this form, we do not argue that any relative or absolute inequality measure implies a *reasonable* welfare measure through one of the above two relations. In particular, to arrive at a reasonable welfare measure, one may want to impose additional restrictions on the inequality measure, e.g., one may want to assume that the relative inequality measure in Equation (30) satisfies $0 \leq \mathsf{I}_{\text{rel}} \leq 1$, and that the absolute inequality measure in Equation (31) satisfies $0 \leq \mathsf{I}_{\text{abs}} \leq \mu$. These conditions are satisfied by many inequality measures; otherwise, they can often be achieved by re-normalization in case the inequality measures are nonnegative and bounded from above, cf. also the discussion in Chakravarty (2009) p.30. Apart from a boundedness condition concerning the relative inequality measures, such restrictions are not needed in our proof verifying Assumption 2.2 for welfare measures obtained through (30) and (31), and are therefore not incorporated into the lemma given below.

Example 4.8. The Gini-welfare measure from Equation (5) is obtained upon choosing $\mathsf{I}_{\text{abs}} = \mathsf{G}_{\text{abs}}$ (cf. Equation (23)) in Equation (31). That $0 \leq \mathsf{G}_{\text{abs}} \leq \mu$ is well known; an argument may be found in the proof of Lemma 4.2.

The following result allows one to use the results from the preceding section in establishing Assumption 2.2 for welfare measures derived via (30) and (31) from an inequality measure.

[9] Historically, the theoretical foundation of inequality measures was based on a social welfare function, e.g., Dalton (1920) and Atkinson (1970). That is, contrary to our presentation, which started with a discussion of inequality measures, inequality measures were derived from a given social welfare functions. For our presentation, however, it is convenient to base the welfare functions on inequality measures. Articles that start with an inequality measure and derive a corresponding welfare measures from it are Blackorby and Donaldson (1978), Blackorby and Donaldson (1980) or Dagum (1990).

Lemma 4.9. *Let $a < b$ be real numbers. Then, the following holds:*

1. *Let the relative inequality measure I_{rel} satisfy Assumption 2.2 with $\mathscr{D}_{rel}$ and C. Suppose further that $|1 - \mathsf{I}_{rel}| \leq \gamma < \infty$ holds. Then the welfare measure W derived via Equation (30) satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}_{rel}$ and constant $\gamma(b-a)+\max(|a|,|b|)C$.*

2. *Let the absolute inequality measure I_{abs} satisfy Assumption 2.2 with $\mathscr{D}_{abs}$ and C. Then the welfare measure W derived via Equation (31) satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}_{abs}$ and with constant $(b-a)+C$.*

Note that if an absolute inequality measure I_{abs} and a relative inequality measure I_{rel} are related via $\mathsf{I}_{\text{abs}}(F) = \mu(F)\mathsf{I}_{\text{rel}}(F)$ for every F (Blackorby and Donaldson (1980) then call I_{abs} a "compromise index", cf. their Section 5), then the welfare measures obtained via Equation (30) and Equation (31), respectively, coincide. One can then verify Assumption 2.2 via Part 1 or Part 2 in Lemma 4.9. We note that in such a situation Part 2 of the lemma will typically imply weaker restrictions.

Together with the results in the preceding section, Lemma 4.9 verifies Assumption 2.2 for many specific welfare measures. For example, Lemma 4.2 can be used to show that the Gini-welfare measure satisfies Assumption 2.2 with $a < b$ real numbers, $\mathscr{D} = D_{cdf}([a,b])$, and constant $C = 2(b-a)$. Similarly, Lemma 4.4 can be used to verify Assumption 2.2 for all welfare measures corresponding to linear inequality measures. The latter class of welfare measures was recently considered in a different context by Kitagawa and Tetenov (2017).

4.4 Poverty measures

Poverty indices are typically based on a *poverty line*, i.e., a threshold z below which an, e.g., income is classified as "poor." There are two basic approaches to defining z: The absolute approach considers z as fixed (i.e., independent of the underlying income distribution F), whereas the relative approach views $\mathsf{z} = \mathsf{z}(F)$ as a functional of the "income distribution" F. In the relative approach, the poverty line adapts to growth or decline of the economy. To make this formal and to give an example, the following poverty line functional combines both approaches (cf. Kakwani (1986) and Lambert (2001), p.139) in taking a convex combination of a fixed amount z_0 and a centrality measure of the underlying income distribution:

$$\mathsf{z}_{\mathsf{m},z_0,\delta}(F) = z_0 + \delta(\mathsf{m}(F) - z_0) \tag{32}$$

where $z_0 > 0$, $0 \leq \delta \leq 1$, and m is a location functional that either coincides with the mean functional μ, or the median functional $q_{1/2}$. Note in particular that $\mathsf{z}_{\mathsf{m},z_0,0} = z_0$ and $\mathsf{z}_{\mathsf{m},z_0,1} = \mathsf{m}$, i.e., this definition nests both an absolute and a relative approach. Lemma C.1 in Appendix C summarizes conditions under which the poverty line functionals in the family (32) satisfy Assumption 2.2.

The first poverty measure we shall consider is the so-called *headcount ratio*, which is the proportion in a population F that, according to a given poverty line z, qualifies as poor:

$$\mathsf{H}_{\mathsf{z}}(F) = F(\mathsf{z}(F)).$$

For the sake of generality, the following lemma establishes conditions under which the headcount ratio satisfies Assumption 2.2 under high-level conditions concerning the poverty line functional z. Specific constants and domains for the concrete family of poverty lines defined in Equation (32) can immediately be obtained with Lemma C.1 in Appendix C. An analogous remark applies to the poverty measures introduced further below, and will not be restated.

Lemma 4.10. *Let $a < b$ be real numbers, and let $\mathsf{z} : D_{cdf}([a,b]) \to \mathbb{R}$ denote a poverty line functional that satisfies Assumption 2.2 with $\mathscr{D}_{\mathsf{z}}$ and constant C_{z}, say. Let $s > 0$. Then, H_{z} satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}_{\mathsf{z}} \cap \mathscr{D}^s([a,b])$ and $C = C_{\mathsf{z}} s + 1$.*

Certain disadvantages of the headcount ratio motivated Sen (1976) to introduce a different family of poverty measures using an axiomatic approach. We shall now discuss this family in the generalized form of Kakwani (1980). Given a poverty line z and a "sensitivity parameter" $\kappa \geq 1$, say, each element of this family of poverty indices is written as

$$\mathsf{P}_{SK}(F; \mathsf{z}, \kappa) = (\kappa + 1) \int_{[0,\mathsf{z}(F)]} \left[1 - \frac{x}{\mathsf{z}(F)}\right] \left[1 - \frac{F(x)}{F(\mathsf{z}(F))}\right]^{\kappa} dF(x),$$

with the convention that $0/0 := 0$. A result discussing conditions under which $\mathsf{P}_{SK}(F; \mathsf{z}, \kappa)$ satisfies Assumption 2.2, and which is again established under high-level assumptions on the poverty line z, is provided next. Note that in case F is supported on $[0, \infty)$, the poverty line in Equation (32) is greater or equal to $(1 - \delta) z_0$, which is positive unless $\delta = 1$.

Lemma 4.11. *Let $a = 0 < b$, $\kappa \geq 1$, and let $\mathsf{z} : D_{cdf}([a,b]) \to \mathbb{R}$ denote a poverty line functional that satisfies Assumption 2.2 with $\mathscr{D}_{\mathsf{z}}$ and constant C_{z}, say. Suppose further that $\mathsf{z} \geq z_* > 0$ holds for some real number z_*. Let $s > 0$. Then $\mathsf{T} = \mathsf{P}_{SK}(\cdot; \mathsf{z}, \kappa)$ satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}_{\mathsf{z}} \cap \mathscr{D}^s([a,b])$ and $C = (\kappa + 1)\{1 + (b z_*^{-2} + 2\kappa s + s) C_{\mathsf{z}} + 4\kappa\}$.*

Finally, we consider a family, each element of which can be written as

$$\mathsf{P}_{FGT}(F; \mathsf{z}, \Lambda) = \int_{[0,\mathsf{z}(F)]} \Lambda\left(1 - [x/\mathsf{z}(F)]\right) dF(x), \tag{33}$$

where $\Lambda : [0,1] \to [0,1]$ is non-decreasing and Lipschitz continuous. This class contains (at least after monotonic transformations), e.g., the measures of Foster et al. (1984) or Chakravarty (1983) as special cases (cf. Lambert (2001) Chapter 6.3, and also the more recent review in Foster et al. (2010)). The following result provides conditions under which P_{FGT} satisfies Assumption 2.2. Again the result is established under high-level assumptions on the poverty line z.

Lemma 4.12. *Let $a = 0 < b$ and let $\mathsf{z} : D_{cdf}([a,b]) \to \mathbb{R}$ denote a poverty line functional that satisfies Assumption 2.2 with $\mathscr{D}_{\mathsf{z}}$ and constant C_{z}, say. Suppose further that $\mathsf{z} \geq z_* > 0$ holds for some real number z_*, and that $\Lambda : [0,1] \to \mathbb{R}$ is non-decreasing, Lipschitz continuous with constant C_{Λ}, and satisfies $\Lambda(0) = 0$. Then, $\mathsf{T} = \mathsf{P}_{FGT}(\cdot; \mathsf{z}, \Lambda)$ satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}_{\mathsf{z}}$ and $C = b z_*^{-2} C_{\Lambda} C_{\mathsf{z}} + \Lambda(1)$.*

As a direct application of Lemma 4.12, we note that given a poverty line z the poverty measure of Foster et al. (1984) is obtained upon setting $\Lambda(x) = x^\alpha$ for some $\alpha \geq 0$ in Equation (33). The conditions in the preceding lemma are satisfied for $\alpha \geq 1$ (in which case $C_\Lambda = \alpha$). The preceding lemma does not cover the case where $\alpha = 0$. However, note that the functional corresponding to $\Lambda(x) = x^\alpha$ with $\alpha = 0$ coincides with the headcount ratio, which is already covered via Lemma 4.10.

References

Abadie, A. (2002): "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association*, 97, 284–292.

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117.

Athey, S. and G. W. Imbens (2017): "The econometrics of randomized experiments," in *Handbook of Economic Field Experiments*, Elsevier, vol. 1, 73–140.

Athey, S. and S. Wager (2017): "Efficient policy learning," *arXiv preprint arXiv:1702.02896*.

Atkinson, A. B. (1970): "On the measurement of inequality," *Journal of Economic Theory*, 2, 244–263.

Audibert, J.-Y. and A. B. Tsybakov (2007): "Fast learning rates for plug-in classifiers," *Annals of Statistics*, 35, 608–633.

Auer, P., N. Cesa-Bianchi, and P. Fischer (2002): "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, 47, 235–256.

Barrett, G. F. and S. G. Donald (2009): "Statistical inference with generalized Gini indices of inequality, poverty, and welfare," *Journal of Business & Economic Statistics*, 27, 1–17.

Besson, L. and E. Kaufmann (2018): "What doubling tricks can and can't do for multi-armed bandits," *arXiv preprint arXiv:1803.06971.*

Bhattacharya, D. and P. Dupas (2012): "Inferring welfare maximizing treatment assignment under budget constraints," *Journal of Econometrics*, 167, 168–196.

Blackorby, C. and D. Donaldson (1978): "Measures of relative equality and their meaning in terms of social welfare," *Journal of Economic Theory*, 18, 59–80.

——— (1980): "A theoretical treatment of indices of absolute inequality," *International Economic Review*, 21, 107–136.

Bubeck, S. and N. Cesa-Bianchi (2012): "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, 5, 1–122.

Cassel, A., S. Mannor, and A. Zeevi (2018): "A general approach to multi-armed bandits under risk criteria," *arXiv preprint arXiv:1806.01380*.

Chakravarty, S. R. (1983): "A new index of poverty," *Mathematical Social Sciences*, 6, 307–313.

——— (2009): *Inequality, Polarization and Poverty*, New York: Springer.

Chamberlain, G. (2000): "Econometrics and decision theory," *Journal of Econometrics*, 95, 255–283.

Chao, M.-T. and W. Strawderman (1972): "Negative moments of positive random variables," *Journal of the American Statistical Association*, 67, 429–431.

Chernozhukov, V. and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

Cowell, F. (2011): *Measuring Inequality*, Oxford: Oxford University Press.

Cowell, F. A. (1980): "Generalized entropy and the measurement of distributional change," *European Economic Review*, 13, 147–159.

Dagum, C. (1990): "On the relationship between income inequality measures and social welfare functions," *Journal of Econometrics*, 43, 91–102.

Dalton, H. (1920): "The measurement of the inequality of incomes," *Economic Journal*, 30, 348–361.

Davidson, R. and J.-Y. Duclos (2000): "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica*, 68, 1435–1464.

Davidson, R. and E. Flachaire (2007): "Asymptotic and bootstrap inference for inequality and poverty measures," *Journal of Econometrics*, 141, 141 – 166.

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

Doob, J. (1936): "Note on probability," *Annals of Mathematics*, 363–367.

Dudley, R. M. (2002): *Real Analysis and Probability*, Cambridge University Press.

Embrechts, P. and M. Hofert (2013): "A note on generalized inverses," *Mathematical Methods of Operations Research*, 77, 423–432.

Folland, G. B. (1999): *Real Analysis: Modern Techniques and their Applications*, New York: Wiley.

Foster, J., J. Greer, and E. Thorbecke (1984): "A class of decomposable poverty measures," *Econometrica*, 52, 761–766.

——— (2010): "The Foster–Greer–Thorbecke (FGT) poverty measures: 25 years later," *Journal of Economic Inequality*, 8, 491–524.

Garivier, A., T. Lattimore, and E. Kaufmann (2016): "On explore-then-commit strategies," in *Advances in Neural Information Processing Systems*, 784–792.

Gastwirth, J. L. (1971): "A general definition of the Lorenz curve," *Econometrica*, 39, 1037–1039.

——— (1974): "Large sample theory of some measures of income inequality," *Econometrica*, 42, 191–196.

Hirano, K. and J. R. Porter (2009): "Asymptotics for statistical treatment rules," *Econometrica*, 77, 1683–1701.

——— (2018): "Statistical decision rules in econometrics," *Working paper.*

Imbens, G. W. and J. M. Wooldridge (2009): "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47, 5–86.

Jacob, C. (1988): *Statistical Power for the Behavioral Sciences*, Lawrence Erlbaum Associates, Publishers.

Kakwani, N. (1980): "On a class of poverty measures," *Econometrica*, 437–446.

——— (1986): *Analyzing Redistribution Policies: A Study Using Australian Data*, Cambridge: Cambridge University Press.

Kallenberg, O. (2001): *Foundations of Modern Probability*, New York: Springer Science & Business Media, 2 ed.

——— (2005): *Probabilistic Symmetries and Invariance Principles*, New York: Springer.

Kitagawa, T. and A. Tetenov (2017): "Equality-minded treatment choice," *Working paper.*

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

Kock, A. B. and M. Thyrsgaard (2017): "Optimal sequential treatment allocation," *arXiv preprint arXiv:1705.09952.*

Kolm, S.-C. (1976a): "Unequal inequalities. I," *Journal of Economic Theory*, 12, 416–442.

——— (1976b): "Unequal inequalities. II," *Journal of Economic Theory*, 13, 82–111.

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

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

Lattimore, T. and C. Szepesvári (2019): "Bandit algorithms," *preprint*.

Lavori, P. W., R. Dawson, and A. J. Rush (2000): "Flexible treatment strategies in chronic disease: clinical and research implications," *Biological psychiatry*, 48, 605–614.

Liese, F. and K. J. Miescke (2008): *Statistical Decision Theory*, New York: Springer.

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

Manski, C. F. (1988): "Ordinal utility models of decision making under uncertainty," *Theory and Decision*, 25, 79–104.

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

——— (2019a): "Remarks on statistical inference for statistical decisions," Tech. rep., Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

——— (2019b): "Treatment choice with trial data: statistical decision theory should supplant hypothesis testing," *American Statistician*, 73, 296–304.

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

Massart, P. (1990): "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *Annals of Probability*, 18, 1269–1283.

McDonald, J. B. (1984): "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 52, 647–663.

McDonald, J. B. and M. Ransom (2008): "The generalized beta distribution as a model for the distribution of income: estimation of related measures of inequality," in *Modeling Income Distributions and Lorenz Curves*, New York: Springer, 147–166.

Mehran, F. (1976): "Linear measures of income inequality," *Econometrica*, 44, 805–809.

Mills, J. A. and S. Zandvakili (1997): "Statistical inference via bootstrapping for measures of inequality," *Journal of Applied Econometrics*, 12, 133–150.

Minassian, D. (2007): "A mean value theorem for one-sided derivatives," *American Mathematical Monthly*, 114, 28.

Murphy, K. R., B. Myors, and A. Wolach (2014): *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*, Routledge.

Murphy, S. A. (2003): "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 331–355.

——— (2005): "An experimental design for the development of adaptive treatment strategies," *Statistics in medicine*, 24, 1455–1481.

Murphy, S. A., M. J. van der Laan, and J. M. Robins (2001): "Marginal mean models for dynamic regimes," *Journal of the American Statistical Association*, 96, 1410–1423.

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

Perchet, V., P. Rigollet, S. Chassang, and E. Snowberg (2016): "Batched bandit problems," *The Annals of Statistics*, 44, 660–681.

Rigollet, P. and A. Zeevi (2010): "Nonparametric bandits with covariates," *Proceedings of COLT*.

Robbins, H. (1952): "Some aspects of the sequential design of experiments," *Bulletin of the Americal Mathematical Society*, 58, 527–535.

Robins, J. M. (1997): "Causal inference from complex longitudinal data," in *Latent variable modeling and applications to causality*, Springer, 69–117.

Rosenbluth, G. (1951): "Note on Mr. Schutz's measure of income inequality," *American Economic Review*, 41, 935–937.

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

Rothe, C. (2010): "Nonparametric estimation of distributional policy effects," *Journal of Econometrics*, 155, 56–70.

——— (2012): "Partial distributional policy effects," *Econometrica*, 80, 2269–2301.

Schluter, C. and K. J. van Garderen (2009): "Edgeworth expansions and normalizing transforms for inequality measures," *Journal of Econometrics*, 150, 16 – 29.

Schutz, R. R. (1951): "On the measurement of income inequality," *American Economic Review*, 41, 107–122.

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

——— (1976): "Poverty: an ordinal approach to measurement," *Econometrica*, 219–231.

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

——— (2009): *Approximation Theorems of Mathematical Statistics*, vol. 162, New York: Wiley.

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

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

——— (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.

Theil, H. (1967): *Economics and Information Theory*, Amsterdam: North-Holland.

Thistle, P. D. (1990): "Large sample properties of two inequality indices," *Econometrica*, 58, 725–728.

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

Tran-Thanh, L. and J. Y. Yu (2014): "Functional bandits," *arXiv preprint arXiv:1405.2432*.

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

——— (2009): *Introduction to Nonparametric Estimation*, New York: Springer.

Witting, H. and U. Müller-Funk (1995): *Mathematische Statistik II*, B.G. Teubner: Stuttgart.

Appendices

Throughout the appendices, the (unique) probability measure on the Borel sets of $\mathbb{R}$ corresponding to a cdf $F \in D_{cdf}(\mathbb{R})$ will be denoted by μ_F (cf., e.g., Folland (1999), p.35).

We shall freely use standard notation and terminology concerning stochastic kernels (also referred to as Markov kernels or probability kernels) and semi-direct products (i.e., the joint distribution corresponding to a stochastic kernel and a probability measure) see, e.g., Appendix A.3 of Liese and Miescke (2008) in particular their Equation A.3. Furthermore, the random variables and vectors appearing in the proofs are defined on an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with corresponding expectation $\mathbb{E}$, which is (without loss of generality) assumed to be rich enough to support all random variables we work with. Furthermore, we shall denote by ω a generic element of Ω.

We also recall from, e.g., Definition 2.5 in Tsybakov (2009), that the Kullback-Leibler divergence between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathfrak{Y})$ is defined as

$$\mathsf{KL}(P,Q) := \begin{cases} \int_{\mathcal{X}} \log(dP/dQ)dP & \text{if } P \ll Q, \\ \infty & \text{else.} \end{cases} \tag{34}$$

The integral appearing in this definition is well-defined, because the negative part of the integrand is P-integrable. The positive part of the integrand is not necessarily P-integrable. Therefore, $\mathsf{KL}(P,Q) = \infty$ might hold even in case $P \ll Q$. Furthermore, $\mathsf{KL}(P,Q)$ is non-negative, and equals 0 if and only if $P = Q$. Proofs for the just-mentioned facts can be found in Section 2.4 of Tsybakov (2009). Note that the definition of KL does not depend on how one defines $\log(0)$ (for completeness, we set $\log(0) := 0$ in the sequel).

A Auxiliary results

This section develops some auxiliary lemmas that will be used in Appendix B. The following result is a general "chain rule" for Kullback-Leibler divergences. Although well-documented under stronger assumptions, we could not find a reference containing a proof of the following statement, which only requires one of the two σ-algebras involved to be countably generated.

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,*

$$\mathsf{KL}(\mathsf{A} \otimes P, \mathsf{B} \otimes Q) = \int_{\mathcal{X}} \mathsf{KL}(\mathsf{A}(\cdot, x), \mathsf{B}(\cdot, x))dP(x) + \mathsf{KL}(P,Q) = \mathsf{KL}(\mathsf{A} \otimes P, \mathsf{B} \otimes P) + \mathsf{KL}(P,Q). \tag{35}$$

Remark A.2. Inspection of the proof of Lemma A.1 shows that the assumption of $\mathfrak{B}$ being countably generated is only used to verify (via Proposition 1.95 in Liese and Miescke (2008)) that (i) $x \mapsto \mathsf{KL}(\mathsf{A}(\cdot, x), \mathsf{B}(\cdot, x))$ is measurable, and (ii) that $\int_{\mathcal{X}} \mathsf{KL}(\mathsf{A}(\cdot, x), \mathsf{B}(\cdot, x))dP(x)$ coincides with $\mathsf{KL}(\mathsf{A} \otimes P, \mathsf{B} \otimes P)$. In situations where $\mathfrak{B}$ fails to be countably generated, the conclusion in the previous lemma still holds if (i) and (ii) are satisfied.

Proof. We conclude from Proposition 1.95 in Liese and Miescke (2008) that the integral $\int_{\mathcal{X}} \mathsf{KL}(\mathsf{A}(\cdot,x),\mathsf{B}(\cdot,x))dP(x)$ appearing in Equation (35) is well-defined (i.e., the non-negative integrand $x \mapsto \mathsf{KL}(\mathsf{A}(\cdot,x),\mathsf{B}(\cdot,x))$ is measurable), and coincides with $\mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes P)$. This proves the second equality in Equation (35).

To prove the first equality in Equation (35), assume first that $\mathsf{A}\otimes P \not\ll \mathsf{B}\otimes Q$. Then, $\mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes Q)=\infty$ by definition of the KL-divergence. Observe that if $P \ll Q$ and $\mathsf{A}(\cdot,x) \ll \mathsf{B}(\cdot,x)$ for P-almost every x would hold, then $\mathsf{A}\otimes P \ll \mathsf{B}\otimes Q$ would follow. Therefore, either $P \not\ll Q$ holds, or $P \ll Q$ and $A(\cdot,x) \not\ll B(\cdot,x)$ for all x in a set of positive P measure. In both cases the statement in the first equality in Equation (35) holds true by definition and non-negativity of the KL-divergence.

Consider now the case $\mathsf{A}\otimes P \ll \mathsf{B}\otimes Q$. Corollary 1.71 of Liese and Miescke (2008) implies $\mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes Q) \geq \mathsf{KL}(P,Q)$. Therefore, if $\mathsf{KL}(P,Q)=\infty$ Equation (35) holds true. Hence, we can assume that $\mathsf{KL}(P,Q)<\infty$. Choose a density $0 \leq a := d(\mathsf{A}\otimes P)/d(\mathsf{B}\otimes Q)$, and let $p := dP/dQ$ denote the corresponding (marginal) Q-density of P. Denote by $[\log(a)]^+$ and $[\log(a)]^-$ the positive and negative parts, respectively, of $\log(a)$. The negative-part $[\log(a)]^-$ is $\mathsf{A}\otimes P$-integrable (cf. the discussion immediately after Equation (34)). Furthermore, since $\mathsf{KL}(P,Q)$ is finite, $\log(p)$ is $\mathsf{A}\otimes P$-integrable, implying that $[\log(a)]^- + \log(p)$ is $\mathsf{A}\otimes P$-integrable, and we can thus write $\mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes Q) - \mathsf{KL}(P,Q)$ (the first summand might be infinite) as

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{Y}} [\log(a)]^+ d(\mathsf{A}\otimes P) - \left[\int_{\mathcal{X}\times\mathcal{Y}} [\log(a)]^- d(\mathsf{A}\otimes P) + \int_{\mathcal{X}\times\mathcal{Y}} \log(p) d(\mathsf{A}\otimes P)\right] \\
&= \int_{\mathcal{X}\times\mathcal{Y}} [\log(a)]^+ d(\mathsf{A}\otimes P) + \int_{\mathcal{X}\times\mathcal{Y}} -\left([\log(a)]^- + \log(p)\right) d(\mathsf{A}\otimes P),
\end{aligned}$$

which, since $[\log(a)]^+$ is clearly non-negative and measurable, equals (cf., e.g., Theorem 4.1.10 in Dudley (2002))

$$\int_{\mathcal{X}\times\mathcal{Y}} [\log(a) - \log(p)] d(\mathsf{A}\otimes P) = \int_{\mathcal{X}\times\mathcal{Y}} \log(a/p)\mathbb{1}\{p>0\}) d(\mathsf{A}\otimes P),$$

the equality following from $\{(x,y) : a(y,x)=0 \text{ or } p(x)=0\}$ being an $\mathsf{A}\otimes P$-null set. Since $(a/p)\mathbb{1}\{p>0\} = d(\mathsf{A}\otimes P)/d(\mathsf{B}\otimes P)$, the right-hand side in the previous display equals $\mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes P)$, establishing that

$$\mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes Q) = \mathsf{KL}(\mathsf{A}\otimes P,\mathsf{B}\otimes P) + \mathsf{KL}(P,Q).$$

The already established second equality in Equation (35) thus establishes the first. $\square$

Lemma A.3. *Consider probability measures μ_i and ν_i for $i=0,\ldots,m$ on a countably generated measurable space $(\mathcal{Y},\mathfrak{B})$. Set $\mu := \sum_{i=0}^m p_i\mu_i$ and $\nu := \sum_{i=0}^m q_i\nu_i$, where $p_i \geq 0$ and $q_i > 0$ hold for every $i=0,\ldots,m$ and $\sum_{i=0}^n p_i = 1 = \sum_{i=0}^n q_i$. Then*

$$\mathsf{KL}(\mu,\nu) \leq \sum_{i=0}^m \left(p_i \mathsf{KL}(\mu_i,\nu_i) + (p_i-q_i)^2/q_i\right).$$

Proof. Define stochastic kernels $\mathsf{A} : \mathcal{B} \times \{0, \ldots, m\} \to [0, 1]$ and $\mathsf{B} : \mathcal{B} \times \{0, \ldots, m\} \to [0, 1]$ via $\mathsf{A}(A, i) = \mu_i(A)$ and $\mathsf{B}(A, i) = \nu_i(A)$, respectively. Let P be the measure on the power set of $\{0, \ldots, m\}$ defined via $P(i) = p_i$, and let Q be the measure on the power set of $\{0, \ldots, m\}$ defined via $Q(i) = q_i$. From Theorem 1.70 in Liese and Miescke (2008) and the Chain Rule from Lemma A.1 we obtain

$$\mathsf{KL}(\mu, \nu) \leq \mathsf{KL}(\mathsf{A} \otimes P, \mathsf{B} \otimes Q) = \sum_{i=0}^{m} p_i \mathsf{KL}(\mu_i, \nu_i) + \mathsf{KL}(P, Q).$$

But $\mathsf{KL}(P, Q)$ is not greater than $\chi^2(P, Q)$, the χ^2-divergence between P and Q (cf., e.g., Lemma 2.7 in Tsybakov (2009)), the latter being equal to $\sum_{i=0}^{m}(p_i - q_i)^2/q_i$. □

Lemma A.4. *Suppose Assumption 2.9 holds. Then there exist H and H' in $D_{cdf}([a, b])$, $c_- > 0$ and $\varepsilon \in (0, 1/2)$ such that the following properties hold:*

1. *Letting $H_v := (1/2 - v)H + (1/2 + v)H'$, the set $\mathcal{H} := \{H_v : v \in [-1/2, 1/2]\}$ is contained in $\{J_\tau : \tau \in [0, 1]\}$.*
2. *The function $v \mapsto \mathsf{T}(H_v)$ defined on $[-1/2, 1/2]$ is Lipschitz continuous.*
3. *For every $v \in [0, \varepsilon]$ it holds that*

$$\mathsf{T}(H_0) - \mathsf{T}(H_{-v}) \geq c_- v \quad \text{and} \quad \mathsf{T}(H_v) - \mathsf{T}(H_0) \geq c_- v, \tag{36}$$

and that

$$\mathsf{KL}^{1/2}(\mu_{H_{-v}}, \mu_{H_v}) \leq \frac{2}{\sqrt{0.5^2 - \varepsilon^2}} v. \tag{37}$$

Proof. Without loss of generality, we can assume that H_1 and H_2 in Assumption 2.9 satisfy $\mathsf{T}(H_2) < \mathsf{T}(H_1)$; otherwise swap the indices. From Assumption 2.2, which is imposed through Assumption 2.9, it follows that the function $h(\tau) := \mathsf{T}(J_\tau)$ for $\tau \in [0, 1]$ is Lipschitz continuous (recall the definition of J_τ from Equation (9)), and hence almost everywhere differentiable. Furthermore, since $\mathsf{T}(H_2) < \mathsf{T}(H_1)$, the derivative of h must be positive at some point $\tau^* \in (0, 1)$, say. Consequently, there exists a $c > 0$ (e.g., half the derivative of h at τ^*) and an $\varepsilon \in (0, 1/2)$ satisfying $[\tau^* - \varepsilon, \tau^* + \varepsilon] \subseteq (0, 1)$, such that

$$\frac{h(\tau) - h(\tau^*)}{\tau - \tau^*} \geq c \quad \text{for every } \tau \in [\tau^* - \varepsilon, \tau^* + \varepsilon]. \tag{38}$$

Finally, let $H := J_{\tau^* - \varepsilon}$ and $H' := J_{\tau^* + \varepsilon}$ and set $c_- = 2c\varepsilon$. Note that $H_v = J_{\tau^* + 2v\varepsilon}$ for every $v \in [-1/2, 1/2]$. Hence, the first part of the present lemma follows. The second part follows from $\mathsf{T}(H_v) = \mathsf{T}(J_{\tau^* + 2v\varepsilon}) = h(\tau^* + 2v\varepsilon)$, recalling that h is Lipschitz continuous. The statements in Equation (36) follow immediately from Equation (38). Lemma A.3 (applied with $m = 1$, $\mu_0 = \nu_0 = \mu_H$ and $\mu_1 = \nu_1 = \mu_{H'}$, $p_0 = 1/2 + v$ and $q_0 = 1/2 - v$) and a simple calculation shows that $\mathsf{KL}(\mu_{H_{-v}}, \mu_{H_v}) \leq \frac{4v^2}{0.5^2 - \varepsilon^2}$, which establishes (37). □

B Proofs of results in Sections 2.2, 2.3, 2.4 and 3

B.1 Proofs of results in Section 2.2

B.1.1 Proof of Theorem 2.10

Let π be an explore-then-commit policy as in Definition 2.7 that satisfies the corresponding exploration condition with $\eta \in (0,1)$. Fix $n \geq 2$, and fix the randomization measure $\mathbb{P}_G$ (i.e., a probability measure on the Borel sets of $\mathbb{R}$). Since n is fixed, we shall abbreviate $\pi_{n,t} = \pi_t$ in the sequel. Furthermore, we write $\pi^c = \pi^c_n$. By Lemma A.4 there exists a one-parametric family $\mathcal{H} = \{H_v : v \in [-1/2, 1/2]\} \subseteq \{J_\tau : \tau \in [0,1]\} \subseteq \mathscr{D}$, a real number $c_- > 0$, and an $\varepsilon \in (0, 1/2)$, such that for every $v \in [0, \varepsilon]$

$$\mathsf{T}(H_0) - \mathsf{T}(H_{-v}) \geq c_- v, \mathsf{T}(H_v) - \mathsf{T}(H_0) \geq c_- v, \text{ and } \mathsf{KL}^{1/2}(\mu_{H_{-v}}, \mu_{H_v}) \leq \frac{2}{\sqrt{0.5^2 - \varepsilon^2}} v. \quad (39)$$

Fix $v \in (0, \varepsilon]$. We need some further notation: For $j \in \{-v, v\}$ and every $t = 1, \ldots, n$, we denote by $\tilde{\mathbb{P}}^t_{\pi,j}$ the distribution of $Z_t = (Y_{\pi_t(Z_{t-1},G_t),t}, \ldots, Y_{\pi_1(G_1),1})$ on the Borel sets of $\mathbb{R}^t$, for Y_t i.i.d. $\mu_{H_0} \otimes \mu_{Hj}$ and G_t i.i.d. $\mathbb{P}_G$. The expectation corresponding to $\tilde{\mathbb{P}}^t_{\pi,j}$ will be denoted by $\tilde{\mathbb{E}}^t_{\pi,j}$. We shall use $z_t \in \mathbb{R}^t$ as a generic symbol for a realization of Z_t, and $g_t \in \mathbb{R}$ as a generic symbol for realization of G_t. We abbreviate $G^t := (G_t, \ldots, G_1)$. Furthermore, we denote by $R^j_n(\pi)$ the regret of policy π under Y_t i.i.d. $\mu_{H_0} \otimes \mu_{Hj}$ and G_t i.i.d. $\mathbb{P}_G$. We abbreviate $n_1(n) = n_1$ and $n_2 = n - n_1$, and denote the joint distribution of (Z_n, G^n) (under Y_t i.i.d. $\mu_{H_0} \otimes \mu_{H_j}$ and G_t i.i.d. $\mathbb{P}_G$) by $\mathbb{P}^n_{\pi,j}$, with corresponding expectation $\mathbb{E}^n_{\pi,j}$.

From Equation (39) we conclude $\mathsf{T}(H_{-v}) < \mathsf{T}(H_0) < \mathsf{T}(H_v)$. Hence, Treatment 2 is inferior under $\mu_{H_0} \otimes \mu_{H_{-v}}$, but superior under $\mu_{H_0} \otimes \mu_{H_v}$. Therefore, recalling the definition of $S_{i,n}(t)$ and the corresponding notational convention in case $t = n$ from Equation (6) and using the expression for $R_n(\pi)$ given in Equation (7), we obtain (with some abuse of notation[10])

$$\begin{aligned}
\sup_{j \in \{-v,v\}} \mathbb{E}^n_{\pi,j} R^j_n(\pi) &\geq \frac{1}{2} \left(\mathbb{E}^n_{\pi,-v} R^{-v}_n(\pi) + \mathbb{E}^n_{\pi,v} R^v_n(\pi) \right) \\
&= \frac{1}{2} \Big((\mathsf{T}(H_0) - \mathsf{T}(H_{-v})) \mathbb{E}^n_{\pi,-v} S_2(n) + (\mathsf{T}(H_v) - \mathsf{T}(H_0)) \mathbb{E}^n_{\pi,v} S_1(n) \Big) \\
&\geq \frac{c_- v}{2} \Big(\mathbb{E}^n_{\pi,-v} S_2(n) + \mathbb{E}^n_{\pi,v} S_1(n) \Big),
\end{aligned}$$

where the third inequality follows from (39). Using Definition 2.7, the last expression equals

$$\frac{c_- v}{2} \Big(\mathbb{E}^{n_1}_{\pi,-v} S_{2,n}(n_1) + \mathbb{E}^{n_1}_{\pi,v} S_{1,n}(n_1) \Big) + \frac{c_- v}{2} n_2 \Big(\mathbb{E}^{n_1}_{\pi,-v} \mathbb{1}\{\pi^c(z_{n_1}) = 2\} + \mathbb{E}^{n_1}_{\pi,v} \mathbb{1}\{\pi^c(z_{n_1}) = 1\} \Big);$$

furthermore, for $j \in \{-v, v\}$ and $i = 1, 2$, it holds that $\mathbb{E}^{n_1}_{\pi,j} S_{i,n}(n_1) \geq \eta n_1$. Therefore,

$$\frac{c_- v}{2} \Big(\mathbb{E}^{n_1}_{\pi,-v} S_{2,n}(n_1) + \mathbb{E}^{n_1}_{\pi,-v} S_{1,n}(n_1) \Big) \geq c_- v \eta n_1.$$

[10] Here and at many other places in the appendices, it is occasionally convenient to interpret quantities such as $R^j_n(\pi)$ and $S_1(n)$ as functions on the image space of (Z_n, G^n), as opposed to the random variables obtained by plugging (Z_n, G^n) into these functions.

Noting that

$$\mathbb{E}^{n_1}_{\pi,-v}\mathbb{1}\left\{\pi^c(z_{n_1})=2\right\}+\mathbb{E}^{n_1}_{\pi,v}\mathbb{1}\left\{\pi^c(z_{n_1})=1\right\}=\tilde{\mathbb{E}}^{n_1}_{\pi,-v}\mathbb{1}\left\{\pi^c(z_{n_1})=2\right\}+1-\tilde{\mathbb{E}}^{n_1}_{\pi,v}\mathbb{1}\left\{\pi^c(z_{n_1})=2\right\},$$

which is the sum of Type 1 and Type 2 errors of the test $\mathbb{1}\left\{\pi^c(z_{n_1})=2\right\}$ for the testing problem $\tilde{\mathbb{P}}^{n_1}_{\pi,-v}$ against $\tilde{\mathbb{P}}^{n_1}_{\pi,v}$, it follows from Theorem 2.2(iii) in Tsybakov (2009) that

$$\mathbb{E}^{n_1}_{\pi,-v}\mathbb{1}\left\{\pi^c(z_{n_1})=2\right\}+\mathbb{E}^{n_1}_{\pi,v}\mathbb{1}\left\{\pi^c(z_{n_1})=1\right\}\geq\frac{1}{4}\exp\left(-\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1}_{\pi,v})\right).$$

Summarizing, we obtain

$$\sup_{\substack{F^i\in\{J_\tau:\tau\in[0,1]\}\\ i=1,2}}\mathbb{E}[R_n(\pi)]\geq\frac{c_-v\eta}{8}\left[n_1+(n-n_1)\exp\left(-\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1}_{\pi,v})\right)\right].\tag{40}$$

To obtain an upper bound on $\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1}_{\pi,v})$, we argue as follows: Let $j\in\{-v,v\}$, let Y_t be i.i.d. $\mu_{H_0}\otimes\mu_{Hj}$, and let G_t be i.i.d. $\mathbb{P}_G$. Let $t\in\{1,\ldots,n_1\}$. It is easy to verify that the stochastic kernel

$$(A,(z_{t-1},g_t))\mapsto\mu_{H_0}(A)\mathbb{1}_{\{\pi_t(z_{t-1},g_t)=1\}}+\mu_{H_j}(A)\mathbb{1}_{\{\pi_t(z_{t-1},g_t)=2\}}$$

defines a regular conditional distribution (as defined in, e.g., Liese and Miescke (2008) Definition A.36) of $Y_{\pi_t(Z_{t-1},G_t),t}$ given (Z_{t-1},G_t) (dropping the quantities with index $t-1$ in case $t=1$). Since the joint distribution of (Z_{n_1-1},G_{n_1}) is $\tilde{\mathbb{P}}^{n_1-1}_{\pi,j}\otimes\mathbb{P}_G$, we can write $\tilde{\mathbb{P}}^{n_1}_{\pi,j}\otimes\mathbb{P}_G$ as the semi-direct product

$$\tilde{\mathbb{P}}^{n_1}_{\pi,j}\otimes\mathbb{P}_G=\left(\mu_{H_0}\mathbb{1}_{\{\pi_{n_1}(z_{n_1-1},g_{n_1})=1\}}+\mu_{H_j}\mathbb{1}_{\{\pi_{n_1}(z_{n_1-1},g_{n_1})=2\}}\right)\otimes(\tilde{\mathbb{P}}^{n_1-1}_{\pi,j}\otimes\mathbb{P}_G),$$

where in case $n_1=1$ the arguments z_{n_1-1} and the factor $\tilde{\mathbb{P}}^{n_1-1}_{\pi,j}$ need to be dropped. The chain rule in Lemma A.1 (and Tonelli's theorem) hence implies

$$\begin{aligned}\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1}_{\pi,v})&=\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v}\otimes\mathbb{P}_G,\tilde{\mathbb{P}}^{n_1}_{\pi,v}\otimes\mathbb{P}_G)\\&=\mathsf{KL}(\tilde{\mathbb{P}}^{n_1-1}_{\pi,-v}\otimes\mathbb{P}_G,\tilde{\mathbb{P}}^{n_1-1}_{\pi,v}\otimes\mathbb{P}_G)+\tilde{\mathbb{E}}^{n_1-1}_{\pi,-v}\mathbb{E}_G\left(\mathbb{1}_{\{\pi_{n_1}(z_{n_1-1},g_{n_1})=2\}}\right)\mathsf{KL}(\mu_{H_{-v}},\mu_{H_v})\\&\leq\mathsf{KL}(\tilde{\mathbb{P}}^{n_1-1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1-1}_{\pi,v})+\mathsf{KL}(\mu_{H_{-v}},\mu_{H_v}).\end{aligned}$$

By induction, it follows that

$$\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1}_{\pi,v})\leq n_1\mathsf{KL}(\mu_{H_{-v}},\mu_{H_v})\leq c^+v^2n_1$$

for $c^+=c^+(\varepsilon):=\frac{4}{(0.5^2-\varepsilon^2)}$, the second estimate following from (39). This upper bound on $\mathsf{KL}(\tilde{\mathbb{P}}^{n_1}_{\pi,-v},\tilde{\mathbb{P}}^{n_1}_{\pi,v})$ and Equation (40) imply that for every $v\in[0,\varepsilon]$ we have

$$\begin{aligned}\sup_{\substack{F^i\in\{J_\tau:\tau\in[0,1]\}\\ i=1,2}}\mathbb{E}[R_n(\pi)]&\geq\frac{c_-v\eta}{8}\left[n_1+(n-n_1)\exp\left(-c^+v^2n_1\right)\right]\\&\geq\frac{c_-v\eta}{8}n\exp\left(-c^+v^2n_1\right).\end{aligned}\tag{41}$$

To establish the first claim in the theorem, we use that the supremum in Equation (41) is bounded from below by the average of the first lower bound appearing in that Equation applied to $v = \varepsilon$ and to $v = \varepsilon/\sqrt{n_1}$. In particular, after dropping two nonnegative terms in this average, the supremum is found to be bounded from below by

$$\frac{c_- \eta \varepsilon}{16} \left[n_1 + \frac{n - n_1}{\sqrt{n_1}} \exp(-c^+ \varepsilon^2) \right]. \tag{42}$$

We consider two cases: On the one hand, if $n_1 \geq n/2$, the quantity in (42) is not smaller than $\frac{c_- \eta \varepsilon}{32} n$. On the other hand, if $n_1 < n/2$, then the quantity in (42) is not smaller than

$$\frac{c_- \eta \varepsilon}{16} \left[n_1 + \frac{n}{2\sqrt{n_1}} \exp(-c^+ \varepsilon^2) \right] \geq \frac{c_- \eta \varepsilon}{16} \inf_{z \in (0,\infty)} \left[z^2 n + \frac{\sqrt{n}}{2z} \exp(-c^+ \varepsilon^2) \right].$$

The infimum is attained at $z^* = c(\varepsilon) n^{-1/6}$ for $c(\varepsilon) := 4^{-1/3} \exp(-c^+(\varepsilon)\varepsilon^2/3)$, implying the lower bound

$$\frac{c_- \eta \varepsilon}{16} \left[c^2(\varepsilon) + \frac{1}{2c(\varepsilon)} \exp(-c^+(\varepsilon)\varepsilon^2) \right] n^{2/3}.$$

Combining the two cases proves the first statement with constant $c_l = \frac{c_- \varepsilon}{16} \min(0.5, c^2(\varepsilon) + \frac{1}{2c(\varepsilon)} \exp(-c^+(\varepsilon)\varepsilon^2))$.

Upon replacing n_1 by n^* and setting $v = \varepsilon$ in the second line in Equation (41), the second statement in the theorem follows with constant $c_l(n^*) = \frac{c_- \varepsilon}{8} \exp(-c^+(\varepsilon)\varepsilon^2 n^*)$.

B.1.2 Proof Theorem 2.11

Let Y_t be i.i.d. such that the marginal $Y_{i,t}$ has cdf $F^i \in \mathscr{D}$ for $i = 1, \ldots, K$. We denote $\bar{\Delta} := \{i : \Delta_i > 0\}$. If $\bar{\Delta} = \emptyset$, then the statement in the theorem trivially holds. Thus, we assume henceforth that $\bar{\Delta} \neq \emptyset$. Let $n \in \mathbb{N}$ be fixed. In the following, we will abbreviate $\tilde{\pi}_{n,t} = \tilde{\pi}_t$ for $t = 1, \ldots, n$, and will write $n_1 = n_1(n) := \min(K\lceil n^{2/3} \rceil, n)$. We consider two cases:

1) Suppose that $n \leq K\lceil n^{2/3} \rceil$: Note that trivially $R_n(\tilde{\pi}) \leq Cn \leq CK\lceil n^{2/3} \rceil \leq 2CKn^{2/3}$, where we used that Assumption 2.2 implies $\Delta_i \leq C$. Hence, Equation (10) holds.

2) Suppose that $n > K\lceil n^{2/3} \rceil$: Note that $n_1 = K\lceil n^{2/3} \rceil$. Equations (6) and (7) show that

$$R_n(\tilde{\pi}) = \sum_{i \in \bar{\Delta}} \Delta_i S_i(n) = \sum_{i \in \bar{\Delta}} \Delta_i \sum_{t=1}^{n} \mathbb{1}\{\tilde{\pi}_t(Z_{t-1}, G_t) = i\}.$$

Decomposing the last sum, and using $\Delta_i \leq C$ yields

$$R_n(\tilde{\pi}) \leq C \sum_{t=1}^{n_1} \mathbb{1}\{\tilde{\pi}_t(Z_{t-1}, G_t) \in \bar{\Delta}\} + \sum_{i \in \bar{\Delta}} \Delta_i \sum_{t=n_1+1}^{n} \mathbb{1}\{\tilde{\pi}_t(Z_{t-1}, G_t) = i\}.$$

We thus obtain

$$\mathbb{E}R_n(\tilde{\pi}) \leq Cn_1 + \sum_{t=n_1+1}^{n} \sum_{i \in \bar{\Delta}} \Delta_i \mathbb{P}(\tilde{\pi}_t(Z_{t-1}, G_t) = i).$$

By definition, for $t = n_1 + 1, \dots, n$,

$$\tilde{\pi}_t(Z_{t-1}, G_t) = \min \arg\max\{\mathsf{T}(\hat{F}_{i,n_1,n}) : S_{i,n}(n_1) > 0\}, \tag{43}$$

which, in particular, is constant in $t = n_1 + 1, \dots, n$. Hence,

$$\mathbb{E}R_n(\tilde{\pi}) \leq Cn_1 + (n - n_1) \sum_{i \in \bar{\Delta}} \Delta_i \mathbb{P}(\tilde{\pi}_{n_1+1}(Z_{n_1}, G_{n_1+1}) = i). \tag{44}$$

We now develop an upper bound for the probabilities appearing in the previous display. Note that $\bar{\Delta} \neq \{1, \dots, K\}$, fix $i^* \in \bar{\Delta}^c := \{1, \dots, K\} \backslash \bar{\Delta}$, and let $i \in \bar{\Delta}$. From Equation (43) it follows that $\mathbb{P}(\tilde{\pi}_{n_1+1}(Z_{n_1}, G_{n_1+1}) = i)$ is bounded from above by

$$\mathbb{P}(\mathsf{T}(\hat{F}_{i,n_1,n}) \geq \mathsf{T}(\hat{F}_{i^*,n_1,n}), S_{i^*,n}(n_1) > 0, S_{i,n}(n_1) > 0) + \mathbb{P}(S_{i^*,n}(n_1) = 0).$$

For $j \in \{i, i^*\}$, we denote by ι_j the n_1-dimensional random vector with t-th coordinate equal to 1 if $G_t = j$, and equal to 0 otherwise. Furthermore, for every $c \in \{0,1\}^{n_1}$ such that $c \neq 0$, we define the empirical cdf $\hat{G}(c, j) := \|c\|_1^{-1} \sum_{t: c_t \neq 0} \mathbb{1}\{Y_{j,t} \leq \cdot\}$, $\|\cdot\|_1$ denoting the 1-norm. We now define the event $\{\mathsf{T}(\hat{F}_{i,n_1,n}) \geq \mathsf{T}(\hat{F}_{i^*,n_1,n})\} =: M(i, i^*)$, and write the first probability in the previous display as

$$\sum_{a \in \{0,1\}^{n_1} \backslash \{0\}} \sum_{b \in \{0,1\}^{n_1} \backslash \{0\}} \mathbb{P}(M(i, i^*), \iota_i = a, \iota_{i^*} = b) \tag{45}$$

Recall that for every $t = 1, \dots, n_1$ we have

$$\tilde{\pi}_t(Z_{t-1}, G_t) = G_t, \quad \text{with } G_t \text{ uniformly distributed on } \mathcal{I} = \{1, \dots, K\}. \tag{46}$$

On the event where $\iota_i = a$ and $\iota_{i^*} = b$, we can use Equations (8) and (46) to write

$$M(i, i^*) = \big\{\mathsf{T}(G(a, i)) \geq \mathsf{T}(G(b, i^*))\big\},$$

Because, for $j \in \{i, i^*\}$, the random vector ι_j is a measurable function of $G^{n_1} = (G_1, \dots, G_{n_1})$, and since $Y_1, \dots, Y_{n_1}$ is independent of G^{n_1}, it follows that $M(i, i^*)$ and $\{\iota_i = a, \iota_{i^*} = b\}$ are independent, and we can write the double sum in Equation (45) as

$$\sum_{a \in \{0,1\}^{n_1} \backslash \{0\}} \sum_{b \in \{0,1\}^{n_1} \backslash \{0\}} \mathbb{P}\big(\mathsf{T}(G(a, i)) \geq \mathsf{T}(G(b, i^*))\big) \mathbb{P}(\iota_i = a, \iota_{i^*} = b). \tag{47}$$

Since $\Delta_i = \mathsf{T}(F^{i^*}) - \mathsf{T}(F^i)$, we can bound every $\mathbb{P}\big(\mathsf{T}(G(a, i)) \geq \mathsf{T}(G(b, i^*))\big)$ from above by

$$\begin{aligned} & \mathbb{P}\big(|\mathsf{T}(G(a, i)) - \mathsf{T}(F^i)| + |\mathsf{T}(F^{i^*}) - \mathsf{T}(G(b, i^*))| \geq \Delta_i\big) \\ \leq \; & \mathbb{P}\big(|\mathsf{T}(G(a, i)) - \mathsf{T}(F^i)| \geq \Delta_i/2\big) + \mathbb{P}\big(|\mathsf{T}(G(b, i^*)) - \mathsf{T}(F^{i^*})| \geq \Delta_i/2\big). \end{aligned}$$

Using Assumption 2.2, we can bound the latter sum by

$$\mathbb{P}\big(\|G(a, i) - F^i\|_\infty \geq \Delta_i/(2C)\big) + \mathbb{P}\big(\|G(b, i^*) - F^{i^*}\|_\infty \geq \Delta_i/(2C)\big)$$

Hence, the double sum in Equation (47) is seen to be bounded from above by

$$\begin{aligned}&\sum_{a\in\{0,1\}^{n_1}\setminus\{0\}} \mathbb{P}\big(\|G(a,i)-F^i\|_\infty > \Delta_i/(2C)\big)\mathbb{P}(\iota_i = a)\\ &+\sum_{b\in\{0,1\}^{n_1}\setminus\{0\}} \mathbb{P}\big(\|G(b,i^*)-F^{i^*}\|_\infty > \Delta_i/(2C)\big)\mathbb{P}(\iota_{i^*} = b).\end{aligned} \tag{48}$$

The Dvoretzky-Kiefer-Wolfowitz-Massart inequality (note that Equation 1.5 in Massart (1990) obviously remains valid if ">" is replaced by "≥") implies that the first sum in Equation (48) is bounded from above by

$$\sum_{m=1}^{n_1} 2e^{-m\Delta_i^2/(2C^2)}\mathbb{P}(\|\iota_i\|_1 = m) \leq \frac{\sqrt{2}C}{\Delta_i}\sum_{m=1}^{n_1}\frac{1}{\sqrt{m}}\mathbb{P}(\|\iota_i\|_1 = m),$$

where, to obtain the inequality, we used that $ze^{-z^2} \leq (2e)^{-1/2} < 1/2$ for every $z > 0$. An analogous upper bound holds for the second sum in Equation (48). Noting that $\|\iota_j\|_1 = S_{j,n}(n_1)$, and since the distribution of $S_{j,n}(n_1)$ does not depend on $j \in \{i, i^*\}$ (cf. Equation (46)), we therefore see that the double sum in Equation (48) is bounded from above by $\frac{4C}{\Delta_i}\sum_{m=1}^{n_1}\frac{1}{\sqrt{2m}}\mathbb{P}(\|\iota_{i^*}\|_1 = m)$. Summarizing the argument we started after Equation (44), we now obtain

$$\mathbb{E}R_n(\tilde{\pi}) \leq Cn_1 + (n-n_1)\sum_{i\in\bar{\Delta}}\left[4C\left(\sum_{m=1}^{n_1}\frac{1}{\sqrt{2m}}\mathbb{P}(S_{i^*,n}(n_1) = m)\right) + \Delta_i\mathbb{P}(S_{i^*,n}(n_1) = 0)\right],$$

which (using that $\sqrt{2m} \geq \sqrt{m+1}$ for $m \geq 1$, $\Delta_i \leq C$, and Jensen's inequality) gives

$$\mathbb{E}R_n(\tilde{\pi}) \leq Cn_1 + 4C(n-n_1)K\left[\mathbb{E}[1/(S_{i^*,n}(n_1)+1)]\right]^{1/2}.$$

From Equations (6) and (46) it follows that $S_{i^*,n}(n_1)$ is Bernoulli distributed with success probability K^{-1} and "sample size" n_1. Equation 3.4 in Chao and Strawderman (1972) establishes

$$\mathbb{E}\left(1/[S_{i^*,n}(n_1)+1]\right) = K\frac{1-(1-K^{-1})^{n_1+1}}{n_1+1} \leq \frac{K}{n_1}.$$

Therefore,

$$\mathbb{E}R_n(\tilde{\pi}) \leq Cn_1 + 4C(n-n_1)K^{3/2}n_1^{-1/2} \leq 2CKn^{2/3} + 4CK(n^{2/3}-n^{1/3}) \leq 6CKn^{2/3},$$

where the second inequality was obtained from $2Kn^{2/3} \geq n_1 = K\lceil n^{2/3}\rceil \geq Kn^{2/3}$.

B.2 Proofs of results in Section 2.3

B.2.1 Proof of Theorem 2.12

Let $n \in \mathbb{N}$ be fixed, and let $F^i \in \mathscr{D}$ for $i = 1, \ldots, K$. We need to show that $\mathbb{E}[R_n(\hat{\pi})] \leq c(Kn\overline{\log}(n))^{0.5}$ for $c = c(\beta, C)$ as defined in the statement of the theorem. Note that this inequality trivially holds if $\mathsf{T}(F^1) = \ldots = \mathsf{T}(F^K)$. Therefore, we will assume that $\mathsf{T}(F^i)$ is not constant in $i \in \{1, \ldots, K\}$. Because $\hat{\pi}$ is an anytime policy, we shall make use of the notational simplifications discussed right after Assumption 2.6 (e.g., we write $S_i(t)$ instead of $S_{i,n}(t)$ and $\hat{F}_{i,t}$ instead of $\hat{F}_{i,t,n}$).

We now claim that for every i with $\Delta_i > 0$ it holds that

$$\mathbb{E}[S_i(n)] \leq \frac{2C^2\beta\log(n)}{\Delta_i^2} + \frac{\beta+2}{\beta-2}. \tag{49}$$

Before proving this claim, recall from Equation (7) that $\mathbb{E}[R_n(\hat{\pi})] = \sum_{i:\Delta_i>0} \Delta_i \mathbb{E}[S_i(n)]$, which, together with the claim in Equation (49) and $\Delta_i \leq C$ (by Assumption 2.2), yields

$$\begin{aligned}
\mathbb{E}[R_n(\hat{\pi})] &= \sum_{i:\Delta_i>0} \sqrt{\Delta_i^2 \mathbb{E}[S_i(n)]}\sqrt{\mathbb{E}[S_i(n)]} \\
&\leq \sqrt{2C^2\beta\log(n) + C^2(\beta+2)/(\beta-2)} \sum_{i:\Delta_i>0} \sqrt{\mathbb{E}[S_i(n)]} \leq \left[c(\beta, C)\sqrt{\overline{\log}(n)}\right]\sqrt{Kn},
\end{aligned}$$

the last inequality following from the Cauchy-Schwarz inequality and $\sum_{i:\Delta_i>0} \mathbb{E}[S_i(n)] \leq n$. Therefore, to conclude the proof, it remains to prove the statement in Equation (49).

To this end, let i be such that $\Delta_i > 0$. We first note that if $n \leq K$, then $S_i(n) \leq 1$, hence Equation (49) is trivially satisfied in this case. Consider now the case where $n > K$. We set $i^* := \min \arg\max_{i=1,\ldots,K} \mathsf{T}(F^i)$. For $t > K$, from the definition of $\hat{\pi}$, it follows that $S_i(t-1) \geq 1$. We now abbreviate $\{\hat{\pi}_t(Z_{t-1}) = i\}$ by $\{\hat{\pi}_t = i\}$, and will argue that for $t > K$ we have $\{\hat{\pi}_t = i\} \subseteq A_t \cup B_{i,t} \cup C_{i,t}$, where

$$\begin{aligned}
A_t &:= \left\{\mathsf{T}(\hat{F}_{i^*,t-1}) + C\sqrt{\beta\log(t)/(2S_{i^*}(t-1))} \leq \mathsf{T}(F^{i^*})\right\}, \\
B_{i,t} &:= \left\{\mathsf{T}(\hat{F}_{i,t-1}) > \mathsf{T}(F^i) + C\sqrt{\beta\log(t)/(2S_i(t-1))}\right\}, \\
C_{i,t} &:= \left\{\Delta_i < 2C\sqrt{\beta\log(n)/(2S_i(t-1))}\right\} = \left\{S_i(t-1) < 2\beta C^2\log(n)/\Delta_i^2\right\}.
\end{aligned} \tag{50}$$

Indeed, on the complement of $A_t \cup B_{i,t} \cup C_{i,t}$ we have

$$\begin{aligned}
\mathsf{T}(\hat{F}_{i^*,t-1}) + C\sqrt{\beta\log(t)/(2S_{i^*}(t-1))} > \mathsf{T}(F^{i^*}) &= \mathsf{T}(F^i) + \Delta_i \\
&\geq \mathsf{T}(F^i) + 2C\sqrt{\beta\log(n)/(2S_i(t-1))} \\
&\geq \mathsf{T}(F^i) + 2C\sqrt{\beta\log(t)/(2S_i(t-1))} \\
&\geq \mathsf{T}(\hat{F}_{i,t-1}) + C\sqrt{\beta\log(t)/(2S_i(t-1))},
\end{aligned}$$

which implies $\hat{\pi}_t(Z_{t-1}) \neq i$. Hence $\{\hat{\pi}_t = i\} \subseteq A_t \cup B_{i,t} \cup C_{i,t}$ for $t > K$. Setting $u := \lceil 2C^2\beta\log(n)/\Delta_i^2 \rceil$, we therefore obtain (recalling that $n \geq K+1$, and by definition of $\hat{\pi}$)

$$\begin{aligned}
S_i(n) &= \sum_{t=1}^{K} \mathbb{1}_{\{\hat{\pi}_t=i\}} + \sum_{t=K+1}^{n} \mathbb{1}_{\{\hat{\pi}_t=i\}} = 1 + \sum_{t=K+1}^{n} \mathbb{1}_{\{\hat{\pi}_t=i\}} \\
&= 1 + \sum_{t=K+1}^{n} \mathbb{1}_{\{\hat{\pi}_t=i\}\cap C_{i,t}} + \sum_{t=K+1}^{n} \mathbb{1}_{\{\hat{\pi}_t=i\}\cap C_{i,t}^c} \leq u + \sum_{t=K+1}^{n} \mathbb{1}_{A_t \cup B_{i,t}},
\end{aligned} \tag{51}$$

where, to obtain the inequality, we used $1 + \sum_{t=K+1}^{n} \mathbb{1}_{\{\hat{\pi}_t = i\} \cap C_{i,t}} \leq u$. To see the latter inequality, we consider two cases: On the one hand, if $\omega \in \Omega$ is such that $\omega \notin C_{i,t}$ for every $t = K+1, \ldots, n$, then the inequality trivially holds, because $u \geq 1$. On the other hand, denoting by t^* the largest $t \in \{K+1, \ldots, n\}$ such that $\omega \in C_{i,t}$, it follows that

$$1 + \sum_{t=K+1}^{n} \mathbb{1}_{\{\hat{\pi}_t = i\} \cap C_{i,t}}(\omega) \leq \sum_{t=1}^{t^*} \mathbb{1}_{\{\hat{\pi}_t = i\}}(\omega) = S_i(t^*)(\omega) \leq S_i(t^* - 1)(\omega) + 1 \leq u,$$

where, for the last inequality, we used the second expression for $C_{i,t}$ in Equation (50). From the upper bound in Equation (51) we get

$$\mathbb{E}[S_i(n)] \leq u + \sum_{t=K+1}^{n} \left[\mathbb{P}(A_t) + \mathbb{P}(B_{i,t})\right].$$

We will show further below that for $t = K+1, \ldots, n$ we have:

$$\begin{aligned} \mathbb{P}(A_t) &\leq \sum_{s=1}^{t} \mathbb{P}\big(\mathsf{T}(F_{i^*,s}) + C\sqrt{\beta \log(t)/(2s)} \leq \mathsf{T}(F^{i^*})\big) \\ \mathbb{P}(B_{i,t}) &\leq \sum_{s=1}^{t} \mathbb{P}\big(\mathsf{T}(F_{i,s}) > \mathsf{T}(F^{i}) + C\sqrt{\beta \log(t)/(2s)}\big), \end{aligned} \tag{52}$$

where for every $s \in \{1, \ldots, t\}$ and every $l \in \{i, i^*\}$ we define $F_{l,s} := s^{-1} \sum_{j=1}^{s} \mathbb{1}_{\{Y_{l,j} \leq \cdot\}}$. From Equation (52), Assumption 2.2 and the Dvoretzky-Kiefer-Wolfowitz-Massart inequality (note that Equation 1.5 in Massart (1990) obviously remains valid if "$>$" is replaced by "$\geq$"), we then obtain

$$\begin{aligned} \mathbb{P}(A_t) &\leq \sum_{s=1}^{t} \mathbb{P}(||F_{i^*,s} - F^{i^*}||_\infty \geq \sqrt{\beta \log(t)/(2s)}) \leq 2 \sum_{s=1}^{t} \frac{1}{t^\beta} = \frac{2}{t^{\beta-1}} \\ \mathbb{P}(B_{i,t}) &\leq \sum_{s=1}^{t} \mathbb{P}(||F_{i,s} - F^{i}||_\infty > \sqrt{\beta \log(t)/(2s)}) \leq 2 \sum_{s=1}^{t} \frac{1}{t^\beta} = \frac{2}{t^{\beta-1}}. \end{aligned}$$

The identity

$$\sum_{t=K+1}^{n} \frac{1}{t^{\beta-1}} \leq \int_K^\infty \frac{1}{x^{\beta-1}} dx = \frac{1}{(\beta-2)K^{\beta-2}} \leq \frac{1}{\beta-2} \tag{53}$$

combined with $u \leq 1 + 2C^2 \beta \log(n)/\Delta_i^2$ now establishes (49).

It remains to verify the two inequalities in Equation (52). We need the following two results, the proofs of which are given below.

Lemma B.1. *For every $i \in \{1, \ldots, K\}$, every $r \in \mathbb{N}$, and every $\omega \in \Omega$ we have*

$$t_{i,r}(\omega) := \inf\left\{ s \in \mathbb{N} : \sum_{j=1}^{s} \mathbb{1}_{\{\hat{\pi}_j(Z_{j-1}) = i\}}(\omega) = r \right\} \in \mathbb{N}.$$

The preceding lemma and Doob's optional skipping theorem (cf. Doob (1936), see also Kallenberg (2005) for a modern formulation) establishes the following result.

Lemma B.2. *For every $i \in \{1, \ldots, K\}$ and every $m \in \mathbb{N}$ the joint distribution of $Y_{i,1}, \ldots, Y_{i,m}$ coincides with the joint distribution of $Y_{i,t_{i,1}}, \ldots, Y_{i,t_{i,m}}$.*

Now, to obtain the upper bounds in Equation (52), let $t \in \{K+1, \ldots, n\}$, and note that

$$\begin{aligned}\mathbb{P}(B_{i,t}) &= \mathbb{P}(\mathsf{T}(\hat{F}_{i,t-1}) > \mathsf{T}(F^i) + C\sqrt{\beta \log(t)/(2S_i(t-1))}) \\ &= \sum_{s=1}^{t} \mathbb{P}(\mathsf{T}(\hat{F}_{i,t-1}) > \mathsf{T}(F^i) + C\sqrt{\beta \log(t)/(2s)}, S_i(t-1) = s).\end{aligned}$$

On the event $\{S_i(t-1) = s\}$, we have $\hat{F}_{i,t-1} = s^{-1}\sum_{j=1}^{s} \mathbb{1}\{Y_{i,t_{i,j}} \leq \cdot\}$. Hence, the sum in the second line of the previous display is not greater than

$$\sum_{s=1}^{t} \mathbb{P}\left[\mathsf{T}(s^{-1}\sum_{j=1}^{s} \mathbb{1}\{Y_{i,t_{i,j}} \leq \cdot\}) > \mathsf{T}(F^i) + C\sqrt{\beta \log(t)/(2s)}\right].$$

Lemma B.2 now shows that the joint distribution of $Y_{i,t_{i,1}}, \ldots, Y_{i,t_{i,s}}$ coincides with the joint distribution of $Y_{i,1}, \ldots, Y_{i,s}$. It thus follows that we can replace $Y_{i,t_{i,1}}, \ldots, Y_{i,t_{i,s}}$ by $Y_{i,1}, \ldots, Y_{i,s}$ in the previous display. In other words, we can replace $s^{-1}\sum_{j=1}^{s} \mathbb{1}\{Y_{i,t_{i,j}} \leq \cdot\}$ by $F_{i,s}$ (defined after Equation (52)), from which the upper bound on $\mathbb{P}(B_{i,t})$ in Equation (52) follows. The upper bound on $\mathbb{P}(A_t)$ is obtained analogously.

Proof of Lemma B.1: We argue by contradiction. Suppose there would exist a triple l, r, ω such that $\sum_{j=1}^{s} \mathbb{1}_{\{\hat{\pi}_j(Z_{j-1})=l\}}(\omega) = S_l(s)(\omega) < r$ for every $s \in \mathbb{N}$, implying that

$$1 \leq \sum_{j=1}^{\infty} \mathbb{1}_{\{\hat{\pi}_j(Z_{j-1})=l\}}(\omega) =: \kappa(\omega) < r, \tag{54}$$

where we used that $t_{l,1}(\omega) = l$. From the definition of $\hat{\pi}$ it follows for $t > K$ that

$$\hat{\pi}_t(Z_{t-1}(\omega)) \in \arg\max_{j \in \mathcal{I}} \left\{\mathsf{T}(\hat{F}_{j,t-1}(\cdot)(\omega)) + C\sqrt{\beta \log(t)/(2S_j(t-1)(\omega))}\right\}.$$

For notational convenience, we shall write $\hat{F}_{j,t-1}(\omega)$ instead of $\hat{F}_{j,t-1}(\cdot)(\omega)$. From Equation (54) and the previous display, it follows that for all t large enough

$$\begin{aligned}&\mathsf{T}(\hat{F}_{\hat{\pi}_t(Z_{t-1}(\omega)),t-1}(\omega)) + C\sqrt{\beta \log(t)/(2S_{\hat{\pi}_t(Z_{t-1}(\omega))}(t-1)(\omega))} \\ \geq\ &\mathsf{T}(\hat{F}_{l,t-1}(\omega)) + C\sqrt{\beta \log(t)/(2S_l(t-1)(\omega))},\end{aligned}$$

which is equivalent to (recall that $C > 0$ from the discussion in the first paragraph of the present subsection)

$$a_t\left[\mathsf{T}(\hat{F}_{\hat{\pi}_t(Z_{t-1}(\omega)),t-1}(\omega)) - \mathsf{T}(\hat{F}_{l,t-1}(\omega))\right] \geq \left[[S_l(t-1)(\omega)]^{-1/2} - [S_{\hat{\pi}_t(Z_{t-1}(\omega))}(t-1)(\omega)]^{-1/2}\right],$$

where $a_t := [C\sqrt{\beta \log(t)/2}]^{-1} \to 0$ as $t \to \infty$. The sequence on the left hand side of the previous inequality converges to 0 as $t \to \infty$. To see this, let $F \in \mathscr{D}$ and note that

$$
\begin{aligned}
|\mathsf{T}(\hat{F}_{\hat{\pi}_t(Z_{t-1}(\omega)),t-1}(\omega)) - \mathsf{T}(\hat{F}_{l,t-1}(\omega))| \leq & |\mathsf{T}(\hat{F}_{\hat{\pi}_t(Z_{t-1}(\omega)),t-1}(\omega)) - \mathsf{T}(F)| + |\mathsf{T}(F) - \mathsf{T}(\hat{F}_{l,t-1}(\omega))| \\
\leq & 2C.
\end{aligned}
$$

It thus follows that

$$
\limsup_{t\to\infty} \left[[S_l(t-1)(\omega)]^{-1/2} - [S_{\hat{\pi}_t(Z_{t-1}(\omega))}(t-1)(\omega)]^{-1/2} \right] \leq 0,
$$

or equivalently, noting that $\lim_{t\to\infty} S_l(t-1)(\omega) = \kappa(\omega)$ by Equation (54), that

$$
\limsup_{t\to\infty} S_{\hat{\pi}_t(Z_{t-1}(\omega))}(t-1)(\omega) \leq \kappa(\omega).
$$

This, however, implies that $\lim_{t\to\infty} S_j(t-1)(\omega) < \infty$ for every $j = 1, \ldots, K$. To see the latter, suppose $\lim_{t\to\infty} S_j(t-1)(\omega) = \infty$ holds for treatment j. Define the subsequence $t' := \{t \in \mathbb{N} : \hat{\pi}_t(Z_{t-1}(\omega)) = j\}$ of $\mathbb{N}$ (if $\lim_{t\to\infty} S_j(t-1)(\omega) = \infty$ this is indeed a subsequence). Next, observe that the sequences $S_j(t'-1)(\omega) = S_{\hat{\pi}_{t'}(Z_{t'-1}(\omega))}(t'-1)(\omega)$, a contradiction to the previous display. □

Proof of Lemma B.2: We apply Doob's optional skipping theorem, cf. Proposition 4.1 in Kallenberg (2005). To verify the conditions there, denote by $\mathcal{F}$ the natural filtration corresponding to the i.i.d. sequence $(Y_t)_{t\in\mathbb{N}}$. By Lemma B.1, for every $r \in \mathbb{N}$, the function $t_{i,r}$ takes its values in $\mathbb{N}$; furthermore, it is easy to see that $\{t_{i,r} = s\} \in \mathcal{F}_{s-1}$ for every $s \in \mathbb{N}$. Therefore, $t_{i,r}$ is an $\mathcal{F}$-predictable time. In addition, $t_{i,1} < \cdots < t_{i,m}$ holds by definition. The statement in the lemma now follows from Proposition 4.1 in Kallenberg (2005) (the remaining assumptions there following immediately as $(Y_t)_{t\in\mathbb{N}}$ is i.i.d). □

B.2.2 Proof of Theorem 2.13

Proof of Theorem 2.13. Let π be a policy and let $n \in \mathbb{N}$. Fix the randomization measure $\mathbb{P}_G$. Since n is fixed, we shall abbreviate $\pi_{n,t} = \pi_t$ in the sequel. As in the proof of Theorem 2.10, we obtain from Lemma A.4 a one-parametric family $\mathcal{H} \subseteq \{J_\tau : \tau \in [0,1]\}$, a $c_- > 0$ and an $\varepsilon \in (0, 1/2)$, such that the statement in Equation (39) holds for every $v \in [0, \varepsilon]$, from which it follows that

$$
\mathsf{KL}^{1/2}(\mu_{H_{-v}}, \mu_{H_v}) \leq \frac{2}{c_-\sqrt{0.5^2 - \varepsilon^2}} \min_{j\in\{-v,v\}} |\mathsf{T}(H_j) - \mathsf{T}(H_0)| \quad \text{for every } v \in [0, \varepsilon]. \tag{55}
$$

For ease of notation, we set $\zeta := \frac{2}{c_-\sqrt{0.5^2-\varepsilon^2}}$ and define $f(v) := \zeta \min_{j\in\{-v,v\}} |\mathsf{T}(H_j) - \mathsf{T}(H_0)|$ for every $v \in [0, \varepsilon]$. Note that $f(0) = 0$, and that $f(\varepsilon) \geq 2\varepsilon/\sqrt{0.5^2 - \varepsilon^2} > 0$ by (39). By continuity of f (following from continuity of $v \mapsto \mathsf{T}(H_v)$ as guaranteed by Lemma A.4) and the intermediate-value theorem, we can choose $a_n \in (0, \varepsilon]$ such that $f^2(a_n) = \min(f^2(\varepsilon), 8)/n$.

As in the proof of Theorem 2.10, for $j \in \{-a_n, a_n\}$ and every $t = 1, \ldots, n$, we denote by $\tilde{\mathbb{P}}^t_{\pi,j}$ the distribution induced by $Z_t = (Y_{\pi_t(Z_{t-1},G_t),t}, \ldots, Y_{\pi_1(G_1),1})$ for Y_t i.i.d. $\mu_{H_0} \otimes \mu_{Hj}$ and G_t i.i.d. $\mathbb{P}_G$. We abbreviate $G^t := (G_t, \ldots, G_1)$; denote by $R^j_n(\pi)$ the regret of policy π

and denote the joint distribution of (Z_n, G^n) by $\mathbb{P}^n_{\pi,j}$ (each under Y_t i.i.d. $\mu_{H_0} \otimes \mu_{H_j}$ and G_t i.i.d. $\mathbb{P}_G$), with corresponding expectation $\mathbb{E}^n_{\pi,j}$.

Arguing similarly as around the second display in the proof of Theorem 2.10, we obtain

$$\mathbb{E}^n_{\pi,-a_n} R_n^{-a_n}(\pi) \geq \frac{f(a_n)}{\zeta} \mathbb{E}^n_{\pi,-a_n} S_2(n) \quad \text{and} \quad \mathbb{E}^n_{\pi,a_n} R_n^{a_n}(\pi) \geq \frac{f(a_n)}{\zeta}(n - \mathbb{E}^n_{\pi,a_n} S_2(n)), \tag{56}$$

and

$$\sup_{j \in \{-a_n, a_n\}} \mathbb{E}^n_{\pi,j} R_n^j(\pi) \geq \frac{f(a_n)}{2\zeta} \left(\mathbb{E}^n_{\pi,-a_n} S_2(n) + \left[n - \mathbb{E}^n_{\pi,a_n} S_2(n) \right] \right) \geq \frac{f(a_n) n}{8\zeta} e^{-\mathsf{KL}(\mathbb{P}^n_{\pi,-a_n}, \mathbb{P}^n_{\pi,a_n})},$$

the second inequality following from Theorem 2.2(iii) in Tsybakov (2009), using that $n^{-1} S_2(n)$ is a test for $H_0 : \mathbb{P}^n_{\pi,-a_n}$ against $H_1 : \mathbb{P}^n_{\pi,a_n}$. By the same argument as used after Equation (40) (but now with n instead of "n_1", and with a_n instead of "v"), we obtain $\mathsf{KL}(\mathbb{P}^n_{\pi,-a_n}, \mathbb{P}^n_{\pi,a_n}) = \mathsf{KL}(\mu_{H_{-a_n}}, \mu_{H_{a_n}}) \mathbb{E}^n_{\pi,-a_n} S_2(n) \leq f^2(a_n) \mathbb{E}^n_{\pi,-a_n} S_2(n)$, the inequality being a consequence of Equation (55). Together with the first inequality in Equation (56), which also provides a lower bound for the supremum in Equation (11), this shows that the supremum in (11) is bounded from below by

$$\max\left(\frac{f(a_n) n}{8\zeta} \exp\left(-f^2(a_n) \mathbb{E}^n_{\pi,-a_n} S_2(n) \right), \frac{f(a_n)}{\zeta} \mathbb{E}^n_{\pi,-a_n} S_2(n) \right).$$

Using $\max(x, y) \geq (x+y)/2$ and recalling $f^2(a_n) = \min(f^2(\varepsilon), 8)/n$ leads to the lower bound

$$\begin{aligned} \sqrt{n}\, \frac{1}{2\zeta} \frac{1}{\sqrt{\min(f^2(\varepsilon), 8)}} \inf_{z \geq 0} \left(\frac{\min(f^2(\varepsilon), 8)}{8} \exp(-z) + z \right) &\geq \sqrt{n}\, \frac{1}{16\zeta} \sqrt{\min(f^2(\varepsilon), 8)} \\ &= \sqrt{n}\, \frac{c_- \sqrt{0.5^2 - \varepsilon^2}}{32} \sqrt{\min(f^2(\varepsilon), 8)}, \end{aligned}$$

where we used that $w \exp(-z) + z \geq w$ for every $z \geq 0$ and every $0 \leq w \leq 1$, and recalled that $\zeta = \frac{2}{c_- \sqrt{0.5^2 - \varepsilon^2}}$. □

B.3 Proofs of the claims in Subsection 2.4

B.3.1 Null rejection probability of the test used in the ETC-T policy

Let F^1 and F^2 in $\mathscr{D} = D_{cdf}([0,1])$ be such that $\mathsf{W}(F^1) = \mathsf{W}(F^2)$. Then, for every natural number $n_1 \geq 2$, we can bound $\mathbb{P}(|\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(\hat{F}_{2,n_1})| \geq c_\alpha)$ from above by

$$\sum_{i=1}^{2} \mathbb{P}\left(|\mathsf{W}(\hat{F}_{i,n_1}) - \mathsf{W}(F^i)| \geq \frac{c_\alpha}{2} \right) \leq \sum_{i=1}^{2} \mathbb{P}\left(\|\hat{F}_{i,n_1} - F^i\|_\infty \geq \frac{c_\alpha}{2C} \right),$$

where, to obtain the inequality, we used that W satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([0,1])$ and C. Now, noting that the cyclical assignment rule leads to $\hat{F}_{i,n_1}$ being based on at least $\lfloor n_1/2 \rfloor$ independent observations from F^i, we can use the Dvoretzky-Kiefer-Wolfowitz-Massart (DKWM) inequality to further bound the double sum to the right in the previous display by $4\exp(-\lfloor n_1/2 \rfloor c_\alpha^2/(2C^2)) = \alpha$, recalling that by definition $c_\alpha = \sqrt{2\log(4/\alpha) C^2/\lfloor n_1/2 \rfloor}$.

B.3.2 Power guarantee concerning the choice of n_1 in the ETC-T policy

Let F^1 and F^2 in $\mathscr{D} = D_{cdf}([0,1])$ satisfy $\Delta = |\mathsf{W}(F^1) - \mathsf{W}(F^2)| > 0$, let $\eta \in (0,1)$, and set $n_1 = 2\lceil 8 \log(4/\min(\alpha,\eta))C^2/\Delta^2 \rceil$. Assume first that $\Delta = \mathsf{W}(F^1) - \mathsf{W}(F^2)$ (the other case is handled similarly). Then, the probability that the test does not reject equals

$$\begin{aligned}
\mathbb{P}(|\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(\hat{F}_{2,n_1})| < c_\alpha) &\leq \mathbb{P}(\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(\hat{F}_{2,n_1}) < c_\alpha) \\
&= \mathbb{P}(\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(F^1) + \mathsf{W}(F^2) - \mathsf{W}(\hat{F}_{2,n_1}) < c_\alpha - \Delta) \\
&\leq \mathbb{P}(\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(F^1) + \mathsf{W}(F^2) - \mathsf{W}(\hat{F}_{2,n_1}) < -\Delta/2),
\end{aligned}$$

where we used $c_\alpha \leq \Delta/2$ to obtain the last inequality. This can be upper bounded by

$$\begin{aligned}
&\mathbb{P}(\mathsf{W}(\hat{F}_{1,n_1}) - \mathsf{W}(F^1) < -\Delta/4) + \mathbb{P}(\mathsf{W}(F^2) - \mathsf{W}(\hat{F}_{2,n_1}) < -\Delta/4) \\
\leq\ &\sum_{i=1}^{2} \mathbb{P}(|\mathsf{W}(\hat{F}_{i,n_1}) - \mathsf{W}(F^i)| > \Delta/4) \ \leq\ 4\exp(-\lfloor n_1/2 \rfloor \Delta^2/(8C^2)) \leq \min(\alpha,\eta) \leq \eta,
\end{aligned}$$

where (as in the previous subsection) we used Assumption 2.2 and the DKWM inequality.

B.3.3 Regret guarantee concerning the choice of n_1 in the ETC-ES policy

Let F^1 and F^2 in $\mathscr{D} = D_{cdf}([0,1])$ be arbitrary, let $\delta > 0$, and set $n_1 = 2\lceil 16C^2/(\delta^2 \exp(1)) \rceil$. If $\mathsf{W}(F^1) = \mathsf{W}(F^2)$, then $\mathbb{E}(\max_{i \in \mathcal{I}} \mathsf{W}(F^i) - \mathsf{W}(F^{\pi_n^c(Z_{n_1})})) = 0$. Suppose next that $\Delta := \mathsf{W}(F^1) - \mathsf{W}(F^2) > 0$. Then,

$$\begin{aligned}
\mathbb{E}(\max_{i \in \mathcal{I}} \mathsf{W}(F^i) - \mathsf{W}(F^{\pi_n^c(Z_{n_1})})) &= \Delta \mathbb{P}(\mathsf{W}(\hat{F}_{2,n_1}) > \mathsf{W}(\hat{F}_{1,n_1})) \\
&= \Delta \mathbb{P}(\mathsf{W}(\hat{F}_{2,n_1}) - \mathsf{W}(F^2) + \mathsf{W}(F^1) - \mathsf{W}(\hat{F}_{1,n_1}) > \Delta),
\end{aligned}$$

which, by Assumption 2.2 and the DKWM inequality, is not greater than

$$4\Delta \exp(-\lfloor n_1/2 \rfloor \Delta^2/(2C^2)) \leq 4 \max_{z>0} \left[z \exp(-\lfloor n_1/2 \rfloor z^2/(2C^2)) \right] \leq 4\sqrt{C^2/\lfloor n_1/2 \rfloor} e^{-1/2} \leq \delta.$$

The remaining case where $\mathsf{W}(F^2) > \mathsf{W}(F^1)$ is established analogously.

B.4 Proofs of results in Section 3

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 (16) 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 $\pi_{n,t}(X_t, Z_{t-1})$ in this section.

We provide two auxiliary results that will be useful in the proofs of Theorems 3.6 and 3.12.

Lemma B.3. *Suppose that Assumptions 2.2 and 3.4 are satisfied (the latter with $\gamma \in (0,1]$ and $L > 0$), and assume that the inclusion in Equation* (12) *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 Theorem 3.6, 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*

$$|\mathsf{T}(F^i(\cdot,x))-\mathsf{T}(F^i(\cdot,\tilde{x}))|\le CLV_{n,j}^{\gamma} \quad \text{and} \quad |\mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big)-\mathsf{T}\big(F^{\pi^\star(\tilde{x})}(\cdot,\tilde{x})\big)|\le CLV_{n,j}^{\gamma}; \tag{57}$$

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

$$|\mathsf{T}(F^i_{n,j})-\mathsf{T}(F^i(\cdot,x))|\le CLV_{n,j}^{\gamma} \quad \text{and} \quad |\mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^*_{n,j})|\le CLV_{n,j}^{\gamma}. \tag{58}$$

Proof. Fix i, j, x and $\tilde{x}$ as in the statement of the lemma. By Assumption 3.4

$$||F^i(\cdot,x)-F^i(\cdot,\tilde{x})||_\infty\le L||x-\tilde{x}||^\gamma\le LV_{n,j}^{\gamma} \tag{59}$$

Assumption 2.2 and (12) thus imply the first inequality in (57), and the second follows from

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

Next, assume that $\mathbb{P}_X(B_{n,j})>0$. For every $y\in\mathbb{R}$, from Equation (59), we obtain

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

The first inequality in (58) is now a direct consequence of Assumption 2.2 and (12) (noting that $F^i_{n,j}\in D_{cdf}([a,b])$), and the second inequality follows via

$$|\mathsf{T}(F^{\pi^\star(x)}(\cdot,x))-\mathsf{T}(F^*_{n,j})|=\big|\max_{i\in\mathcal{I}}\mathsf{T}(F^i(\cdot,x))-\max_{i\in\mathcal{I}}\mathsf{T}(F^i_{n,j})\big|\le\max_{i\in\mathcal{I}}|\mathsf{T}(F^i(\cdot,x))-\mathsf{T}(F^i_{n,j})|.$$

$\square$

Lemma B.4. *Suppose Assumption 2.2 is satisfied and that* $\mathscr{D}$ *is convex. Suppose further that* $\mathbb{P}_{Y,X}$ *is such that Equation* (12) *holds, and that Assumption 3.4 is satisfied. Then, for every Borel set* $B\subseteq[0,1]^d$ *that satisfies* $\mathbb{P}_X(B)>0$ *and every* $i=1,\ldots,K$, *the cdf*

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

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

Proof. Let $i\in\{1,\ldots,K\}$. 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 $B_{m,1},\ldots,B_{m,l_m}$ for $m\in\mathbb{N}$ be a triangular array of partitions of $[0,1]^d$, such that the maximal diameter $v_m:=\sup_{i=1,\ldots,l_m}\sup_{x_1,x_2\in B_{m,i}}\|x_1-x_2\|\to0$ 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

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

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 (12). From Assumption 3.4, 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,

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

□

B.4.1 Proof of Theorem 3.6

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 (12), and Assumption 3.4 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

$$\tilde{R}_j(\bar{\pi}) := \sum_{t=1}^n \left[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot, X_t)\big) - \mathsf{T}\big(F^{\bar{\pi}_t(X_t)}(\cdot, X_t)\big)\right] \mathbb{1}_{\{X_t \in B_j\}}, \tag{60}$$

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 2.2) 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 (16), which we here abbreviate as F^i_j, each summand in (60) can be written as

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

which, by Lemma B.3, is not greater than $\mathsf{T}(F^*_j) - \mathsf{T}(F_j^{\bar{\pi}_t(X_t)}) + 2CLV_j^\gamma$. Therefore, we obtain

$$\tilde{R}_j(\bar{\pi}) \leq \sum_{t=1}^n \left[\mathsf{T}\big(F^*_j\big) - \mathsf{T}\big(F_j^{\bar{\pi}_t(X_t)}\big)\right] \mathbb{1}_{\{X_t \in B_j\}} + 2CLV_j^\gamma \sum_{t=1}^n \mathbb{1}_{\{X_t \in B_j\}}. \tag{61}$$

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 $c = c(\beta, C)$ (cf. Theorem 2.12) it holds that

$$\mathbb{E}\left(\sum_{t=1}^n \left[\mathsf{T}\big(F^*_j\big) - \mathsf{T}\big(F_j^{\bar{\pi}_t(X_t)}\big)\right] \mathbb{1}_{\{X_t \in B_j\}}\right) \leq c\sqrt{Kn\mathbb{P}_X(B_j)\overline{\log}(n\mathbb{P}_X(B_j))}. \tag{62}$$

To this end we will use a conditioning argument in combination with Theorem 2.12. Define for every $v = (v_1, \ldots, v_n) \in \{0, 1\}^n$ the event

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

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

$$\mathbb{E}(f) = \sum_{v \in \{0,1\}^n} \mathbb{E}(\mathbb{1}_{\Omega(v)} f) = \sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v))\mathbb{E}(f|\Omega(v)), \tag{63}$$

where (as usual) we define

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

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 3) to rewrite

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

By definition, $W^s = (Y_{\hat{\pi}_{s-1}(W^{s-1}),t_s}, W^{s-1})$ (where W^0 is the empty vector). 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

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

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

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

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

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

where $Z_s^* = (Y^*_{\hat{\pi}_s(Z_{s-1}^*),s}, \ldots, Z_{s-1}^*)$ (and where Z_0^* is the empty vector). The r-th marginal of $\mathbb{Q}$ has cdf F_j^r, which by Lemma B.4 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 2.12, applied with $\mathrm{cl}(\mathscr{D})$ (cf. Remark 2.4) 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 (63) (noting that f vanishes on $\Omega(0)$) we see that

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

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 (62).

B.4.2 Proof of Corollary 3.7

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 (12), Assumption 3.1 with $\underline{c}$ and $\bar{c}$, and Assumption 3.4 with L and γ. We shall apply Theorem 3.6 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 3.1, we obtain $\mathbb{P}_X(B_{n,j}) \leq \bar{c}P^{-d}$. Therefore, Equation (17) implies the upper bound

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

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

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

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

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

which proves the theorem.

B.4.3 Proof of Theorem 3.12

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 (19) 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 3.11 with $\delta = c_1P^{-\gamma}$ in the arguments below). To see this, note that, by Assumption 2.2, 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 3.12 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 (12), Assumption 3.1 with $\underline{c}$ and $\bar{c}$, Assumption 3.4 with L and γ, and Assumption 3.11 with $\alpha \in (0,1)$ and $C_0 > 0$.

We establish $\mathbb{E}[R_n(\bar{\pi})] \le cK\overline{\log}(n)n^{1-\frac{\gamma(1+\alpha)}{2\gamma+2}}$ 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

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

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 $\mathcal{J}$, $\mathcal{J}_s$, $\mathcal{J}_w$ constitute a partition of the set of all P^d bins B_j, and we can thus write

$$\mathbb{E}(R_n(\bar{\pi})) = \sum_{j \in \mathcal{J}_s} \mathbb{E}(\tilde{R}_j(\bar{\pi})) + \sum_{j \in \mathcal{J}_w} \mathbb{E}(\tilde{R}_j(\bar{\pi})) + \sum_{j \in \mathcal{J}} \mathbb{E}(\tilde{R}_j(\bar{\pi})), \tag{64}$$

where, as in Equation (60), we define

$$\tilde{R}_j(\bar{\pi}) := \sum_{t=1}^{n} \Big[\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big) - \mathsf{T}\big(F^{\bar{\pi}_t(X_t)}(\cdot,X_t)\big) \Big] \mathbb{1}_{\{X_t \in B_j\}}. \tag{65}$$

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 B.3 yields

$$\begin{aligned}
\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{i}(\cdot,x)) &= \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{i}(\cdot,x)) - [\mathsf{T}(F^{\pi^\star(\bar{x})}(\cdot,\bar{x})) - \mathsf{T}(F^{i}(\cdot,\bar{x}))] \\
&\le 2CLd^{\gamma/2}P^{-\gamma} \le c_1 P^{-\gamma}.
\end{aligned} \tag{66}$$

First of all, this shows that $\mathcal{J}$ and $\mathcal{J}_s$ are disjoint. Furthermore, from Equations (65) and (66), we obtain

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

From Condition 3.11 we hence obtain:

$$\begin{aligned}\sum_{j\in\mathcal{J}_s}\mathbb{E}[\tilde{R}_j(\bar{\pi})] &\le c_1 nP^{-\gamma}\mathbb{P}_X\big(0<\mathsf{T}\big(F^{\pi^\star(X)}(\cdot,X)\big)-\mathsf{T}\big(F^{\pi^\sharp(X)}(\cdot,X)\le c_1P^{-\gamma}\big)\\ &\le C_0c_1^{1+\alpha}nP^{-\gamma(1+\alpha)}.\end{aligned}\tag{67}$$

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 3.11, recall that $\mathbb{P}(X_t\in B_j)\ge\frac{\underline{c}}{P^d}$, that

$$\begin{aligned}|\mathcal{J}_w|\frac{\underline{c}}{P^d}\le\sum_{j\in\mathcal{J}_w}\mathbb{P}(X_t\in B_j)&\le\mathbb{P}\big(0<\mathsf{T}\big(F^{\pi^\star(X_t)}(\cdot,X_t)\big)-\mathsf{T}\big(F^{\pi^\sharp(X_t)}(\cdot,X_t)\big)\le c_1P^{-\gamma}\big)\\ &\le C_0c_1^{\alpha}P^{-\gamma\alpha},\end{aligned}$$

which yields $|\mathcal{J}_w|\le(C_0c_1^\alpha/\underline{c})P^{d-\gamma\alpha}$. Using (61) and (62) 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 B.4.2)

$$\mathbb{E}[\tilde{R}_j(\bar{\pi})]\le c'\left(\sqrt{Kn\overline{\log}(n)}P^{-d/2}+nP^{-\gamma-d}\right),\tag{68}$$

where c' depends on $d,L,\gamma,\bar{c},C,\beta$, but *not* on n. Combining (68) with $|\mathcal{J}_w|\le(C_0c_1^\alpha/\underline{c})P^{d-\gamma\alpha}$ leads to

$$\sum_{j\in\mathcal{J}_w}\mathbb{E}[\tilde{R}_j(\bar{\pi})]\le c''\big(\sqrt{Kn\overline{\log}(n)}P^{d/2-\gamma\alpha}+nP^{-\gamma(1+\alpha)}\big),\tag{69}$$

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

$$\mathsf{T}(F^{\pi^\star(x_j)}(\cdot,x_j))-\mathsf{T}(F^{\pi^\sharp(x_j)}(\cdot,x_j))>c_1P^{-\gamma}.\tag{70}$$

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

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

Clearly $\pi^\star(x_j)\in I_j^\star$ and $\pi^\sharp(x_j)\in I_j^0$ (cf. (70)). 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 (65) as the sum of

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

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

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 B.3 yield

$$\begin{aligned}
0 &< \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot,x)) \\
&\leq \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x)) \\
&= \mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^{\pi^\star(x_j)}(\cdot,x_j)) + \mathsf{T}(F^i(\cdot,x_j)) - \mathsf{T}(F^i(\cdot,x)) \leq 2CLd^{\gamma/2}P^{-\gamma} \leq c_1 P^{-\gamma},
\end{aligned}$$

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 $x \in B_j$

$$\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x)) \leq c_1 P^{-\gamma} \mathbb{1}_{\{v: 0 < \mathsf{T}(F^{\pi^\star(v)}(\cdot,v)) - \mathsf{T}(F^{\pi^\sharp(v)}(\cdot,v)) \leq c_1 P^{-\gamma}\}}(x).$$

We deduce

$$\mathbb{E}[\tilde{R}_{j,I_j^\star}(\bar{\pi})] \leq \mathbb{E}\sum_{t=1}^{n} c_1 P^{-\gamma} \mathbb{1}_{\{0 < \mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t)) \leq c_1 P^{-\gamma}\}} \mathbb{1}_{\{X_t \in B_j\}} \leq n c_1 P^{-\gamma} q_j \tag{71}$$

where $q_j := \mathbb{P}(0 < \mathsf{T}(F^{\pi^\star(X_t)}(\cdot,X_t)) - \mathsf{T}(F^{\pi^\sharp(X_t)}(\cdot,X_t)) \leq c_1 P^{-\gamma}, X_t \in B_j)$, 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 B.3, 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 $F_{n,j}^i$ by F_j^i)

$$\mathsf{T}(F^{\pi^\star(x)}(\cdot,x)) - \mathsf{T}(F^i(\cdot,x)) \leq \left[\mathsf{T}(F_j^*) - \mathsf{T}(F_j^i)\right] + c_1 P^{-\gamma}, \tag{72}$$

from which it follows that

$$\mathbb{E}[\tilde{R}_{j,I_j^0}(\bar{\pi})] \leq \sum_{i \in I_j^0} \Delta_j^i \mathbb{E}S(i,n,j) + c_1 P^{-\gamma} \sum_{i \in I_j^0} \mathbb{E}S(i,n,j), \tag{73}$$

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)

$$\mathbb{E}S(i,n,j) \leq \frac{2C^2\beta \log(\bar{c}nP^{-d})}{[\Delta_j^i]^2} + \frac{\beta+2}{\beta-2}. \tag{74}$$

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 (72), and from using the definition of I_j^0. Next, noting that $\max_{i \in I_j^0} \Delta_j^i \leq 2C$ by Assumption 2.2, and combining Equations (73) and (74), we obtain the bound

$$\mathbb{E}[\tilde{R}_{j,I_j^0}(\bar{\pi})] \leq K \frac{2C^2\beta\log(\bar{c}nP^{-d})}{\underline{\Delta}_j}\left(1 + \frac{c_1 P^{-\gamma}}{\underline{\Delta}_j}\right) + (c_1 + 2C)K\frac{\beta+2}{\beta-2}. \tag{75}$$

It remains to prove the claim in Equation (74). To this end we apply a conditioning argument as in the proof of Theorem 3.6. We shall now use some quantities (in particular the sets $\Omega(v)$) that were defined in that proof: Note that

$$\mathbb{E}S(i,n,j) = \sum_{v \in \{0,1\}^n} \mathbb{P}(\Omega(v))\mathbb{E}(S(i,n,j)|\Omega(v)). \tag{76}$$

Arguing as in the proof of Theorem 3.6, it is now easy to see that $\mathbb{E}(S(i,n,j)|\Omega(v))$ can be written as the expected number of times arm 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 B.4). We can hence (cf. Remark 2.4) apply the bound established in Equation (49), to the just mentioned problem, to obtain

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

We can now combine the obtained inequality with Equation (76) to see that

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

The claim in (74) now follows from Jensen's inequality, and (cf. the end of the proof of Theorem 3.6) $\sum_{v\in\{0,1\}^n}\mathbb{P}(\Omega(v))\bar{m} \leq \bar{c}nP^{-d}$.

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 B.3 with $V_j = \sqrt{d}P^{-1}$ shows that $c_1P^{-\gamma}$ is smaller than

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

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

$$\mathsf{T}\big(F^{\pi^\star(x)}(\cdot,x)\big) - \mathsf{T}(F^i(\cdot,x)) > (1 + 2CLd^{\gamma/2})P^{-\gamma}. \tag{77}$$

[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 $x \in B_j$, (77) and Lemma B.3 (recalling that $\mathbb{P}_X(B_j) > \underline{c}P^{-d} > 0$) imply

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

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 (71) and (75) (with the just observed $\underline{\Delta}_j > P^{-\gamma}$) to see that for any $j \in \mathcal{J}$

$$\mathbb{E}[\tilde{R}_j(\bar{\pi})] \leq nc_1P^{-\gamma}q_j + \frac{2C^2(c_1+1)K\beta\log(\bar{c}nP^{-d})}{\underline{\Delta}_j} + (c_1+2C)K\frac{\beta+2}{\beta-2}. \tag{78}$$

Step 4d: A bound for $\sum_{j\in\mathcal{J}}\mathbb{E}[\tilde{R}_j(\bar{\pi})]$. Using Equation (78) and $|\mathcal{J}| \leq P^d$ we obtain

$$\sum_{j\in\mathcal{J}}\mathbb{E}[\tilde{R}_j(\bar{\pi})] \leq (c_1+2C)K\frac{\beta+2}{\beta-2}P^d + nc_1P^{-\gamma}\sum_{j\in\mathcal{J}}q_j + \sum_{j\in\mathcal{J}}\frac{2C^2(c_1+1)K\beta\log(\bar{c}nP^{-d})}{\underline{\Delta}_j}. \tag{79}$$

Since the B_j are disjoint, we obtain, recalling the definition of q_j after Equation (71), that

$$\frac{nc_1}{P^\gamma}\sum_{j\in\mathcal{J}} q_j \leq \frac{nc_1}{P^\gamma}\mathbb{P}\big(0 < \mathsf{T}(F^{\pi^\star(X_1)}(\cdot, X_1)) - \mathsf{T}(F^{\pi^\sharp(X_1)}(\cdot, X_1)) < c_1 P^{-\gamma}\big) \leq C_0 c_1^{1+\alpha} n P^{-\gamma(1+\alpha)}, \tag{80}$$

where we used Assumption 3.11 to obtain the last inequality.

To deal with the last sum in the upper bound in (79), we need a better lower bound on the $\underline{\Delta}_j$-s than the already available $P^{-\gamma}$. For notational simplicity, let's suppose that the well-behaved bins are indexed as $\mathcal{J} = \{1, 2, \ldots, j_1\}$ such that $0 < P^{-\gamma} \leq \underline{\Delta}_1 \leq \underline{\Delta}_2 \leq \ldots \leq \underline{\Delta}_{j_1}$. Fix $j \in \mathcal{J}$. Then, for any $k = 1, \ldots, j$, we claim that:

$$B_k \subseteq \left\{x : 0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot, x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot, x)) < \underline{\Delta}_j + 2CLd^{\gamma/2}P^{-\gamma}\right\}. \tag{81}$$

To see (81), 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 B.3 and B.4 and Remark 2.4 yield (the first inequality following from the observation after Equation (77))

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

and thus x is an element of the set on the right-hand-side of (81). Since all bins B_k are disjoint and $\underline{\Delta}_j + 2CLd^{\gamma/2}P^{-\gamma} \leq c_1\underline{\Delta}_j$ (obtained by recalling $c_1 = 4CLd^{\gamma/2}+1$, and using $\underline{\Delta}_j > P^{-\gamma}$), the inclusion (81) yields that for any $j \in \mathcal{J}$:

$$\mathbb{P}_X\big(x : 0 < \mathsf{T}(F^{\pi^\star(x)}(\cdot, x)) - \mathsf{T}(F^{\pi^\sharp(x)}(\cdot, x)) < c_1\underline{\Delta}_j\big) \geq \sum_{k=1}^{j}\mathbb{P}_X(B_k) \geq \frac{\underline{c}j}{P^d}. \tag{82}$$

Let's denote $j_2 := \max\{j \in \mathcal{J} : \underline{\Delta}_j \leq 1/c_1\}$ (here interpreting the maximum of an empty set as 0). Then, for each $j \in \{1, \ldots, j_2\}$ by Assumption 3.11 :

$$\mathbb{P}_X\big(0 < \mathsf{T}(F^{\pi^\star(X)}(\cdot, X)) - \mathsf{T}(F^{\pi^\sharp(X)}(\cdot, X)) < c_1\underline{\Delta}_j\big) \leq C_0(c_1\underline{\Delta}_j)^\alpha. \tag{83}$$

Combining (82), (83), and $\underline{\Delta}_j > P^{-\gamma}$, for any $j \in \{1, \ldots, j_2\}$ we get $\underline{\Delta}_j \geq \max\left(c_*\big(jP^{-d}\big)^{1/\alpha}, P^{-\gamma}\right)$, 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

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

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

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

for $c_{**} := [2 + c_*^{-1}(\alpha^{-1} - 1)^{-1}]$, where we used $\sum_{j=\tilde{P}+1}^{\infty} j^{-1/\alpha} \leq (\alpha^{-1} - 1)^{-1}\tilde{P}^{1-\alpha^{-1}}$ (cf. (53)). Hence, Equations (79) and (80), and the bounds in the previous two displays imply

$$\sum_{j \in \mathcal{J}} \mathbb{E}[\tilde{R}_j(\bar{\pi})] \leq c''' \left(nP^{-\gamma(1+\alpha)} + K\overline{\log}(nP^{-d})P^d + K\overline{\log}(nP^{-d})P^{d+\gamma(1-\alpha)} \right), \tag{84}$$

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

Step 5: Combining. From Equations (64), (67), (69) and (84) we obtain

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

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

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

from which the conclusion follows.

B.4.4 Proof of Theorem 3.13

To prove the theorem we just combine Theorem 3.12 and the following lemma, which allows one to upper bound the number of suboptimal assignments made by any policy.

Lemma B.5. *Suppose Assumptions 2.2 and 3.11 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, any randomization measure, and for all $(Y_t, X_t) \sim \mathbb{P}_{Y,X}$, such that $\mathbb{P}_{Y,X}$ satisfies Assumption 3.4, it holds that*

$$\mathbb{E}[R_n(\pi)] \geq \tilde{C}(\alpha, D_0, C_0) n^{-1/\alpha} \big(\mathbb{E}[S_n(\pi)]\big)^{1+1/\alpha} \quad \textit{for every } n \in \mathbb{N}. \tag{85}$$

Remark B.6. In Lemma B.5 we impose Assumptions 2.2 and 3.4 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 3.11.

Proof. 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 3.1. We write $\pi_{n,t} = \pi_t$. Let $\mathbb{P}_G$ be a randomization measure. We show that

$$\mathbb{E}[R_n(\pi)] \geq \tilde{C} n^{-1/\alpha} \big(\mathbb{E}[S_n(\pi)]\big)^{1+1/\alpha} \tag{86}$$

for $\tilde{C} = \tilde{C}(\alpha, D_0, C_0)$. If $\mathbb{E}[S_n(\pi)] = 0$, (86) trivially holds. Thus, suppose that $\mathbb{E}[S_n(\pi)] > 0$.

Note that for any $\delta > 0$,

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

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 3.11 yields

$$
\mathbb{E}[R_n(\pi)] \geq \delta(\mathbb{E}[S_n(\pi)] - C_0 n\delta^\alpha) = \delta(1 - 1/D_0)\mathbb{E}[S_n(\pi)] = \tilde{C} n^{-1/\alpha}\big(\mathbb{E}[S_n(\pi)]\big)^{1+1/\alpha},
$$

which proves (86). □

B.4.5 Proof of Theorem 3.14

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 (20) and (21) now proceeds in 5 steps:

Step 0: Preliminary observations and some notation. (a) From the maintained assumptions and Assumption 2.2 (imposed through Assumption 3.2) it follows that

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

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 the Appendix)

$$
\mathsf{KL}^{1/2}(\mu_{H_v}, \mu_{H_w}) \leq \zeta \left(\mathsf{T}(H_w) - \mathsf{T}(H_v)\right) \quad \text{ for every } v \leq w \text{ in } [-\varepsilon, \varepsilon]. \tag{87}
$$

By definition $\mathsf{T}(H_w) - \mathsf{T}(H_v) = h(w) - h(v)$. Hence, the statement in (87) 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 A.3 (and is established similarly as the last claim in Lemma A.4).

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 (18), 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 $j \in \{1, \ldots, P^d\}$ 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 $\|x\|_\infty := \max_{1 \leq i \leq d} |x_i|$ 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 $f_\sigma \in \mathcal{C}$

$$|f_\sigma(x_1) - f_\sigma(x_2)| \leq c_{-}\varepsilon 2^{-1} ||x_1 - x_2||^\gamma \quad \text{for every } x_1, x_2 \in [0,1]^d, \tag{88}$$

with $\| \cdot \|$ denoting the Euclidean norm. We note that for any pair $x_1, x_2 \in [0,1]^d$ one has $|\phi(x_1) - \phi(x_2)| \leq ||x_1 - x_2||_\infty^\gamma \leq ||x_1 - x_2||^\gamma$; the second inequality is obvious, and the first inequality follows from $|p^\gamma - q^\gamma| \leq |p - q|^\gamma$ for $p, q \geq 0$ and $0 < \gamma \leq 1$, together with the reverse triangle inequality. Now, to show (88), we consider two cases: First, if $x_1, x_2 \in B_j$ for $j \in \{1, \ldots, P^d\}$, the definition of f_σ and $|\phi(x_1) - \phi(x_2)| \leq ||x_1 - x_2||^\gamma$ lead to (note that if $j > m$, the following inequality trivially holds)

$$[c_{-}\varepsilon]^{-1} |f_\sigma(x_1) - f_\sigma(x_2)| \leq |\varphi_j(x_1) - \varphi_j(x_2)| \leq \frac{2^\gamma}{4} ||x_1 - x_2||^\gamma \leq \frac{1}{2} ||x_1 - x_2||^\gamma. \tag{89}$$

We remark that by continuity of f_σ, equation (89) continues to hold if x_1 and x_2 are elements of the closure of B_j, i.e., of $\bar{B}_j$. Secondly, suppose that $x_1 \in B_j, x_2 \in B_k$ for $j \neq k$. 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 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 $i = 1, \ldots, m$ and $\bar{\sigma}_i = 0$ for $i > m$. We obtain

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

where for the second inequality we made use of the second inequality in (89) (cf. also the remark immediately after (89)), and for the third inequality we combined $(a^\gamma + b^\gamma) \leq 2^{1-\gamma}(a + b)^\gamma$ for $0 < \gamma \leq 1$ and $a, b \geq 0$ with $||x_1 - y_1|| + ||y_2 - x_2|| \leq ||x_1 - y_1|| + ||y_1 - y_2|| + ||y_2 - x_2|| = ||x_1 - x_2||$. Since the hypercubes $B_1, \ldots, B_{P^d}$ define a partition of $[0,1]^d$ this establishes Equation (88).

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 $h(\varepsilon) - h(0) \geq c_-\varepsilon$ and similarly $h(0) - h(-\varepsilon) \geq c_-\varepsilon$. 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 (88) shows that A_f is Hölder-continuous with constant $\varepsilon/2$ and exponent γ. Note that by definition

$$f(x) = h\left(h^{-1} \circ f(x)\right) = h(A_f(x)) = \mathsf{T}\left(H_{A_f(x)}\right) \quad \text{for every } x \in [0,1]^d \text{ and every } f \in \mathcal{C}. \tag{90}$$

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 $\|H_v - H_w\|_\infty \leq |v - w|$ 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

$$\mathbb{P}_f := \mu_{H_0} \otimes [\mu_{H_{A_f(\cdot)}} \otimes \mathbb{P}_X]; \tag{91}$$

noting that the product in brackets is a semi-direct 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 (12) is satisfied as a consequence of Assumption 3.2. Now, for every $t = 1, \ldots, n$, denote by $\mathbb{P}^t_{\pi,f}$ the probability measure on the Borel sets of $\mathbb{R}^{(d+1)t}$ induced by the (recursively defined) random vector $Z_t = (Y_{\pi_t(X_t, Z_{t-1}, G_t),t}, X_t, \ldots, Y_{\pi_1(X_1,G_1),1}, X_1)$ 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 $\mathbb{R}^{(d+1)t}$ (i.e., a "realization" of the random vector Z_t).

We close this step with an important observation: For a vector $u = (u_1, \ldots, u_l)$ we denote $\Pi_-(u) = (u_1, \ldots, u_{l-1})$. Note that $\bar{K}_{t,f} : \mathcal{B}(\mathbb{R}) \times [0,1]^d \times \mathbb{R}^{(t-1)(d+1)} \times \mathbb{R}$ defined via

$$B \times x \times z_{t-1} \times g \mapsto \mu_{H_0}(B)\mathbb{1}\{\pi_t(x, z_{t-1}, g) = 1\} + \mu_{H_{A_f(x)}}(B)\mathbb{1}\{\pi_t(x, z_{t-1}, g) = 2\} \tag{92}$$

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

$$\mathbb{P}^t_{\pi,f} = \left(\bar{K}_{t,f} \otimes [\mathbb{P}_X \otimes \mathbb{P}^{t-1}_{\pi,f} \otimes \mathbb{P}_G]\right) \circ \Pi_-, \tag{93}$$

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 (92) (to see (93), note that the probability measure in parentheses in (93) is the joint distribution of Z_t and G_t; therefore the distribution of Z_t is obtained as the *image measure* of this joint distribution under the map Π_-). 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, Corollary 1.71 in Liese and Miescke (2008) and the Chain Rule of Lemma A.1 imply that for $f_1, f_2 \in \mathcal{C}$ and any $t = 1, \ldots, n$ we have

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

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

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

Using Equation (92) this sum further simplifies to

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

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

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

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

$$\mathsf{KL}\left(\mathbb{P}^{t}_{\pi,f_1}, \mathbb{P}^{t}_{\pi,f_2}\right) \le \int \sum_{i=1}^{t} \mathbb{1}\{\pi_i = 2\}\mathsf{KL}(\mu_{H_{A_{f_1}(x)}}, \mu_{H_{A_{f_2}(x)}})d(\mathbb{P}_X\otimes\mathbb{P}^{n}_{\pi,f_1}\otimes\mathbb{P}_G)(x,z_n,g). \quad (94)$$

Step 3: Verifying Assumptions 3.4 and 3.11 for every $\mathbb{P}_f$. Fix $f = f_\sigma \in \mathcal{C}$. To verify Assumption 3.4 (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 (91), note that

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

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 (90); 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 3.11 (with α and $C_0 = 8d[c_-\varepsilon]^{-\alpha}$ as given in the theorem), it suffices to show (recall that $K = 2$) that

$$\mathbb{P}_X\left(x\in[0,1]^d : 0 < |\mathsf{T}(H_{A_f(x)}) - \mathsf{T}(H_0)| \le c_-\varepsilon\delta\right) \le 8d\delta^\alpha \text{ for all } \delta \ge 0. \quad (95)$$

The statement in (95) is trivial for $\delta = 0$. Let $\delta > 0$. We use Equation (90) 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

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

which we can write as

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

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 (20) **and** (21)**.** We start with Equation (21). We already know that for every $f \in \mathcal{C}$ the measure $\mathbb{P}_f$ satisfies the inclusion in Equation (12) and Assumptions 3.4 and 3.11. It therefore suffices to verify

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

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 (90) 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 $\mathbb{P}_X \otimes \mathbb{P}^{t-1}_{\pi, f_\sigma} \otimes \mathbb{P}_G$. 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 t-th summand in the previous display depends on z_t only via z_{t-1}, we obtain

$$
\begin{aligned}
\sup_{f \in \mathcal{C}} \mathbb{E}_{(\mathbb{P}_f \otimes \mathbb{P}_G)^n}[S_n(\pi)] &= \sup_{\sigma \in \Sigma_m} \sum_{t=1}^n \mathbb{E}^{t-1}_{\pi, f_\sigma} \mathbb{E}_G \left[\mathbb{P}_X \left(x : \bar{f}_\sigma(x) \neq 0, \ \check{\pi}_t(x, z_{t-1}, g_t) \neq \operatorname{sign}(\bar{f}_\sigma(x)) \right) \right] \\
&\geq \sup_{\sigma \in \Sigma_m} \sum_{j=1}^m \sum_{t=1}^n \mathbb{E}^{t-1}_{\pi, f_\sigma} \mathbb{E}_G [\mathbb{P}_X (x \in B_j : \check{\pi}_t(x, z_{t-1}, g_t) \neq \sigma_j)] \\
&\geq \frac{1}{2^m} \sum_{j=1}^m \sum_{t=1}^n \sum_{\sigma \in \Sigma_m} \mathbb{E}^{t-1}_{\pi, f_\sigma} \mathbb{E}_G [\mathbb{P}_X (x \in B_j : \check{\pi}_t(x, z_{t-1}, g_t) \neq \sigma_j)], \qquad (96)
\end{aligned}
$$

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 (93)). For every $j \in \{1, \ldots, m\}$ and $t \in \{1, \ldots, n\}$,

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

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

$$
\frac{1}{P^d} \left(\mathbb{E}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}} \mathbb{E}_G \mathbb{E}^j_X \mathbb{1}_{\{\check{\pi}_t(x, z_{t-1}, g) = 1\}} + 1 - \mathbb{E}^{t-1}_{\pi, f_{\sigma^1_{-j}}} \mathbb{E}_G \mathbb{E}^j_X \mathbb{1}_{\{\check{\pi}_t(x, z_{t-1}, g) = 1\}} \right) =: \frac{1}{P^d} e(\sigma, j, t). \qquad (97)
$$

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

$$
\begin{aligned}
e(\sigma, j, t) &\geq \frac{1}{4} \exp \left[-\mathsf{KL} \left(\mathbb{P}^j_X \otimes \mathbb{P}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}} \otimes \mathbb{P}_G, \mathbb{P}^j_X \otimes \mathbb{P}^{t-1}_{\pi, f_{\sigma^1_{-j}}} \otimes \mathbb{P}_G \right) \right] \\
&= \frac{1}{4} \exp \left[-\mathsf{KL} \left(\mathbb{P}^{t-1}_{\pi, f_{\sigma^{-1}_{-j}}}, \mathbb{P}^{t-1}_{\pi, f_{\sigma^1_{-j}}} \right) \right],
\end{aligned} \qquad (98)
$$

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 (94) 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 that $(f_1(x), f_2(x)) = (h(0) - c_-\varepsilon\varphi_j(x), h(0) + c_-\varepsilon\varphi_j(x))$ for $x \in B_j$, from which it follows from Equations (87) (note that $A_{f_1(x)} \leq A_{f_2(x)}$ follows from strict monotonicity of h^{-1}, cf. Step 0) and (90) that

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

Since $[2\zeta c_-\varepsilon\varphi_j(x)]^2 \leq [\zeta c_-\varepsilon 2^{-1}P^{-\gamma}]^2 =: \bar{r}P^{-2\gamma}$ holds for $x \in B_j$, Equation (94) delivers

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

with $G(i,j) := \{(x, z_n, g) : x \in B_j, \pi_i(x, z_{i-1}, g) = 2\}$, $N_{j,\sigma_{-j}} := \int \sum_{i=1}^n \mathbb{1}\{G(i,j)\} d(\mathbb{P}_X \otimes \mathbb{P}^n_{\pi,f_{\sigma^{-1}_{-j}}} \otimes \mathbb{P}_G)$. The dependence of $N_{j,\sigma_{-j}}$ on π has been suppressed. In combination with Equations (97) and (98) we hence obtain

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

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

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

Combining the lower bounds in the previous two displays with (96) yields

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

which can further be lower-bounded by

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

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 $w\exp(-\varrho)+\varrho\geq w$ for every $\varrho\geq 0$ and every $0<w\leq 1$ 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

$$\frac{n^{1-\frac{\alpha\gamma}{d+2\gamma}}}{16}2^{-\gamma\alpha}\geq n^{1-\frac{\alpha\gamma}{d+2\gamma}}/32.$$

This establishes Equation (21). Finally, Lemma B.5 (cf. Step 3, which verifies the assumptions needed) with $D_0=2+C_0^{-1}$ shows that the lower bound established in Lemma B.5 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 B.5 and the already established Equation (21) proves Equation (20).

B.4.6 Proof of Theorem 3.3

Because Assumption 3.4 (for any $\gamma\in(0,1]$ and any $L>0$) implies the assumption in Equation (14), the statement follows immediately from the lower bound in Equation (20) in Theorem 3.14 upon letting $\gamma\to 0$.

B.4.7 Proof of Theorem 3.8

The statement follows from the first lower bound established in Theorem 3.14, 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.

B.4.8 Proof of Theorem 3.5

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., π is as defined in Equation (1). 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 2.2 and 3.4 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 (15), note that by definition, and since π is a policy that does not depend on covariates,

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

Note furthermore that

$$\begin{aligned} &\big[\{X_t \in A_{1,\bar{m}}\} \cap \{\pi_t(Z_{t-1}, G_t) \neq 1\}\big] \cup \big[\{X_t \in A_{2,\bar{m}}\} \cap \{\pi_t(Z_{t-1}, G_t) \neq 2\}\big] \\ \subseteq\ &\{\pi^\star(X_t) \neq \pi_t(Z_{t-1}, G_t)\}. \end{aligned}$$

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

$$R_n(\pi) \geq \bar{m}^{-1} \sum_{t=1}^n \left(\mathbb{1}_{A_{1,\bar{m}}}(X_t)\mathbb{1}_{\{\pi_t(Z_{t-1},G_t)\neq 1\}} + \mathbb{1}_{A_{2,\bar{m}}}(X_t)\mathbb{1}_{\{\pi_t(Z_{t-1},G_t)\neq 2\}}\right).$$

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

C Proofs of results in Section 4

Proof of Lemma 4.1: Given $F, G \in D_{cdf}([a, b])$ it holds that $|\mathsf{S}_{\text{abs}}(F) - \mathsf{S}_{\text{abs}}(G)|$ is not greater than 1/2-times

$$\int_{[a,b]} \big||x - \mu(F)| - |x - \mu(G)|\big|\, dF(x) + \left|\int_{[a,b]} |x - \mu(G)| dF(x) - \int_{[a,b]} |x - \mu(G)| dG(x)\right|.$$

Using the reverse triangle inequality, the first integral in the previous display can be bounded from above by $|\mu(F) - \mu(G)| \leq (b - a)\|F - G\|_\infty$ (cf. Example D.3 for the inequality). Using Lemma D.2, the remaining expression to the right in the previous display is seen not to be greater than $(b - a)\|F - G\|_\infty$. Hence, the first statement follows (noting that S_{abs} is obviously well defined on all of $D_{cdf}([a, b])$).

Concerning the second claim, we first observe that for every $F \in D_{cdf}([a, b])$ it holds that

$$\frac{1}{2} \int_{[a,b]} |x - \mu(F)| dF(x) = \int_{[a,\mu(F)]} (\mu(F) - x) dF(x). \tag{99}$$

Next, let $s > 0$, $\delta \in (a,b)$, $F \in \mathscr{D}(s,\delta)$ and $G \in D_{cdf}([a,b])$. We consider two cases, and start with the case where $\mu(G) = 0$ (implying that $a = 0$ and that G is the cdf corresponding to point mass at 0). Then, by convention, $\mathsf{S}_{\mathrm{rel}}(G) = 0$, and it follows from Equation (99) (recalling that $\mu(F) \geq \delta > 0$) that

$$|\mathsf{S}_{\mathrm{rel}}(F) - \mathsf{S}_{\mathrm{rel}}(G)| \leq \int_{[a,\mu(F)]} |1 - x/\mu(F)| dF(x) \leq F(\mu(F)).$$

Since F is continuous $0 = F(0) = F(\mu(G))$ holds. It follows that $F(\mu(F)) = |F(\mu(F)) - F(\mu(G))|$. Using the mean-value theorem of Minassian (2007) and Example D.3 we conclude that $|F(\mu(F)) - F(\mu(G))| \leq s(b-a)\|F - G\|_\infty$.

Next, we turn to the case where $\mu(G) > 0$. First, we note that

$$|\mathsf{S}_{\mathrm{rel}}(F) - \mathsf{S}_{\mathrm{rel}}(G)| \leq |F(\mu(F)) - G(\mu(G))| + \left| \int_{[a,\mu(F)]} \frac{x}{\mu(F)} dF(x) - \int_{[a,\mu(G)]} \frac{x}{\mu(G)} dG(x) \right|.$$

Consider the first term in absolute values in the previous display: By the triangle inequality:

$$|F(\mu(F)) - G(\mu(G))| \leq |F(\mu(F)) - F(\mu(G))| + \|F - G\|_\infty.$$

From the mean-value theorem for right-differentiable functions as in Minassian (2007), and the definition of $\mathscr{C}^s([a,b])$, we obtain $|F(\mu(F)) - F(\mu(G))| \leq s|\mu(F) - \mu(G)| \leq s(b-a)\|F - G\|_\infty$, the second inequality following from Example D.3. Now, note that (incorporating the considerations in case $\mu(G) = 0$) it remains to show that

$$\left| \int_{[a,\mu(F)]} \frac{x}{\mu(F)} dF(x) - \int_{[a,\mu(G)]} \frac{x}{\mu(G)} dG(x) \right| \leq ((s + \delta^{-1})(b-a) + 4)\|F - G\|_\infty. \tag{100}$$

To this end, denote $m := \min(\mu(F), \mu(G))$, $M := \max(\mu(F), \mu(G))$, let $\tilde{F}$ denote a cdf in $\{F, G\}$ which realizes the latter maximum, and rewrite the difference of integrals inside the absolute value to the left in the preceding display as

$$\int_{[a,m]} \frac{x}{\mu(F)} dF(x) - \int_{[a,m]} \frac{x}{\mu(F)} dG(x) \pm \int_{(m,M]} \frac{x}{\mu(\tilde{F})} d\tilde{F}(x) + \int_{[a,m]} \left[\frac{x}{\mu(F)} - \frac{x}{\mu(G)} \right] dG(x),$$

where "$\pm$" is to be interpreted as "$+$" in case $\tilde{F} = F$ and as "$-$" in case $\tilde{F} = G$. Next, denote the difference of the first two integrals in the previous display by A, the third integral by B and the fourth by D, respectively. First, Lemma D.2 (applied with $k = 1$, $c = a$, $d = m$ and $\varphi(x) = x/\mu(F)$) implies (working with the upper bounds $|M^*| \leq 1$ and $C \leq 1$ in Lemma D.2 for the special case under consideration) that $|A| \leq 2\|F - G\|_\infty$. Second, note that the integrand in B is smaller than 1, hence

$$|B| \leq \tilde{F}(M) - \tilde{F}(m) \leq F(M) - F(m) + 2\|F - G\|_\infty \leq s|\mu(F) - \mu(G)| + 2\|F - G\|_\infty$$

where we used $\|\tilde{F} - F\| \leq \|F - G\|_\infty$ for the first inequality, and the mean-value theorem of Minassian (2007) for the second. To obtain an upper bound for $|B|$ we now use Example D.3

to see that the right hand side in the previous display is not greater than $[s(b-a)+2]\|F-G\|_\infty$. Concerning $|D|$ note that (cf. Example D.3)

$$|D| \leq \int_{[a,\mu(G)]} \left| \frac{\mu(G)}{\mu(F)} - 1 \right| dG(x) \leq \left| \frac{\mu(G)}{\mu(F)} - 1 \right| \leq \delta^{-1}|\mu(G) - \mu(F)| \leq \delta^{-1}(b-a)\|F-G\|_\infty.$$

Summarizing,

$$|A| + |B| + |D| \leq ((s+\delta^{-1})(b-a)+4)\|F-G\|_\infty,$$

which proves the statement in Equation (100). □

Proof of Lemma 4.2: The first statement follows from Example D.6. To prove the statement concerning G_{rel}, we first note that G_{rel} is well defined on $D_{cdf}([a,b])$ (note that $\mu(F) \leq 0$ implies that $a = 0$ and that μ_F is point mass at 0, implying that $\mathsf{G}_{\text{rel}}(F) = 0$). Let $F, G \in D_{cdf}([a,b])$, and assume $\mu(F) \geq \delta$, where $\delta \in (a,b)$. Consider first the case where $\mu(G) = 0$. Then, $\mathsf{G}_{\text{rel}}(G) = 0$ and

$$|\mathsf{G}_{\text{rel}}(F) - \mathsf{G}_{\text{rel}}(G)| = \mathsf{G}_{\text{rel}}(F) \leq \delta^{-1}[\mu(F) - \mu(G)] \leq \delta^{-1}(b-a)\|F-G\|_\infty,$$

where we used that $\mathsf{G}_{\text{abs}}(F) \leq \mu(F)$ (just note that $|x_1 - x_2| = (x_1 + x_2) - 2\min(x_1, x_2)$), and Example D.3.

If, on the other hand, $\mu(G) > 0$ (recall that $a \geq 0$), we abbreviate $\varphi(x_1, x_2) = |x_1 - x_2|$, and write

$$|\mathsf{G}_{\text{rel}}(F) - \mathsf{G}_{\text{rel}}(G)| \leq (A+B)/2,$$

where

$$A := \delta^{-1} \left| \int_{[a,b]} \int_{[a,b]} \varphi(x_1,x_2) dF(x_1) dF(x_2) - \int_{[a,b]} \int_{[a,b]} \varphi(x_1,x_2) dG(x_1) dG(x_2) \right|,$$

which, by Example D.6 is not greater than $\delta^{-1}2(b-a)\|F-G\|_\infty$, and

$$B := \int_{[a,b]} \int_{[a,b]} |(\mu(F)^{-1} - \mu(G)^{-1})\varphi(x_1,x_2)| dG(x_1) dG(x_2) \leq 2 \left| [\mu(G)/\mu(F)] - 1 \right|,$$

where we used $\mathsf{G}_{\text{abs}}(G) \leq \mu(G)$. Note that $\left|[\mu(G)/\mu(F)] - 1\right| \leq \delta^{-1}(b-a)\|F-G\|_\infty$ (cf. the end of the proof of Lemma 4.1). Hence, in case $\mu(G) \neq 0$, we obtain that

$$|\mathsf{G}_{\text{rel}}(F) - \mathsf{G}_{\text{rel}}(G)| \leq 2\delta^{-1}(b-a)\|F-G\|_\infty.$$

Together with the first case, this proves the result.

□

Proof of Lemma 4.3: The functional in Equation (26) is well defined on $D_{cdf}([a,b])$, because of Lemma D.13, and since $a > 0$ is assumed. Next, we apply Lemma D.12 together with Lemma D.14 to obtain that for every $u \in [0,1]$ the functional $F \mapsto L(F,u)$ satisfies Assumption 2.2 with a, b and $\mathscr{D}$ (as in the statement of the present lemma) and with constant $a^{-1}(r^{-1} + (b-a)a^{-1}b)$. The statement immediately follows. □

Proof of Lemma 4.4: Arguing similarly as in the proof of Lemma 4.3, the triangle inequality, together with Example D.3 and Lemma D.14 (which is applicable due to Lemma D.12) immediately yield the claimed result. $\square$

Proof of Lemma 4.5: We start with the first statement. The functional T is obviously everywhere defined on $D_{cdf}([a,b])$ (in case $\mu(F)=0$ it follows that $a=0$ and that F corresponds to point mass 1 at 0 in which case $\mathsf{T}(F)=0$, by definition). Next, let $\delta \in (a,b)$, let $F \in \mathscr{D}(\delta)$ and let $G \in D_{cdf}([a,b])$. We consider first the case where $\mu(G)=0$ (implying that $\mathsf{T}(G)=0$ and $a=0$). We conclude that $|\mathsf{T}(F)-\mathsf{T}(G)| = \mathsf{T}(F)$, the latter being not greater than

$$\frac{1}{c|c-1|\delta^c}\left|\int_{[a,b]} x^c dF(x) - \int_{[a,b]} x^c dG(x)\right| \leq \frac{b^c-a^c}{c|c-1|\delta^c}\|F-G\|_\infty,$$

where we used Example D.4 (recall that $a=0$) for the last inequality. Next, consider the case where $\mu(G)>0$. We note that

$$\left|\int_{[a,b]} (x/\mu(F))^c dF(x) - \int_{[a,b]} (x/\mu(G))^c dG(x)\right| \tag{101}$$

can be upper bounded by $A+B$ with

$$A := \left|\int_{[a,b]} (x/\mu(F))^c dF(x) - \int_{[a,b]} (x/\mu(F))^c dG(x)\right| \leq \frac{b^c-a^c}{\delta^c}\|F-G\|_\infty$$

the inequality following from Lemma D.2, and

$$B := |(1/\mu(F))^c - (1/\mu(G))^c| \int_{[a,b]} x^c dG(x) \leq |(\mu(G)/\mu(F))^c - 1|,$$

the inequality following from Jensen's inequality (recalling that $c \in (0,1)$). It remains to observe that the simple inequality $|z^c-1| \leq |z-1|$ for $z>0$ implies

$$|(\mu(G)/\mu(F))^c - 1| \leq |\mu(G)/\mu(F) - 1| \leq \delta^{-1}(b-a)\|F-G\|_\infty,$$

where the second inequality follows from Example D.3 together with $\mu(F) \geq \delta$. Hence, in case $\mu(G)>0$ we see that

$$|\mathsf{T}(F)-\mathsf{T}(G)| \leq (c|c-1|)^{-1}\left[\frac{b^c-a^c}{\delta^c} + \delta^{-1}(b-a)\right]\|F-G\|_\infty,$$

which proves the first claim.

We now prove the second claim. Since $a>0$ holds in this case, $\mu(G)$ and $\mu(F)$ can not be smaller than a. Hence the functional is well defined on all of $D_{cdf}([a,b])$. Furthermore, the expression in Equation (101) is not greater than $A+B$, where A and B have been defined above. By Lemma D.2 it holds that A is not greater than $a^{-c}|b^c-a^c|\|F-G\|_\infty$. Furthermore, B is not greater than

$$\begin{aligned}\max\left((a/b)^c,(b/a)^c\right)|(\mu(F)/\mu(G))^c-1| &\leq |c|\max\left((a/b)^{2c-1},(b/a)^{2c-1}\right)|\mu(F)/\mu(G)-1| \\ &\leq |c|\max\left((a/b)^{2c-1},(b/a)^{2c-1}\right)a^{-1}(b-a)\|F-G\|_\infty,\end{aligned}$$

the first inequality following from $|z^c - 1| \leq |c| \max((a/b)^{c-1}, (b/a)^{c-1})|z-1|$ for $z \in [a/b, b/a]$ (noting that this interval contains 1 and recalling that $c \notin [0,1]$), and the second inequality following from Exercise D.3.

We now turn to the last case where $c \in \{0,1\}$ (and $a > 0$ guaranteeing that the functional is then well defined on all of $D_{cdf}([a,b])$). We consider first the case where $c = 0$. Without loss of generality, assume $\mu(F) \leq \mu(G)$. The statement follows after noting that $|\mathsf{T}(F) - \mathsf{T}(G)|$ is not greater than $C + D$ with

$$C := \left| \int_{[a,b]} \log(x) dF(x) - \int_{[a,b]} \log(x) dG(x) \right| \leq \log(b/a) \|F - G\|_\infty,$$

the inequality following from Lemma D.2, and (using Example D.3)

$$D := \log(\mu(G)/\mu(F)) \leq \log(1 + \frac{b-a}{a} \|F - G\|_\infty) \leq \frac{b-a}{a} \|F - G\|_\infty. \tag{102}$$

In case $c = 1$, set $f(x) := (x/\mu(F)) \log(x/\mu(F))$ and $g(x) := (x/\mu(G)) \log(x/\mu(G))$. Write

$$|\mathsf{T}(F) - \mathsf{T}(G)| \leq \left| \int_{[a,b]} f(x) dF(x) - \int_{[a,b]} f(x) dG(x) \right| + \int_{[a,b]} |f(x) - g(x)| dG(x).$$

From Lemma D.2 it follows that the first absolute value in the upper bound is not greater than $\|F - G\|_\infty$ times the total variation of f on $[a,b]$, the latter being bounded from above by $\int_{[a/b, b/a]} |1 + \log(x)| dx$. Finally, noting that for every $x \in [a,b]$ we have

$$\begin{aligned} |f(x) - g(x)| &\leq |x| \left\{ |\mu^{-1}(F) - \mu^{-1}(G)| |\log(x/\mu(F))| + \mu^{-1}(G) \left| \log(\mu(F)/\mu(G)) \right| \right\} \\ &\leq \frac{b(b-a)}{a^2} \left\{ \log(b/a) + 1 \right\} \|F - G\|_\infty, \end{aligned}$$

where (in addition to $a > 0$) we used Example D.3 and (102). The final claim follows. $\square$

Proof of Lemma 4.6: We start with Part 1: Let $\delta \in (a,b)$. From the first part of Lemma 4.5 we obtain that in case $\varepsilon \in (0,1)$ the functional $\varepsilon(\varepsilon - 1)\mathsf{E}_{c(\varepsilon)} + 1$ satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{D}(\delta)$ and constant $[\delta^{-c(\varepsilon)}(b^{c(\varepsilon)} - a^{c(\varepsilon)}) + \delta^{-1}(b-a)]$. It remains to observe that the function

$$z \mapsto 1 - z^{1/c(\varepsilon)} \tag{103}$$

is Lipschitz continuous on $[0,1]$ with constant $c(\varepsilon)^{-1}$. The claim then follows from Lemma D.1 (with $m = 1$), and the representation in Equation (28) together with the observation that $0 \leq \varepsilon(\varepsilon - 1)\mathsf{E}_{c(\varepsilon)}(F) + 1 \leq 1$ holds for every $F \in D_{cdf}([a,b])$ as a consequence of Jensen's inequality.

For Part 2 we argue similarly as in Part 1. From the second part of Lemma 4.5 we obtain that in case $\varepsilon \in (1, \infty)$ the functional $\varepsilon(\varepsilon - 1)\mathsf{E}_{c(\varepsilon)} + 1$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$ and constant (note that $2c(\varepsilon) - 1 < 0$) equal to

$$[a^{-c(\varepsilon)}(a^{c(\varepsilon)} - b^{c(\varepsilon)}) + c(\varepsilon)(a/b)^{2c(\varepsilon)-1} a^{-1}(b-a)].$$

The function in Equation (103) is Lipschitz continuous on $[(b/a)^{c(\varepsilon)}, (a/b)^{c(\varepsilon)}]$ with constant $(\varepsilon - 1)^{-1}(b/a)^\varepsilon$. From Equation (28), and because $(b/a)^{c(\varepsilon)} \leq \varepsilon(\varepsilon - 1)\mathsf{E}_{1-\varepsilon}(F) + 1 \leq (a/b)^{c(\varepsilon)}$ trivially holds for every $F \in D_{cdf}([a,b])$ ($a > 0$ and $c(\varepsilon) < 0$), the claim follows from Lemma D.1 (with $m = 1$). $\square$

Proof of Lemma 4.7: Clearly T is well defined on all of $D_{cdf}([a,b])$. Let $F \in D_{cdf}([a,b])$. Then, by Jensen's inequality:

$$\int_{[a,b]} e^{\kappa[\mu(F)-x]} dF(x) \geq 1.$$

Since $x \mapsto \log(x)$ restricted to $[1,\infty)$ is Lipschitz continuous with constant 1, we obtain for any $G \in D_{cdf}([a,b])$ that $\kappa|\mathsf{T}(F) - \mathsf{T}(G)|$ is bounded from above by

$$\begin{aligned}
&\left|\int_{[a,b]} e^{\kappa[\mu(F)-x]} dF(x) - \int_{[a,b]} e^{\kappa[\mu(G)-x]} dG(x)\right| \\
= \; &\left|(e^{\kappa\mu(F)} - e^{\kappa\mu(G)}) \int_{[a,b]} e^{-\kappa x} dF(x) + e^{\kappa\mu(G)} \Big(\int_{[a,b]} e^{-\kappa x} dF(x) - \int_{[a,b]} e^{-\kappa x} dG(x)\Big)\right| \\
\leq \; &\left|e^{\kappa\mu(F)} - e^{\kappa\mu(G)}\right| e^{-\kappa a} + e^{\kappa b}\left|\int_{[a,b]} e^{-\kappa x} dF(x) - \int_{[a,b]} e^{-\kappa x} dG(x)\right| \\
\leq \; &\Big(\kappa e^{\kappa(b-a)}[b-a] + e^{\kappa b}[e^{-\kappa a} - e^{-\kappa b}]\Big) \|F - G\|_\infty ,
\end{aligned}$$

where we used Example D.3 and Lemma D.2 in bounding in each of the summands on the left hand side of the last inequality (as well as the mean-value theorem for the first summand). □

Proof of Lemma 4.9: Obviously, the welfare function W is well defined on $D_{cdf}([a,b])$ in both parts of the lemma. The first statement follows from the assumptions and Example D.3, noting that $x_1x_2 - y_1y_2 = (x_1 - y_1)x_2 - y_1(y_2 - x_2)$ holds for real numbers x_i, y_i, $i = 1, 2$. The second statement follows directly from the assumptions and Example D.3. □

Lemma C.1. *Let $a < b$ be real numbers, $z_0 > 0$ and $0 \leq \delta \leq 1$. Then, the following holds:*

1. *If $\delta = 0$, then $\mathsf{z}_{\mathsf{m},z_0,\delta}$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$, and any $C > 0$.*

2. *If $\delta > 0$ and $\mathsf{m} \equiv \mu(\cdot)$, then $\mathsf{z}_{\mathsf{m},z_0,\delta}$ satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a,b])$ and $C = \delta(b-a)$.*

3. *If $\delta > 0$ and $\mathsf{m} \equiv q_{1/2}(\cdot)$, then, for every $r > 0$, the poverty line $\mathsf{z}_{\mathsf{m},z_0,\delta}$ satisfies Assumption 2.2 with $\mathscr{D} = \mathscr{C}_r([a,b])$, and $C = r^{-1}\delta$.*

Proof of Lemma C.1: By definition $\mathsf{z}_{\mathsf{m},z_0,\delta}(F) = z_0 + \delta(\mathsf{m}(F) - z_0)$. The first statement is trivial; the second follows directly from Example D.3; and the third follows from Lemma D.12 and Example D.10. □

Proof of Lemma 4.10: Since z satisfies Assumption 2.2 the functional z is well defined on all of $D_{cdf}([a,b])$. Thus H_z is well defined on $D_{cdf}([a,b])$ as well. Finally, given $F \in \mathscr{D}$ and $G \in D_{cdf}([a,b])$, note that by definition and the triangle inequality:

$$|\mathsf{H}_\mathsf{z}(F) - \mathsf{H}_\mathsf{z}(G)| \leq |F(\mathsf{z}(F)) - F(\mathsf{z}(G))| + \|F - G\|_\infty \leq (C_\mathsf{z} s + 1)\|F - G\|_\infty,$$

where we used that z satisfies Assumption 2.2 together with a mean-value theorem as in Minassian (2007) for the last inequality. □

Proof of Lemma 4.11: Obviously, $\mathsf{P}_{SK}(\cdot;\mathsf{z},\kappa)$ is well defined on $D_{cdf}([a,b])$ because $\mathsf{z} \geq z_* > 0$ holds by assumption, and due to our convention that $0/0 := 0$ (noting also that $F(x) = 0$ for every $x \in [0, \mathsf{z}(F)]$ in case $F(\mathsf{z}(F)) = 0$). Next, fix $F \in \mathscr{D}$ and $G \in D_{cdf}([a,b])$. Define for all $x \in \mathbb{R}$

$$f(x) := \max(1 - [x/\mathsf{z}(F)], 0)|1 - [F(x)/F(\mathsf{z}(F))]|^{\kappa},$$

and analogously

$$g(x) := \max(1 - [x/\mathsf{z}(G)], 0)|1 - [G(x)/G(\mathsf{z}(G))]|^{\kappa}.$$

Define $m := \min(\mathsf{z}(F), \mathsf{z}(G))$ and $M := \max(\mathsf{z}(F), \mathsf{z}(G))$, and the following partition of $[a, M]$ (using our convention $0/0 := 0$):

$$A := \left\{x \in [a,m] : \frac{F(x)}{F(\mathsf{z}(F))} > \frac{G(x)}{G(\mathsf{z}(G))}\right\}, \quad B := [a,m]\backslash A, \quad \text{and } D := (m, M],$$

where $D = \emptyset$ in case $m = M$. Next, write

$$\frac{\mathsf{P}_{SK}(F;\mathsf{z},\kappa) - \mathsf{P}_{SK}(G;\mathsf{z},\kappa)}{\kappa+1} = \int_{[a,M]} [f(x)-g(x)]dF(x) + \left[\int_{[a,b]} g(x)dF(x) - \int_{[a,b]} g(x)dG(x)\right],$$

noting that f and g vanish for $x > M$; and denote the right-hand side by $S_1 + S_2$, S_2 denoting the term in brackets to the far right. Since $g([a,b]) \subseteq [0,1]$ and because g is right-continuous (G is a cdf) and non-increasing, it hence follows from Lemma D.7 that $|S_2| \leq \|F - G\|_\infty$.

Concerning S_1, note that for every $x \in [a, m]$ it holds that $|f(x) - g(x)|$ is not greater than the sum of

$$\begin{aligned}|\max([1 - \frac{x}{\mathsf{z}(F)}], 0) - \max([1 - \frac{x}{\mathsf{z}(G)}], 0)| &\leq x|\mathsf{z}(F)^{-1} - \mathsf{z}(G)^{-1}| \\ &\leq x z_*^{-2}|\mathsf{z}(F) - \mathsf{z}(G)| \leq b z_*^{-2} C_{\mathsf{z}} \|F - G\|_\infty,\end{aligned}$$

(where we used that $\mathsf{z} \geq z_*$ to obtain the second inequality, and that z satisfies Assumption 2.2 to obtain the third) and

$$\big||1 - [F(x)/F(\mathsf{z}(F))]|^{\kappa} - |1 - [G(x)/G(\mathsf{z}(G))]|^{\kappa}\big| \leq \kappa|[F(x)/F(\mathsf{z}(F))] - [G(x)/G(\mathsf{z}(G))]|$$

(where we used $\kappa \geq 1$, the mean-value theorem, and the reverse triangle inequality to obtain the upper bound). For $x \in D$, it holds that $|f(x) - g(x)| \leq 1$. It hence follows that $|S_1|$ is bounded from above by the sum of $b z_*^{-2} C_{\mathbf{z}} \|F - G\|_\infty$, κ times

$$\int_A [F(x)/F(\mathsf{z}(F))] - [G(x)/G(\mathsf{z}(G))]dF(x) + \int_B [G(x)/G(\mathsf{z}(G))] - [F(x)/F(\mathsf{z}(F))]dF(x) \tag{104}$$

and $\int_D dF(x)$. From the mean-value theorem in Minassian (2007), and z satisfying Assumption 2.2, we conclude

$$\int_D dF(x) \leq F(M) - F(m) \leq s C_{\mathsf{z}} \|F - G\|_\infty.$$

Before bounding the integrals in Equation (104), we recall that Lemma 4.10 shows that

$$G(\mathsf{z}(G)) - (C_{\mathsf{z}}s+1)\|F-G\|_\infty \leq F(\mathsf{z}(F)) \leq G(\mathsf{z}(G)) + (C_{\mathsf{z}}s+1)\|F-G\|_\infty. \tag{105}$$

To bound the integrals in Equation (104), we now consider different cases:

Consider first the case where $F(\mathsf{z}(F)) = 0$: Then, the convention $0/0 := 0$ implies $A = \emptyset$ and $B = [a, m]$. Furthermore, the integral over B in Equation (104) vanishes in this case, because $m \leq \mathsf{z}(F)$ implies $F(m) = 0$ (and the integrand is non-negative). Hence, the expression in Equation (104) is 0.

Next, consider the case where $G(\mathsf{z}(G)) = 0$ and $F(\mathsf{z}(F)) > 0$. It follows from our convention that then $A = \{x \in [a,m] : F(x)/F(\mathsf{z}(F)) > 0\}$, and that the integral over B in (104) vanishes. The integral over A is not greater than

$$F(m) = F(m) - G(\mathsf{z}(G)) \leq F(\mathsf{z}(F)) - G(\mathsf{z}(G)) \leq (C_{\mathsf{z}}s+1)\|F-G\|_\infty,$$

where we used Equation (105) to obtain the last inequality. We thus see that in this case the expression in Equation (104) does not exceed $(C_{\mathsf{z}}s+1)\|F-G\|_\infty$.

Finally, consider the case where $G(\mathsf{z}(G))$ and $F(\mathsf{z}(F))$ are both positive. Then, we can write the integral over A in Equation (104) as

$$\begin{aligned}
& F(\mathsf{z}(F))^{-1}\int_A F(x) - F(\mathsf{z}(F))\frac{G(x)}{G(\mathsf{z}(G))}dF(x) \\
\leq\ & F(\mathsf{z}(F))^{-1}\int_A [F(x)-G(x)] + (C_{\mathsf{z}}s+1)\|F-G\|_\infty \frac{G(x)}{G(\mathsf{z}(G))}dF(x) \\
\leq\ & F(\mathsf{z}(F))^{-1}\int_A dF(x)(C_{\mathsf{z}}s+2)\|F-G\|_\infty \leq (C_{\mathsf{z}}s+2)\|F-G\|_\infty,
\end{aligned}$$

where we used Equation (105) to obtain the first inequality. Similarly, the integral over B in Equation (104) can be shown not to be greater than $(C_{\mathsf{z}}s+2)\|F-G\|_\infty$. Summarizing, in this last case the expression in Equation (104) does not exceed $[2C_{\mathsf{z}}s+4]\|F-G\|_\infty$. In particular, this bound is bigger than the two bounds in the other two cases. Hence, we conclude that the expression in Equation (104) is not greater than $[2C_{\mathsf{z}}s+4]\|F-G\|_\infty$.

It follows that $|S_1|$ is bounded from above by

$$[(bz_*^{-2} + 2\kappa s + s)C_{\mathsf{z}} + 4\kappa]\|F-G\|_\infty.$$

Recalling $|S_2| \leq \|F-G\|_\infty$, it follows that

$$\frac{\mathsf{P}_{SK}(F;\mathsf{z},\kappa) - \mathsf{P}_{SK}(G;\mathsf{z},\kappa)}{\kappa+1} \leq |S_1| + |S_2| \leq [1 + (bz_*^{-2} + 2\kappa s + s)C_{\mathsf{z}} + 4\kappa]\|F-G\|_\infty.$$

□

Proof of Lemma 4.12: Obviously, $\mathsf{P}_{FGT}(\cdot;\mathsf{z},\Lambda)$ is well defined on $D_{cdf}([a,b])$ because $\mathsf{z} \geq z_* > 0$ is assumed. Next, fix $F \in \mathscr{D}$ and $G \in D_{cdf}([a,b])$. Since $\Lambda(0) = 0$, we can write

$$\mathsf{P}_{FGT}(F;\mathsf{z},\Lambda) = \int_{[a,b]} \Lambda(\max(1 - [x/\mathsf{z}(F)], 0))dF(x).$$

Abbreviating $f(x) := \Lambda(\max(1-[x/\mathsf{z}(F)], 0))$ and $g(x) := \Lambda(\max(1-[x/\mathsf{z}(G)], 0))$, we obtain

$$\mathsf{P}_{FGT}(F; \mathsf{z}, \Lambda) - \mathsf{P}_{FGT}(G; \mathsf{z}, \Lambda) = \int_{[a,b]} [f(x) - g(x)] dF(x) + \left[\int_{[a,b]} g(x) dF(x) - \int_{[a,b]} g(x) dG(x) \right].$$

Denote the first integral on the right by A, and the term in brackets to the far right by B. Because $g : [a, b] \to [0, \Lambda(1)]$ is continuous and non-increasing, Lemma D.7 implies $|B| \leq \Lambda(1) \|F - G\|_\infty$. Concerning A, we use the Lipschitz-continuity of Λ, and the inequality $|\max(1 - z_1, 0) - \max(1 - z_2, 0)| \leq |z_1 - z_2|$ for nonnegative z_1, z_2, to bound

$$|A| \leq bC_\Lambda |[1/\mathsf{z}(F)] - [1/\mathsf{z}(G)]| \leq bz_*^{-2} C_\Lambda |\mathsf{z}(F) - \mathsf{z}(G)| \leq bz_*^{-2} C_\Lambda C_{\mathsf{z}} \|F - G\|_\infty,$$

where we used the Lipschitz-continuity of the map $x \mapsto x^{-1}$ on $[z_*, \infty)$ (with constant z_*^{-2}), and the assumption that z satisfies Assumption 2.2 for obtaining the second inequality. Together with the upper bound on $|B|$ we obtain the claimed statement. $\square$

D General results for establishing Assumption 2.2

In this appendix, we summarize in a self-contained way a body of techniques that turns out to be useful for establishing Assumption 2.2 for empirically relevant functionals T. Once Assumption 2.2 is verified for a given functional T the Dvoretzky-Kiefer-Wolfowitz-Massart inequality delivers a concentration inequality for T of the type

$$\mathbb{P}(|\mathsf{T}(\hat{F}_n) - \mathsf{T}(F)| > \varepsilon) \leq 2e^{-2n\varepsilon^2/C^2} \quad \text{for every } \varepsilon > 0; \tag{106}$$

(here $\hat{F}_n$ denotes the empirical cdf of an i.i.d. sample of size n from the cdf $F \in \mathscr{D}$), a fact which we heavily use after an optional skipping argument, e.g., in the proofs concerning the finite-sample upper bounds on the F-UCB policy. As already mentioned at the end of Section 2.1.1, due to its simplicity and generality, such a concentration inequality could also be of independent interest for, e.g., constructing uniformly valid confidence intervals in finite samples.

Applications of the results in the present section to specific functionals are discussed in detail in Section 4. They include inequality measures (cf. Section 4.2), welfare measures (cf. Section 4.3), and poverty measures (cf. Section 4.4).

The techniques we describe are based on decomposability-properties of the functional, its specific structural (e.g., linearity) properties, and on properties of quantiles and quantile functions, or related quantities such as Lorenz curves. We emphasize that *the results in the present section are elementary, but are difficult to pinpoint in the literature in the form needed.* We start with a short section concerning notation.

D.1 Notation

We denote by $D(\mathbb{R})$ the Banach space of real-valued bounded càdlàg functions equipped with the supremum norm $\|G\|_\infty = \sup\{|G(x)| : x \in \mathbb{R}\}$. The closed convex subset of $D(\mathbb{R})$

consisting of all cumulative distribution functions (cdfs) shall be denoted by $D_{cdf}(\mathbb{R})$. Furthermore, given two real numbers $a < b$, we define the subset $D_{cdf}((a, b])$ of $D_{cdf}(\mathbb{R})$ as follows: $F \in D_{cdf}((a, b])$ if and only if $F \in D_{cdf}(\mathbb{R})$, $F(a) = 0$ and $F(b) = 1$. We also recall the definition of $D_{cdf}([a, b])$ from Section 2.1: $F \in D_{cdf}([a, b])$ if and only if $F \in D_{cdf}(\mathbb{R})$, $F(a-) = 0$ and $F(b) = 1$. Here $F(a-)$ denotes the left-sided limit of F at a. Recall also from the beginning of the Appendix of this article that given a cdf F, we denote by μ_F the (uniquely defined) probability measure on the Borel sets of $\mathbb{R}$ that satisfies

$$\mu_F((-\infty, x]) = F(x) \quad \text{for every } x \in \mathbb{R};$$

as usual, we denote the integral of a μ_F-integrable Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ by $\int_{\mathbb{R}} f(x)dF(x) := \int_{\mathbb{R}} f(x)d\mu_F(x)$.

In the following subsections we shall repeatedly encounter *functionals* T with a domain $\mathscr{T} \subseteq D_{cdf}(\mathbb{R})$, say, and co-domain $\mathbb{R}$, which are Lipschitz continuous ($\mathscr{T}$ being equipped with the metric induced by the supremum norm on $D(\mathbb{R})$): Recall that a functional $\mathsf{T} : \mathscr{T} \to \mathbb{R}$ is called *Lipschitz continuous* if there exists a nonnegative real number C such that for every F and every $G \in \mathscr{T}$ it holds that

$$|\mathsf{T}(F) - \mathsf{T}(G)| \leq C\|F - G\|_\infty.$$

In this case, we call C *a* Lipschitz constant of T. When we say that a functional $\mathsf{T} : \mathscr{T} \to \mathbb{R}$ is Lipschitz continuous with constant C, we do not imply that this is the smallest such constant. Recall from Remark 2.3 that if a functional T is Lipschitz continuous on $\mathscr{T} = D_{cdf}([a, b])$ for real numbers $a < b$, then T satisfies Assumption 2.2 with $\mathscr{D} = D_{cdf}([a, b])$.

D.2 Decomposability

Oftentimes a given functional can be decomposed into a function of several "simpler" functionals. It is a straightforward but useful fact that if a functional can be written as a composition of a number of functionals that satisfy Assumption 2.2 with a Lipschitz continuous function on a suitable intermediating metric space, this composition satisfies Assumption 2.2 as well. A corresponding result is as follows.

Lemma D.1. *Let $a < b$ be real numbers, and let $\emptyset \neq \mathscr{D} \subseteq D_{cdf}([a, b])$. Let $m \in \mathbb{N} \cup \{\infty\}$. For every $i \in \{1, \ldots, m\} \cap \mathbb{N}$ let $\mathsf{T}_i : D_{cdf}([a, b]) \to \mathbb{R}$ satisfy Assumption 2.2 with $\mathscr{D}$ and with constant C_i. Denote by $\bar{C}$ the vector with i-th coordinate C_i, and by $\overline{\mathsf{T}}$ the vector with i-th coordinate T_i. Set $\mathfrak{I} := \{\overline{\mathsf{T}}(F) : F \in D_{cdf}([a, b])\} \subseteq \mathbb{R}^m$. Suppose that for $p \in [1, \infty]$ it holds that $\|\bar{C}\|_p = (\sum_i |C_i|^p)^{1/p} < \infty$. Then, $(x, y) \mapsto \|x - y\|_p$ defines a metric on $\mathfrak{I}$. If the function $G : \mathfrak{I} \to \mathbb{R}$ is Lipschitz continuous with constant C (with respect to the just-mentioned metric), then $\mathsf{T} = G \circ \overline{\mathsf{T}}$ satisfies Assumption 2.2 with $\mathscr{D}$ and constant $C\|\bar{C}\|_p$.*

Proof. We first show that $(x, y) \mapsto \|x - y\|_p$ defines a metric on $\mathfrak{I} \subseteq \mathbb{R}^m$. To this end, we only need to verify that $\|x - y\|_p < \infty$ for every $x, y \in \mathfrak{I}$; all remaining properties of a metric are trivially satisfied. For every $x, y \in \mathfrak{I}$ there exist $F, G \in D_{cdf}([a, b])$ such that $x_i = \mathsf{T}_i(F)$ and $y_i = \mathsf{T}_i(H)$ for every i. Fix an arbitrary element $F^* \in \mathscr{D}$. It then follows from Assumption 2.2 that $|x_i - y_i| \leq |\mathsf{T}_i(F) - \mathsf{T}_i(F^*)| + |\mathsf{T}_i(F^*) - \mathsf{T}_i(H)| \leq 2C_i$.

Hence, $\|\bar{C}\|_p < \infty$ implies $\|x - y\|_p < \infty$. Having established the first claim in the lemma, we move on to the final claim. Let $F \in \mathscr{D}$ and $H \in D_{cdf}([a,b])$. We have $|\mathsf{T}(F) - \mathsf{T}(H)| = |G(\overline{\mathsf{T}}(F)) - G(\overline{\mathsf{T}}(H))| \leq C\|\overline{\mathsf{T}}(F) - \overline{\mathsf{T}}(H)\|_p$, the inequality following from Lipschitz continuity of $G : \mathfrak{I} \to \mathbb{R}$. From the definition of $\|\cdot\|_p$ and Assumption 2.2 it immediately follows that $\|\overline{\mathsf{T}}(F) - \overline{\mathsf{T}}(H)\|_p \leq \|\bar{C}\|_p \|F - H\|_\infty$, which proves the lemma. □

D.3 U-functionals

We here consider "U-functionals" (the corresponding sample plug-in variants being traditionally referred to as U-statistics, hence the name). The following result covers examples such as moments and certain concentration measures or dependence measures, cf. Chapter 5 in Serfling (2009), and see also the subsequent discussion for examples.

Lemma D.2. *Let $a < b$ be real numbers and let $\varphi : [a,b]^k \to \mathbb{R}$ for some $k \in \mathbb{N}$. Suppose that φ is bounded, and is symmetric in the sense that $\varphi(x_1, \ldots, x_k) = \varphi(x_{\pi_1}, \ldots, x_{\pi_k})$ for every permutation $x_{\pi_1}, \ldots, x_{\pi_k}$ of $x_1, \ldots, x_k$. Let $a \leq c < d \leq b$. Suppose that for every $x_2^*, \ldots, x_k^* \in [c,d]^{k-1}$ the function $x \mapsto \varphi(x, x_2^*, \ldots, x_k^*)$ defined on $[c,d]$ is continuous and has total variation not greater than $C \in \mathbb{R}$. For $F \in D_{cdf}([a,b])$ define*

$$\mathsf{m}_{\varphi;c,d}(F) := \int_{[c,d]} \cdots \int_{[c,d]} \varphi(x_1, \ldots, x_k) dF(x_1) \ldots dF(x_k), \tag{107}$$

which we abbreviate as $\mathsf{m}_\varphi(\cdot)$ in case $c = a$ and $d = b$. Then, $\mathsf{m}_{\varphi;c,d}$ is Lipschitz continuous on $D_{cdf}([a,b])$ with constant kC^*, where

$$C^* = \begin{cases} C & \text{if } a = c, b = d \\ C + m^* & \text{if } b = d \\ C + M^* & \text{if } a = c \\ C + m^* + M^* & \text{else,} \end{cases}$$

and where

$$\begin{aligned} m^* &:= \sup\{|\varphi(c, x_2^*, \ldots, x_k^*)| : x_2^*, \ldots, x_k^* \in [c,d]^{k-1}\} \\ M^* &:= \sup\{|\varphi(d, x_2^*, \ldots, x_k^*)| : x_2^*, \ldots, x_k^* \in [c,d]^{k-1}\}. \end{aligned}$$

Proof. Note first that $\mathsf{m}_{\varphi;c,d}(F)$ is well defined (i.e., φ is integrable w.r.t. the k-fold product measure $\bigotimes_{i=1}^k \mu_F$) on $D_{cdf}([a,b])$ because φ is bounded. Next, we reduce the statement to the case $k = 1$: Let $F, G \in D_{cdf}([a,b])$, let μ be a probability measure that dominates μ_F and μ_G, and let f and g denote μ-densities of μ_F and μ_G, respectively. Then,

$$\mathsf{m}_{\varphi;c,d}(F) = \int_{[c,d]} \cdots \int_{[c,d]} \varphi(x_1, \ldots, x_k) \prod_{j=1}^k f(x_j) d\mu(x_1) \ldots d\mu(x_k), \tag{108}$$

and an analogous expression (replacing the density f by the density g) corresponds to $\mathsf{m}_{\varphi;c,d}(G)$. Recall also that for arbitrary real numbers a_j, b_j for $j = 1, \ldots, k$ we may write (e.g., Witting and Müller-Funk (1995) Hilfssatz 5.67(a))

$$\prod_{j=1}^{k} a_j - \prod_{j=1}^{k} b_j = \sum_{j=1}^{k} \left[\left(\prod_{i=1}^{j-1} a_i \right) (a_j - b_j) \prod_{i=j+1}^{k} b_i \right], \tag{109}$$

where empty products are to be interpreted as 1. Equipped with (109), using Equation (108), and Fubini's theorem, we write $\mathsf{m}_{\varphi;c,d}(F) - \mathsf{m}_{\varphi;c,d}(G)$ as

$$\sum_{j=1}^{k} \int_{[c,d]} \cdots \int_{[c,d]} \varphi(x_1, \ldots, x_k)[f(x_j) - g(x_j)] d\mu(x_j) dF(x_1) \ldots dF(x_{j-1}) dG(x_{j+1}) \ldots dG(x_k).$$

Using the triangle inequality to upper bound $|\mathsf{m}_{\varphi;c,d}(F) - \mathsf{m}_{\varphi;c,d}(G)|$, an application of the symmetry condition shows that it suffices to verify that for $x_2^*, \ldots, x_k^*$ in $[c,d]^{k-1}$ arbitrary

$$\left| \int_{[c,d]} \varphi(x, x_2^* \ldots, x_k^*) dF(x) - \int_{[c,d]} \varphi(x, x_2^* \ldots, x_k^*) dG(x) \right| \leq C^* \|F - G\|_\infty. \tag{110}$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function (possibly depending on $x_2^*, \ldots, x_k^*$) such that $f(x) = \varphi(x, x_2^* \ldots, x_k^*)$ holds for every $x \in [c,d]$, and such that $f(x) \to 0$ as $x \to -\infty$. Integration-by-parts (as in, e.g., Exercise 34.b on p.108 in Folland (1999)) gives

$$\int_{[c,d]} \varphi(x, x_2^* \ldots, x_k^*) dF(x) = \int_{[c,d]} f(x) dF(x) = f(d)F(d) - f(c-)F(c-) - \int_{[c,d]} F(x) df(x),$$

an analogous statement holding for F replaced by G. Hence, the quantity to the left in the inequality in (110) is seen to be not greater than

$$|f(d)||F(d) - G(d)| + |f(c)||F(c-) - G(c-)| + \left| \int_{[c,d]} F(x) - G(x) df(x) \right|.$$

Noting that $|f(d)| \leq M^*$, that $|f(c)| \leq m^*$, that $|F(d) - G(d)| = 0$ if $d = b$, that $|F(c-) - G(c-)| = 0$ if $a = c$, and furthermore noting that $|F(d) - G(d)| \leq \|F - G\|_\infty$ and $|F(c-) - G(c-)| \leq \|F - G\|_\infty$ always hold, (110) follows from $\left| \int_{[c,d]} F(x) - G(x) df(x) \right| \leq \|F - G\|_\infty C$, a consequence of the total variation of f on $[c,d]$ being not greater than C. $\square$

Example D.3 (Mean)**.** Let $a < b$ be real numbers. Let $k = 1$ and set $\varphi(x) = x$, i.e., we consider the mean functional $F \mapsto \mu(F)$, say, defined via

$$F \mapsto \int_{[a,b]} x dF(x).$$

Note that φ is bounded on $[a,b]$, is trivially symmetric, and φ satisfies the continuity condition in Lemma D.2. Furthermore, the total variation of φ is $\int_{[a,b]} |\varphi'(x)| dx = (b - a)$. As a consequence of Lemma D.2 the functional m_φ is thus Lipschitz continuous on $D_{cdf}([a,b])$ with constant $(b - a)$.

Example D.4 (Moments)**.** For simplicity, let $a = 0$ and $b > 0$. Let $k = 1$ and set $\varphi(x) = x^p$ for some $p > 0$, i.e., we consider the p-mean functional

$$F \mapsto \int_{[0,b]} x^p dF(x).$$

Note that φ is bounded on $[a, b]$, is trivially symmetric, and φ satisfies the continuity condition in Lemma D.2. Furthermore, by monotonicity, the total variation of φ is b^p. As a consequence of Lemma D.2 the functional m_φ is thus Lipschitz continuous on $D_{cdf}([0, b])$ with constant b^p.

Example D.5 (Variance)**.** Let $a < b$ be real numbers. Let $k = 2$ and set $\varphi(x_1, x_2) = 0.5(x_1 - x_2)^2$, i.e., we consider the variance

$$F \mapsto 0.5 \int_{[a,b]} \int_{[a,b]} (x_1 - x_2)^2 dF(x_1)dF(x_2) = \int_{[a,b]} \left[x_1 - \int_{[a,b]} x_2 dF(x_2) \right]^2 dF(x_1).$$

Note that φ is bounded on $[a, b]^2$, is symmetric, and φ satisfies the continuity condition in Lemma D.2. For every $x_2 \in [a, b]$ the total variation of $x \mapsto 0.5(x - x_2)^2$ is $\int_{[a,b]} |x - x_2| dx \leq (a - b)^2/2$. It follows from Lemma D.2 that the variance functional is Lipschitz continuous with constant $(a - b)^2$.

Example D.6 (Gini-mean difference)**.** Let $a < b$ be real numbers, and let $\varphi(x_1, x_2) = |x_1 - x_2|$. This corresponds to the functional

$$F \mapsto \int_{[a,b]} \int_{[a,b]} |x_1 - x_2| dF(x_1)dF(x_2),$$

which constitutes the numerator of the Gini-index defined in Equation (23) (and equals twice the absolute Gini index G_{abs} defined in Equation (24)), and is sometimes called the Gini-mean difference or absolute mean difference. Clearly, φ is bounded on $[a, b]^2$, symmetric, and satisfies the continuity condition in Lemma D.2. Furthermore, for every $x_2 \in [a, b]$ the total variation of $x \mapsto |x - x_2|$ equals $(b - a)$. It follows from Lemma D.2 that m_φ is Lipschitz continuous on $D_{cdf}([a, b])$ with constant $2(b - a)$.

The following lemma is sometimes useful, because it avoids the continuity condition of the integrand in Lemma D.2 by working with a right-continuity and monotonicity condition.

Lemma D.7. *Let $a < b$ be real numbers and let $\varphi : [a, b] \to \mathbb{R}$ be right-continuous, and be non-decreasing or non-increasing. Then, the functional*

$$F \mapsto \int_{[a,b]} \varphi(x) dF(x)$$

is Lipschitz continuous on $D_{cdf}([a, b])$ with constant $|\varphi(b) - \varphi(a)|$.

Proof. Note first that the functional under consideration is well defined on $D_{cdf}([a,b])$; and that we only need to consider the case where φ is non-decreasing. To this end let $F, G \in D_{cdf}([a,b])$ and note that, by the transformation theorem, we have

$$\int_{[a,b]} \varphi(x) dF(x) - \int_{[a,b]} \varphi(x) dG(x) = \int_{[\varphi(a),\varphi(b)]} x dF_\varphi(x) - \int_{[\varphi(a),\varphi(b)]} x dG_\varphi(x),$$

where $F_\varphi \in D_{cdf}([\varphi(a), \varphi(b)])$ denotes the cdf corresponding to the image measure $\mu_F \circ \varphi$, and $G_\varphi \in D_{cdf}([\varphi(a), \varphi(b)])$ is defined analogously. An application of Example D.3 thus shows that

$$\left| \int_{[a,b]} \varphi(x) dF(x) - \int_{[a,b]} \varphi(x) dG(x) \right| \leq [\varphi(b) - \varphi(a)] \|F_\varphi - G_\varphi\|_\infty.$$

It remains to observe that $\|F_\varphi - G_\varphi\|_\infty \leq \|F - G\|_\infty$, by Lemma D.8. □

Lemma D.8. *Let F and G be cdfs, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be right-continuous, and be non-decreasing. Then $\|F_\varphi - G_\varphi\|_\infty \leq \|F - G\|_\infty$, where F_φ denotes the cdf corresponding to the image measure $\mu_F \circ \varphi$, and G_φ is defined analogously.*

Proof. First of all, note that $\|F_\varphi - G_\varphi\|_\infty = \sup_{z \in C(F,G)} |F_\varphi(z) - G_\varphi(z)|$, where $C(F,G) \subseteq \mathbb{R}$ is defined as the (dense) subset of points at which both F_φ and G_φ are continuous. Next, define $\varphi^-(x) := \inf\{y \in \mathbb{R} : \varphi(y) \geq x\}$, i.e., a generalized inverse of φ. Part (5) of Proposition 1 in Embrechts and Hofert (2013) shows that for every $z \in \mathbb{R}$ we have

$$A(z) := \{x \in \mathbb{R} : \varphi(x) < z\} = \{x \in \mathbb{R} : x < \varphi^-(z)\}.$$

Using this expression for $A(z)$, we can for every $z \in C(F,G)$ rewrite $|F_\varphi(z) - G_\varphi(z)|$ as

$$\begin{aligned} |\mu_{F_\varphi}((-\infty, z)) - \mu_{G_\varphi}((-\infty, z))| &= |\mu_F(A(z)) - \mu_G(A(z))| \\ &= |\mu_F(\{x \in \mathbb{R} : x < \varphi^-(z)\}) - \mu_G(\{x \in \mathbb{R} : x < \varphi^-(z)\})|. \end{aligned}$$

On the one hand, the expression to the far right in the previous display equals $0 \leq \|F - G\|_\infty$ in case $\varphi^-(z) \in \{-\infty, +\infty\}$. On the other hand, if $\varphi^-(z) \in \mathbb{R}$, the same expression is seen to equal $|F(\varphi^-(z)-) - G(\varphi^-(z)-)| \leq \|F - G\|_\infty$. Since this argument goes through for every $z \in C(F,G)$, we are done. □

D.4 Quantiles, quantile functions, L-functionals, Lorenz curve, and truncation

In the present subsection we provide some results concerning quantile-based functionals. For $\alpha \in [0,1]$ we define the α-quantile of a cdf F as usual via $q_\alpha(F) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$. Note that for $\alpha = 0$ we have $q_\alpha(F) = -\infty$, and that (by monotonicity) the quantile function $\alpha \mapsto q_\alpha(F)$ is $\mathcal{B}([0,1]) - \mathcal{B}(\bar{\mathbb{R}})$ measurable. The first result is as follows:

Lemma D.9. *Let $\alpha \in (0,1]$ and let $F \in D_{cdf}([a,b])$ for real numbers $a < b$. Suppose $F(q_\alpha(F)) = \alpha$ and that there exists a positive real number r such that*

$$\begin{aligned} &F(q_\alpha(F) - x) - \alpha \leq -rx \qquad \text{if} \quad x > 0 \text{ and } q_\alpha(F) - x > a, \\ &F(q_\alpha(F) + x) - \alpha \geq rx \qquad \text{if} \quad x > 0 \text{ and } q_\alpha(F) + x < b. \end{aligned} \tag{111}$$

Then, for every $G \in D_{cdf}([a,b])$ it holds that $|q_\alpha(F) - q_\alpha(G)| \leq r^{-1}\|F - G\|_\infty$. Consequently, denoting by $\mathscr{D}$ the set of all cdfs that satisfy the conditions imposed on F above, it follows that q_α satisfies Assumption 2.2 with $a, b, \mathscr{D}$ and constant $C = r^{-1}$.

Proof. To prove the first statement, we may impose the additional assumption that the inequalities to the left in Equation (111) hold strictly for all x in the considered ranges (to see this, just observe that Equation (111) implies the just mentioned strict version for all $0 < r_* < r$, which can then be used to take care of situations where the additional assumption is not satisfied).

Given this additional assumption, let G be an element of $D_{cdf}([a,b])$. The claimed inequality is trivial if $F = G$. Thus, we assume that $F \neq G$. Note that $F(x) = 0 < \alpha$ for every $x < a$, and $F(x) = 1 \geq \alpha$ for every $x \geq b$ implies $q_\alpha(F) \in [a,b]$; and that, by the same reasoning, $q_\alpha(G) \in [a,b]$.

We first show that $q_\alpha(G) \geq q_\alpha(F) - r^{-1}\|G - F\|_\infty$: On the one hand, if $q_\alpha(F) - r^{-1}\|G - F\|_\infty \leq a$, then $q_\alpha(G) \geq q_\alpha(F) - r^{-1}\|G - F\|_\infty$ trivially holds. If, on the other hand, $q_\alpha(F) - r^{-1}\|G - F\|_\infty > a$, then, from the (strict) inequality in the first line of (111) with $x = r^{-1}\|G - F\|_\infty$, one obtains $\alpha > F(q_\alpha(F) - r^{-1}\|G - F\|_\infty) + \|G - F\|_\infty$, thus $\alpha > G(q_\alpha(F) - r^{-1}\|G - F\|_\infty)$ and hence, again, $q_\alpha(G) \geq q_\alpha(F) - r^{-1}\|G - F\|_\infty$.

We next show that $q_\alpha(G) \leq q_\alpha(F) + r^{-1}\|G - F\|_\infty$: On the one hand, if $q_\alpha(F) + r^{-1}\|G - F\|_\infty \geq b$, then $q_\alpha(G) \leq q_\alpha(F) + r^{-1}\|G - F\|_\infty$ trivially holds. If, on the other hand, $q_\alpha(F) + r^{-1}\|G - F\|_\infty < b$, then the second line in (111) with $x = r^{-1}\|G - F\|_\infty$ shows that $F(q_\alpha(F) + r^{-1}\|G - F\|_\infty) - \|G - F\|_\infty \geq \alpha$, thus $G(q_\alpha(F) + r^{-1}\|G - F\|_\infty) \geq \alpha$, and hence, again, $q_\alpha(G) \leq q_\alpha(F) + r^{-1}\|G - F\|_\infty$. Summarizing yields $|q_\alpha(F) - q_\alpha(G)| \leq r^{-1}\|F - G\|_\infty$. The last statement is trivial. $\square$

Example D.10 (Median)**.** The median of a distribution F is defined as its $\alpha = 1/2$ quantile $q_{1/2}(F)$. Let $a < b$ and $r > 0$ be real numbers, and denote by $\mathscr{D}$ the set of cdfs F such that $F(q_{1/2}(F)) = 1/2$, and such that Equation (111) is satisfied for $\alpha = 1/2$ (Lemma D.12 provides a sufficient condition for $F \in \mathscr{D}$). Then, the functional $F \mapsto q_{1/2}(F)$ satisfies Assumption 2.2 with a, b and $\mathscr{D}$ with constant $C = r^{-1}$.

The second result is auxiliary, and concerns not a single quantile, but the whole quantile *function* $F \mapsto q_.(F)$ over closed subintervals of $(0,1]$. It follows immediately from Lemma D.9.

Lemma D.11. *Let $F \in D_{cdf}([a,b])$ for real numbers $a < b$, and let $\alpha_* < \alpha^*$ for α_* and α^* in $(0,1]$. Suppose $F(q_\alpha(F)) = \alpha$ holds for every $\alpha \in [\alpha_*, \alpha^*]$, and that there exists a positive real number r so that Equation* (111) *is satisfied for every $\alpha \in [\alpha_*, \alpha^*]$. Then, for every $G \in D_{cdf}([a,b])$ it holds that*

$$\sup_{\alpha \in [\alpha_*, \alpha^*]} |q_\alpha(F) - q_\alpha(G)| \leq r^{-1}\|F - G\|_\infty.$$

A simple sufficient condition for the assumption on F in Lemma D.11 (and hence also for the assumption on F in Lemma D.9) is that F admits a density that is bounded from below (on the support of F):

Lemma D.12. *Let $a < b$ be real numbers and let $F \in D_{cdf}([a,b])$. Suppose F is continuous, and is right-sided differentiable on (a,b) with right-sided derivative F^+, which furthermore satisfies $F^+(x) \geq r$ for every $x \in (a,b)$ for some $r > 0$. Then, $F(q_\alpha(F)) = \alpha$ and Equation* (111) *holds for every $\alpha \in (0,1]$.*

Proof. The condition $F^+(x) \geq r$ for every $x \in (a,b)$ for a $r > 0$ implies that F is strictly increasing on $[a,b]$, which (together with continuity of F) implies $F(q_\alpha(F)) = \alpha$ for every $\alpha \in (0,1]$. The second claim follows from the mean-value theorem for right-differentiable functions in Minassian (2007) (noting that $q_\alpha(F) \in [a,b]$ for every $\alpha \in (0,1]$, cf. the proof of Lemma D.9). $\square$

The next result, which essentially follows from the previous one, concerns population versions of generalized L-statistics introduced by Serfling (1984) (cf. his Section 2), i.e., L-functionals.

Lemma D.13. *Let ν be a measure on the Borel sets of $[0,1]$, and let $J : [0,1] \to \mathbb{R}$ be such that $\int_{[0,1]} |J(\alpha)| d\nu(\alpha) = c < \infty$. Assume further that $\nu(\{0\}) = 0$. Let $a < b$ be real numbers and define on $D_{cdf}([a,b])$ the functional*

$$\mathsf{T}(F) = \int_{[0,1]} q_\alpha(F) J(\alpha) d\nu(\alpha). \tag{112}$$

Let $F \in D_{cdf}([a,b])$ satisfy $F(q_\alpha(F)) = \alpha$ for every $\alpha \in (0,1]$, and suppose there is a positive real number r so that Equation (111) *holds for every $\alpha \in (0,1]$. Then, for every $G \in D_{cdf}([a,b])$, it holds that*

$$|\mathsf{T}(F) - \mathsf{T}(G)| \leq \frac{c}{r} \|F - G\|_\infty.$$

Consequently, denoting by $\mathscr{D}$ the set of all cdfs that satisfy the conditions imposed on F above, it follows that T defined in Equation (112) *satisfies Assumption 2.2 with $a, b, \mathscr{D}$ and constant $C = c/r$.*

Proof. That $\int_{[0,1]} q_\alpha(F) J(\alpha) d\nu(\alpha)$ exists for every $F \in D_{cdf}([a,b])$ follows from $\nu(\{0\}) = 0$, from $q_\alpha(F) \in [a,b]$ for every $\alpha \in (0,1]$ (cf. the proof of Lemma D.9), and from the integrability condition on J. Next, for F and G as in the statement of the lemma, note that

$$|\mathsf{T}(F) - \mathsf{T}(G)| \leq \int_{(0,1]} |q_\alpha(F) - q_\alpha(G)| |J(\alpha)| d\nu(\alpha). \tag{113}$$

Note that the function $\alpha \mapsto |q_\alpha(F) - q_\alpha(G)|$ is bounded on $(0,1]$. By the monotone convergence theorem, for $\varepsilon \searrow 0$ the integral $\int_{[\varepsilon,1]} |q_\alpha(F) - q_\alpha(G)| |J(\alpha)| d\nu(\alpha)$ converges to the integral in (113). But $\int_{[\varepsilon,1]} |q_\alpha(F) - q_\alpha(G)| |J(\alpha)| d\nu(\alpha) \leq r^{-1} c \|F - G\|_\infty$ by Lemma D.11. The last statement in the lemma is trivial. $\square$

One particularly important application of Lemma D.13 concerns the so-called Lorenz curve associated with a cdf F (cf. Gastwirth (1971)).

Lemma D.14. *Let $a < b$ be real numbers and define on $D_{cdf}([a, b])$ the family of functionals indexed by $u \in [0, 1]$ and defined by*

$$Q(F, u) := \int_{[0,u]} q_\alpha(F) d\alpha;$$

furthermore, if $a > 0$, define the family of functionals indexed by $u \in [0, 1]$ via

$$L(F, u) := \mu(F)^{-1} \int_{[0,u]} q_\alpha(F) d\alpha \tag{114}$$

Let $F \in D_{cdf}([a, b])$ satisfy $F(q_\alpha(F)) = \alpha$ for every $\alpha \in (0, 1]$, and suppose there is a positive real number r such that Equation (111) holds for every $\alpha \in (0, 1]$. Then, for every $G \in D_{cdf}([a, b])$ it holds that

$$|Q(F, u) - Q(G, u)| \leq r^{-1} u \|F - G\|_\infty \leq r^{-1} \|F - G\|_\infty. \tag{115}$$

Consequently, denoting by $\mathscr{D}$ the set of all cdfs that satisfy the conditions imposed on F above, it follows that $\mathsf{T}(\cdot) = Q(\cdot, u)$ satisfies Assumption 2.2 with $a, b, \mathscr{D}$ and constant $C = r^{-1}u$. Furthermore, if $a > 0$, then

$$|L(F, u) - L(G, u)| \leq a^{-1}(r^{-1} + (b - a)a^{-1}b)u\|F - G\|_\infty,$$

and it follows that $\mathsf{T}(\cdot) = L(\cdot, u)$ satisfies Assumption 2.2 with $a, b, \mathscr{D}$ and constant $C = a^{-1}(r^{-1} + (b - a)a^{-1}b)u$.

Proof. For the claim in Equation (115) we just apply Lemma D.13 with ν equal to Lebesgue measure, $J = \mathbb{1}_{[0,u]}$, which satisfies the integrability condition with $c = u \leq 1$. For the second claim, note that $L(\cdot, u)$ is well defined on $D_{cdf}([a, b])$ because $a > 0$. Next, observe that for F and G as in the statement of the lemma we can bound $|L(F, u) - L(G, u)|$ from above by

$$\mu(F)^{-1} \left\{ \big|Q(F, u) - Q(G, u)\big| + |1 - \mu(F)/\mu(G)| \int_{[0,u]} q_\alpha(G) d\alpha \right\}. \tag{116}$$

Since $\mu(G)$ and $\mu(F)$ are not smaller than a, since $q_\alpha(G) \leq b$ for $\alpha \in (0, u]$, and because we already know that

$$\big|Q(F, u) - Q(G, u)\big| \leq r^{-1} u \|F - G\|_\infty,$$

it remains to observe that by Example D.3

$$|1 - (\mu(F)/\mu(G))| \leq (b - a)\|F - G\|_\infty/\mu(G) \leq (b - a)a^{-1}\|F - G\|_\infty$$

to conclude that the expression in (116) is not greater than $a^{-1}\left\{r^{-1} + (b - a)a^{-1}b\right\} u\|F - G\|_\infty$. $\square$

The final result in this section concerns trimmed generalized-mean functionals. We consider one-sidedly trimmed functionals, the trimming affecting the lower or upper tail. Two-sided trimming can be dealt with similarly. We abstain from spelling out the details.

Lemma D.15. *Let $a < b$ be real numbers, let $\varphi : \mathbb{R} \to \mathbb{R}$, let φ restricted to $[a, b]$ be continuous, let the total variation of φ on $[a, b]$ be not greater than C, and let $|\varphi(x)| \leq u$ hold for all $x \in [a, b]$. Furthermore, let $\alpha \in (0, 1)$. For $F \in D_{cdf}([a, b])$ define*

$$\mathsf{m}^{t-}_{\varphi;\alpha}(F) := \int_{[a,q_\alpha(F)]} \varphi(x)dF(x) \quad \textit{and} \quad \mathsf{m}^{t+}_{\varphi;\alpha}(F) := \int_{[q_\alpha(F),b]} \varphi(x)dF(x). \tag{117}$$

Let $F \in D_{cdf}([a, b])$, assume that F is continuous, and right-sided differentiable on (a, b), with right-sided derivative F^+ satisfying $r \leq F^+(x) \leq \kappa$ for every $x \in (a, b)$, and for positive real numbers κ and r. Then, for every $G \in D_{cdf}([a, b])$ it holds that

$$|\mathsf{m}^{t-}_{\varphi;\alpha}(F) - \mathsf{m}^{t-}_{\varphi;\alpha}(G)| \leq [C + u(1 + \kappa r^{-1})]\|F - G\|_\infty,$$

and

$$|\mathsf{m}^{t+}_{\varphi;\alpha}(F) - \mathsf{m}^{t+}_{\varphi;\alpha}(G)| \leq [C + u(1 + \kappa r^{-1})]\|F - G\|_\infty,$$

Consequently, denoting by $\mathscr{D}$ the set of all cdfs that satisfy the conditions imposed on F above, it follows that $m^{t-}_{\varphi;\alpha}$ and $m^{t+}_{\varphi;\alpha}$ satisfy Assumption 2.2 with $a, b, \mathscr{D}$ and constant $C+u(1+\kappa r^{-1})$.

Proof. We only provide an argument for the first claimed inequality, the second is obtained analogously. Furthermore, throughout the proof we write $\mathsf{m}^{t}_{\varphi;\alpha}$ instead of $\mathsf{m}^{t-}_{\varphi;\alpha}$. First, note that the functional $\mathsf{m}^{t}_{\varphi;\alpha}(F)$ is indeed well defined for every $F \in D_{cdf}([a, b])$. This follows from $q_\alpha(F) \in [a, b]$ (cf. the proof of Lemma D.9), and since φ is bounded on $[a, b]$. Next, let F be as in the statement of the lemma and satisfy the conditions imposed. Let $G \in D_{cdf}([a, b])$, implying that $q_\alpha(G) \in [a, b]$. By the triangle inequality, $|\mathsf{m}^{t}_{\varphi;\alpha}(F) - \mathsf{m}^{t}_{\varphi;\alpha}(G)| \leq A + B$, where (using the notation introduced in Equation (107))

$$A := \left|\mathsf{m}_{\varphi;a,q_\alpha(G)}(F) - \mathsf{m}_{\varphi;a,q_\alpha(G)}(G)\right| \leq (C + u)\|F - G\|_\infty,$$

the upper bound following from Lemma D.2, and

$$B := \int g(x)|\varphi(x)|dF(x) \leq u \int g(x)dF(x),$$

where $g(x) = \left|\mathbb{1}_{[a,q_\alpha(F)]}(x) - \mathbb{1}_{[a,q_\alpha(G)]}(x)\right|$. By continuity of F:

$$\int g(x)dF(x) \leq |F(q_\alpha(G)) - F(q_\alpha(F))|.$$

which, by the assumed behavior of the right-derivative of F and a mean-value theorem for right-differentiable functions (for example the one by Minassian (2007)), is not greater than

$$\kappa|q_\alpha(G) - q_\alpha(F)| \leq \kappa r^{-1}\|F - G\|_\infty$$

the last inequality following from Lemma D.9 together with Lemma D.12. This proves the claim. The last statement is trivial. □