The Annals of Statistics
2023, Vol. 51, No. 3, 1030–1057
https://doi.org/10.1214/23-AOS2275

A GENERAL CHARACTERIZATION OF OPTIMAL TIE-BREAKER DESIGNS

BY HARRISON H. LI[a] AND ART B. OWEN[b]

Department of Statistics, Stanford University, [a]*hli90722@stanford.edu,* [b]*owen@stanford.edu*

Tie-breaker designs trade off a measure of statistical efficiency against a short-term gain from preferentially assigning a binary treatment to subjects with higher values of a running variable x. The efficiency measure can be any continuous function of the expected information matrix in a two-line regression model. The short-term gain is expressed as the covariance between the running variable and the treatment indicator. We investigate how to choose design functions $p(x)$ specifying the probability of treating a subject with running variable x in order to optimize these competing objectives, under external constraints on the number of subjects receiving treatment. Our results include sharp existence and uniqueness guarantees, while accommodating the ethically appealing requirement that $p(x)$ be nondecreasing in x. Under this condition, there is always an optimal treatment probability function $p(x)$ that is constant on the sets $(-\infty, t)$ and (t, ∞) for some threshold t and generally discontinuous at $x = t$. When the running variable distribution is not symmetric or the fraction of subjects receiving the treatment is not $1/2$, our optimal designs improve upon a D-optimality objective without sacrificing short-term gain, compared to a typical three-level tie-breaker design that fixes treatment probabilities at 0, $1/2$ and 1. We illustrate our optimal designs with data from Head Start, an early childhood government intervention program.

1. Introduction. Companies, charitable institutions and clinicians often have ethical or economic reasons to prefer assigning a binary treatment to certain individuals. If this preference is expressed by the values of a scalar running variable x, a natural decision is to assign the treatment to a subject if and only if their x is at least some threshold t. This is a regression discontinuity design, or RDD (Thistlethwaite and Campbell (1960)). Unfortunately, treatment effect estimates from an RDD analysis typically have very high variance (Gelman and Imbens (2019), Goldberger (1972), Jacob, Zhu and Somers (2012)), relative to those from a randomized controlled trial (RCT) that does not preferentially treat any individuals.

A tie-breaker design (TBD) provides a compromise between these competing objectives. In a typical TBD, the top ranked subjects get the treatment, the lowest ranked subjects do not get it and the remaining subjects' treatment status is randomized. The earliest tie-breaker reference we are aware of is Campbell (1969) where x was discrete and the randomization broke ties among subjects with identical values of x. Past settings for the TBD include offering remedial English to incoming university students based on their high school English proficiency (Aiken et al. (1998)), assigning arrested juveniles into a diversion program to reduce delinquency (Lipsey, Cordray and Berger (1981)), providing higher education scholarships based on a judgment of the applicants' needs and academic strengths (Abdulkadiroglu et al. (2017), Angrist, Autor and Pallais (2020)), and designing clinical trials (Trochim and Cappelleri (1992)), where they are known as cutoff designs.

The tie-breaker design problem we study involves choosing treatment probabilities p_i for subjects $i = 1, \dots, n$ based on their running variables x_i. These probabilities are chosen

Received October 2022; revised February 2023.
MSC2020 subject classifications. 62K05.
Key words and phrases. Constrained optimal design, compound design, equivalence theorem, cutoff design, D-optimality, regression discontinuity.

before observing the response values $y_1, \dots, y_n$ but with the running variables $x_1, \dots, x_n$ known. The goal is to optimally trade off competing statistical and economic objectives, as discussed above. We assume throughout that $p_i = p_{i'}$ whenever $x_i = x_{i'}$.

Consistent with the optimal experimental design literature, we consider the statistical objective to be an "efficiency" criterion that measures estimation precision. Specifically, our criterion will be a function $\Psi(\cdot)$ of the information (scaled inverse variance) matrix $\mathcal{I}_n(p_1, \dots, p_n)$ for the model parameters $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$ in a two-line regression model relating the response y_i to the running variable x_i and a treatment indicator $z_i \in \{-1, 1\}$:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \varepsilon_i. \tag{1}$$

Despite its simplicity, this working model poses some challenging design problems. In Section 6, we describe some more general modeling settings for tie-breaker models. Throughout, we assume that the running variable is centered, that is, $(1/n)\sum_i x_i = 0$, and that the ε_i have common variance $\sigma^2 > 0$. Here, $z_i = 1$ indicates treatment and so $p_i = \Pr(z_i = 1) = (1 + \mathbb{E}(z_i))/2$. For model (1), the information matrix is $\mathcal{I}_n = \mathbb{E}(\mathcal{X}_i \mathcal{X}_i^\top)$ where $\mathcal{X}_i = (1, x_i, z_i, x_i z_i)^\top \in \mathbb{R}^4$ and the expectation is taken over the treatment assignments z_i, conditional on the running variables x_i (whose values are known). Note that the ordinary least squares estimate $\hat{\beta}$ of β satisfies $\mathbb{E}(\mathrm{Var}(\hat{\beta})^{-1}) = n\mathcal{I}_n/\sigma^2$. Common examples of efficiency criteria $\Psi(\cdot)$ in the literature, such as the D-optimality criterion $\Psi_D(\cdot) = \log(\det(\cdot))$, are concave in both $\mathcal{I}_n$ and $p = (p_1, \dots, p_n)^\top$ (Boyd and Vandenberghe (2004)). However, our theoretical results only require continuity of $\Psi(\cdot)$.

The competing objective is a preference for treating individuals with higher running variables x, as discussed earlier. We express it as an equality constraint on the scaled covariance $\overline{xp} \equiv (1/n)\sum_{i=1}^n x_i p_i$ between treatment and the running variable (recall that the latter is known, hence viewed as nonrandom). Under the two-line model (1), this constraint has the following economic interpretation. We take y to be something like economic value or student success, where larger y is better. We expect that $\beta_3 > 0$ holds in most of our motivating problems. The expected value of y per customer under (1) is then

$$\mathbb{E}(y_i) = \beta_0 + \beta_2 \cdot (2\bar{p} - 1) + \beta_3 \cdot (2\overline{xp} - 1), \tag{2}$$

where $\bar{p} \equiv (1/n)\sum_{i=1}^n p_i$. Equation (2) shows that the expected gain is unaffected by β_0 or β_1. Furthermore, we assume the proportion of treated subjects is fixed by an external budget, that is, an equality constraint $\bar{p} = \widetilde{p}$ for some $\widetilde{p} \in (0, 1)$. For instance, there might be only a set number of scholarships or perks to be given out. The only term affected by the design in (2) is then $\beta_3 \cdot \overline{xp}$, as pointed out by Owen and Varian (2020). For $\beta_3 > 0$, the short-term average value per customer grows with $\overline{xp}$ and we would want that value to be large. Similar functionals are also commonly studied as regret functions in bandit problems (Goldenshluger and Zeevi (2013)) and sequential experimental design (Metelkina and Pronzato (2017)).

We are now ready to formulate the tie-breaker design problem as the following constrained optimization problem. Given ordered running variable values $x_1 \le x_2 \le \dots \le x_n$,

$$\begin{array}{ll} \text{maximize} & \Psi(\mathcal{I}_n(\boldsymbol{p})) \\ \text{over} & \boldsymbol{p} = (p_1, \dots, p_n) \in \mathcal{A} \\ \text{subject to} & n^{-1}\sum_{i=1}^n p_i = \widetilde{p} \quad \text{and} \quad n^{-1}\sum_{i=1}^n x_i p_i = \widetilde{xp} \end{array} \tag{3}$$

for some constants $\widetilde{p}$ and $\widetilde{xp}$. The first equality constraint in (3) is a budget constraint due to the cost of treatment and the second constraint is on the short-term gain mentioned above. We consider two different sets $\mathcal{A}$ in detail. The first is $[0, 1]^n$. The second is $\{\boldsymbol{p} \in [0, 1]^n \mid 0 \le p_1 \le p_2 \le \dots \le p_n \le 1\}$, which requires treatment probabilities to be nondecreasing in the running variable x. Such a monotonicity constraint prevents more qualified students from having a lower chance of getting a scholarship than less qualified ones or more loyal customers having a lower chance for a perk than others. It also eliminates perverse incentives for subjects to lower their x_i. To our knowledge, such a monotonicity constraint has not received much attention in the optimal design literature, though it is enormously appealing in our motivating applications.

When the efficiency criterion $\psi(\cdot)$ is concave in $\boldsymbol{p}$, then a solution to (3) can be found numerically via convex optimization (Metelkina and Pronzato (2017), Morrison and Owen (2022)). However, when x_i is univariate, we show the problem is tractable enough to provide a simple yet complete analytical characterization of the optimal p_i, even if the efficiency criterion is not concave. We show, under general conditions, that we can always find optimal treatment probabilities that are piecewise constant in x, with the number of pieces small and independent of n.

There is a well-developed literature for optimal experiment design in the presence of multiple objectives. Early examples of a constrained optimization problem of the form (3) were designed to account for several of the standard efficiency objectives simultaneously (Lee (1987), Lee (1988), Stigler (1971)). Läuter (1974), Läuter (1976) proposed maximizing a convex combination of efficiency objectives, a practice now typically referred to as a "compound" design approach. It is now well known (Clyde and Chaloner (1996), Cook and Wong (1994)) that in many problems with concave objectives, optimal constrained and compound designs are equivalent. In this paper, we provide another approach to reduce the constrained problem (3) to a compound problem that can handle the monotonicity constraint. At the same time, we provide simple ways to compute our optimal designs that are based directly on the parameters $\widetilde{p}$ and $\widetilde{xp}$ in our constrained formulation (3), and do not require specifying the Lagrange multipliers appearing in the corresponding compound problem. Those Lagrange multipliers involve ratios of information gain to economic gain where each of those quantities is only known up to a multiplicative constant.

Problems similar to (3) have received significant attention in the sequential design of clinical trials. Biased-coin designs, beginning with the simple procedure of Efron (1971), have been developed as a compromise between treatment balance and randomization; see Atkinson (2014) for a review. Covariate-adaptive biased-coin designs often replace the balance objective with an efficiency criterion such as D-optimality (Atkinson (1982), Rosenberger and Sverdlov (2008)). Response-adaptive designs also optimize for some efficiency objective but simultaneously seek to minimize the number of patients receiving the inferior treatment for ethical reasons (Hu and Rosenberger (2006)). Various authors such as Bandyopadhyay and Biswas (2001) and Hu, Zhu and Hu (2015) propose sequential designs to effectively navigate this trade-off. When they also account for covariate information, they are called covariate-adjusted response-adaptive (CARA) designs (Zhang and Hu (2009), Zhang et al. (2007)).

In the CARA literature especially, there has been significant recent interest in optimal design for nonlinear models (Biswas and Bhattacharya (2018), Metelkina and Pronzato (2017), Sverdlov, Rosenberger and Ryeznik (2013)). Unlike optimal designs in linear models such as (1), designs in nonlinear models can typically only be locally optimal, meaning that their optimality depends on the values of the unknown parameters (Chernoff (1953)). While we may be able to obtain increasingly reliable estimates of these parameters over time in sequential settings—further motivating response-adaptive designs—our motivating problems are primarily those where subjects are treated in batches. For instance, when measuring the impact of a scholarship on future educational attainment, we know the running variables for all subjects before designing the experiment, but it can take several years to collect a single set of responses on which to compute a parameter estimate. Response-adaptive and locally optimal designs are not well suited to this setting, and so we focus on optimal design under the linear model (1) in a nonsequential setting, which presents a sufficient challenge.

The existing literature on problems like (3) typically considers the running variable x to be random. For example, Section 7 of Owen and Varian (2020) investigates tie-breaker designs under the assumption that the running variable is either uniform or Gaussian, and exactly half the subjects are to be treated. They consider the typical three-level tie-breaker design where subjects with running variable x above some threshold Δ always get the treatment, subjects with running variable below $-\Delta$ never get the treatment, and the remaining subjects are randomized into treatment with probability $1/2$. They find that several c-optimality criteria of statistical efficiency—corresponding to minimizing the variance in estimating various linear combinations of treatment effect coefficients β_2 and β_3 in the two line model (1)—are monotonically increasing in the width Δ of the randomization window, with the RCT ($\Delta \to \infty$) most efficient and the RDD ($\Delta = 0$) least efficient. Conversely, short-term gain is decreasing in Δ. They also show the three-level design is optimal for any given level of short-term gain. In this article, we show strong advantages to moving away from that three-level design when the running variable is not symmetric, or we cannot treat half of the subjects.

Metelkina and Pronzato (2017) studied a further generalization of the optimal tie-breaker design problem, motivated by CARA designs. They allow for an arbitrary number $K > 2$ of treatments and consider locally optimal designs for possibly nonlinear models. Additionally, their running variable can be multidimensional and they allow for sequential treatment allocation schemes. In particular, their Example 1 is similar, though not quite identical, to a random-x generalization of (3). Crucially, however, their proof does not generalize to the case where we require the treatment probabilities be monotone. Even without the monotonicity constraint, we are able to provide a much sharper characterization of the solutions to our more specific problem (Section 3.1).

In Section 2, we pose a random-x generalization of the fixed-x problem (3). The solution is then a function $p(\cdot)$ mapping the running variable x to $[0, 1]$, specifying the probability of treatment for a subject with running variable x. We show this problem reduces precisely to (3) when the running variable distribution is assumed to be uniform on the known $\{x_1, \dots, x_n\}$. The random-x formulation is also more consistent with previous work (Metelkina and Pronzato (2017), Owen and Varian (2020)) and enables us to use classical results from hypothesis testing in Section 3 to precisely characterize optimal designs. In particular, Theorem 2 shows that, under the monotonicity constraint on p, there always exists a solution to (3) with a threshold t' such that $p(x_i)$ takes one value for all $x_i < t'$ and (when $\widetilde{xz} > 0$) a strictly larger value for all $x_i > t'$. The value of p at any $x_i = t'$ is between those other two values. Section 4 presents some results on the trade-off between a specific D-optimality efficiency criterion and short-term gain for these optimal designs. Examples of this trade-off for several running variable distributions are given in Section 5. That section also includes a fixed-x application based on Head Start—a government assistance program for low-income children—as well as a description of how to compute our optimal designs when x is either fixed or random. Finally, Section 6 summarizes the main results.

2. Random running variable.

Our random-x generalization of (3) assumes the running variables x_i are samples from a common distribution F, which we hereafter identify with the corresponding cumulative distribution function (CDF). It considers an information matrix that averages over both the random treatment assignments and randomness in the running variables. Before stating the problem, we briefly review the standard setting of optimal design in multiple linear regression models; see, for example, Atkinson, Donev and Tobias (2007) for further background. In the simplest case, the user assumes the standard linear model $y_i = \boldsymbol{x}_i^\mathsf{T}\boldsymbol{\beta} + \varepsilon_i$ with the goal of selecting covariate values $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^p$ to optimize an efficiency criterion that is a function of the information matrix $\mathcal{X}^\mathsf{T}\mathcal{X}$, where $\mathcal{X} \in \mathbb{R}^{n\times p}$ is the design matrix with ith row $\boldsymbol{x}_i^\mathsf{T}$. Perhaps the most common such criterion is D-optimality, which corresponds to maximizing $\log(\det(\mathcal{X}^\mathsf{T}\mathcal{X}))$. Another popular choice is c-optimality, which minimizes $c^\mathsf{T}(\mathcal{X}^\mathsf{T}\mathcal{X})^{-1}c$ for some choice of $c \in \mathbb{R}^p$. This can be interpreted as minimizing $\mathrm{Var}(c^\top \hat{\beta}|\mathcal{X})$, where $\hat{\beta}$ is the ordinary least squares estimator.

The study of optimal design is often simplified by the use of *design measures*. A design measure ξ is a probability distribution from which to generate the covariates $\boldsymbol{x}_i$. The relaxed optimal design problem involves selecting a design measure ξ instead of a finite number of covariate values $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$. The objective is to optimize for the desired functional of the *expected* information matrix $\mathcal{I}(\xi) \equiv \mathbb{E}_\xi[\mathcal{X}^\mathsf{T}\mathcal{X}]$ over some space Ξ of design measures ξ. For instance, a design measure ξ^* is D-optimal (for the relaxed problem) if $\xi^* \in \arg\max_{\xi\in\Xi} \det(\mathcal{I}(\xi))$, and c-optimal if $\xi^* \in \arg\min_{\xi\in\Xi} c^\top \mathcal{I}(\xi)^{-1}c$. The original optimal design problem restricts Ξ to only consist of discrete probability distributions supported on at most n distinct points with probabilities that are multiples of $1/n$.

For the tie-breaker design problem, our regression model (1) includes both the running variables x_i and the treatment indicators z_i as covariates. But the experimenter does not have control over the entire joint distribution of (x_i, z_i). The running variable is externally determined, so they can only specify the conditional distribution of the treatment indicator z_i given the running variable x_i. This conditional distribution is specified by a *design function* $p : \mathbb{R} \to [0, 1]$ such that $p(x) \equiv \Pr(z_i = 1|x_i = x)$. As mentioned above, we assume $x_i \sim F$ for a known, fixed distribution F. This allows us to drop subscripts i when convenient. For any two design functions p and p', we say $p = p'$ whenever $\Pr_F(\{x : p(x) = p'(x)\}) = 1$. We only need a minimal assumption on F, which can be continuous, discrete or neither.

ASSUMPTION 1. $0 < \mathrm{Var}_F(x) < \infty$ with $\mathbb{E}_F(x) = 0$.

The mean-centeredness part of Assumption 1 loses no generality, due to the translation invariance of estimation under the two line model (1). All expectations involving x hereafter omit the subscript F from all such expectations with the implicit understanding that $x \sim F$.

The random-x tie-breaker design problem is as follows:

$$
\begin{array}{ll}
\text{maximize} & \Psi(\mathcal{I}(p)) \\
\text{over} & p \in \mathcal{F} \\
\text{subject to} & \mathbb{E}_p(z) = \widetilde{z} \\
\text{and} & \mathbb{E}_p(xz) = \widetilde{xz}.
\end{array} \tag{4}
$$

Here, $\widetilde{z}$ and $\widetilde{xz}$ are constants analogous to $\bar{p}$ and $\overline{xp}$, respectively, in (3), $\mathcal{F}$ is a collection of design functions, and $\mathcal{I}(p)$ is the expected information matrix under the model (1), averaging over both $x \sim F$ and $z|x \sim p$. This problem can be viewed as a constrained relaxed optimal design problem under the regression model (1) where the set Ξ of allowable design measures is indexed by the design functions $p \in \mathcal{F}$.

To interpret the equality constraints in (4), it is helpful to note that

$$
\mathbb{E}_p(x^a z) = \mathbb{E}(x^a \mathbb{E}_p(z|x)) = \mathbb{E}(x^a(2p(x) - 1)) = 2\mathbb{E}(x^a p(x)) - \mathbb{E}(x^a) \tag{5}
$$

for any $a \geq 0$ with $\mathbb{E}(|x|^a) < \infty$. In particular, for each positive integer a, there exists an invertible linear mapping $\varphi_a : \mathbb{R}^{a+1} \to \mathbb{R}^{a+1}$ that does not depend on the design p and maps $(\mathbb{E}_p(z), \mathbb{E}_p(xz), \dots, \mathbb{E}_p(x^a z))$ to $(\mathbb{E}(p(x)), \mathbb{E}(xp(x)), \dots, \mathbb{E}(x^a p(x)))$. For example, $\varphi_1(x, y) = (1/2)(1 + x, y)$. Taking $a = 0$ in (5), we see that the constraint $\mathbb{E}_p(z) = \widetilde{z}$ in (4) is equivalent to a budget constraint requiring the expected proportion of subjects to be treated to be $(1 + \widetilde{z})/2$. Taking $a = 1$ in (5) shows that the second constraint $\mathbb{E}_p(xz) = \widetilde{xz}$ in (4) sets the expected level of short-term gain. In Section 2.2, we provide some guidance on how to choose $\widetilde{xz}$ in practice.

From computing the expected information matrix $\mathcal{I}(p)$ in Section 2.3, we will see that the problem (4) reduces to the finite-dimensional problem (3) when F is discrete, placing probability mass n^{-1} on each of the known running variable values $x_1, \dots, x_n$. Thus, to solve (3) it suffices to solve the problem (4) for any F satisfying Assumption 1, which must hold for any discrete distribution with finite support.

2.1. *Some design functions.* For convenience, we introduce some notation for certain forms of the design function p. We will commonly encounter designs of the form

$$p_A(x) \equiv \mathbf{1}(x \in A) \tag{6}$$

for a set $A \subseteq \mathbb{R}$. Another important special case consists of two-level designs,

$$p_{\ell,u,t}(x) \equiv \ell \mathbf{1}(x < t) + u \mathbf{1}(x \geq t), \tag{7}$$

for treatment probabilities $0 \leq \ell \leq u \leq 1$ and a threshold $t \in \overline{\mathbb{R}}$. For example, $p_{0,1,t}$ is a sharp RDD with threshold t, while for any t, $p_{\theta,\theta,t}$ is an RCT with treatment probability θ.

The condition $\ell \leq u$ ensures that $p(x)$ is nondecreasing in x; we refer to such designs as *monotone*. Under a monotone design, a subject cannot have a lower treatment probability than another subject with lower x. We also define a *symmetric* design to be one for which $p(-x) = 1 - p(x)$; for instance, p might be the CDF of a symmetric random variable. Finally, the three-level tie-breaker design from Owen and Varian (2020) is both monotone and symmetric and defined for $\Delta \in [0, 1]$ by

$$p_{3,\Delta}(x) \equiv 0.5 \times \mathbf{1}\big(|x| \leq \Delta\big) + \mathbf{1}(x > \Delta) \tag{8}$$

when F is the $\mathbb{U}(-1, 1)$ distribution. Note that for all Δ, $p_{3,\Delta}$ always treats half the subjects, that is, $\widetilde{z} = 0$. The generalization to other $\widetilde{z}$ and running variable distribution functions F is

$$p_{3;\widetilde{z},\Delta}(x) = 0.5 \times \mathbf{1}\big(a(\widetilde{z}, \Delta) < x < b(\widetilde{z}, \Delta)\big) + \mathbf{1}\big(x \geq b(\widetilde{z}, \Delta)\big), \tag{9}$$

where $a(\widetilde{z}, \Delta) = F^{-1}((1 - \widetilde{z})/2 - \Delta)$ and $b(\widetilde{z}, \Delta) = F^{-1}((1 - \widetilde{z})/2 + \Delta)$.

2.2. *Bounds on short-term gain.* Before studying optimal designs, we impose lower and upper bounds on the possible short-term gain constraints $\widetilde{xz}$ to consider, for each possible $\widetilde{z} \in (-1, 1)$. For an upper bound, we use $\widetilde{xz}_{\max}(\widetilde{z})$, the maximum $\widetilde{xz}$ that can be attained by *any* design function p satisfying the treatment fraction constraint $\mathbb{E}_p(z) = \widetilde{z}$. It turns out that this upper bound is always uniquely attained. If the running variable distribution F is continuous, it is uniquely attained by a sharp RDD. We remind the reader that uniqueness of a design function satisfying some property means that for any two design functions p and p' with that property, we must have $\Pr(p(x) = p'(x)) = 1$ under $x \sim F$.

LEMMA 1. *For any $\widetilde{z} \in [-1, 1]$ and running variable distribution F, there exists a unique design $p_{\widetilde{z}}$ satisfying*

$$\mathbb{E}_{p_{\widetilde{z}}}(z) = \widetilde{z} \quad \textit{and} \tag{10}$$

$$p_{\widetilde{z}}(x) = \begin{cases} 1, & x > t, \\ 0, & x < t, \end{cases} \tag{11}$$

for some $t \in \mathbb{R}$. Any p that satisfies the treatment fraction constraint (10) also satisfies

$$\mathbb{E}_p(xz) \leq \mathbb{E}_{p_{\widetilde{z}}}(xz) \equiv \widetilde{xz}_{\max}(\widetilde{z}) \tag{12}$$

with equality if and only if $p = p_{\widetilde{z}}$, that is, $\Pr(p(x) = p_{\widetilde{z}}(x)) = 1$ *under $x \sim F$.*

REMARK 1. Notice that Equation (11) does not specify $p_{\tilde{z}}(x)$ at $x = t$. If F is continuous, then any value for $p_{\tilde{z}}(t)$ yields an equivalent design function, but if F has an atom at $x = t$ then we will require a specific value for $p_{\tilde{z}}(t) \in [0, 1]$. We must allow F to have atoms to solve the finite-dimensional problem (3). While we specify $p_{\tilde{z}}(t)$ in the proof of Lemma 1 below, later results of this type do not give the values of design functions at such discontinuities.

REMARK 2. If F is continuous, then $p_{\tilde{z}}$ is an RDD: $p_{\tilde{z}} = p_{0,1,t}$ for $t = F^{-1}((1-\tilde{z})/2)$. We call the design $p_{\tilde{z}}$ a *generalized RDD* for general F satisfying Assumption 1.

REMARK 3. The threshold t in (11) is essentially unique. If there is an interval (t, s) with $\Pr(t < x < s) = 0$, then all step locations in $[t, s)$ provide equivalent generalized RDDs.

PROOF OF LEMMA 1. If $\tilde{z} \in \{-1, 1\}$, then the only design functions (again, up to uniqueness w.p.1 under $x \sim F$) are the constant functions $p(x) = 0$ and $p(x) = 1$, and the result holds trivially. Thus, we can assume that $\tilde{z} \in (-1, 1)$. By (5), the existence of $p_{\tilde{z}}$ follows by taking $t = \inf\{s : F(s) \geq (1-\tilde{z})/2\}$ and

$$p_{\tilde{z}}(t) = \begin{cases} 0 & \text{if } \Pr(x = t) = 0, \\ \dfrac{F(t) - (1-\tilde{z})/2}{\Pr(x = t)} & \text{if } \Pr(x = t) > 0. \end{cases}$$

To show (12), fix any design p satisfying (10) and notice that $\mathbb{E}(p(x) - p_{\tilde{z}}(x)) = 0$ means

$$\mathbb{E}\big(p(x)\mathbf{1}(x < t)\big) + \big(p(t) - p_{\tilde{z}}(t)\big)\Pr(x = t) + \mathbb{E}\big((p(x) - 1)\mathbf{1}(x > t)\big) = 0.$$

Then $\mathbb{E}(x(p_{\tilde{z}}(x) - p(x)))$ equals

$$\begin{aligned} &\mathbb{E}\big(-xp(x)\mathbf{1}(x < t)\big) + t\big(p_{\tilde{z}}(t) - p(t)\big)\Pr(x = t) + \mathbb{E}\big(x\big(1 - p(x)\big)\mathbf{1}(x > t)\big) \\ &\quad \geq t\big[\mathbb{E}\big(-p(x)\mathbf{1}(x < t)\big) + \big(p_{\tilde{z}}(t) - p(t)\big)\Pr(x = t) + \mathbb{E}\big(\big(1 - p(x)\big)\mathbf{1}(x > t)\big)\big] \\ &\quad = 0 \end{aligned}$$

with equality iff $(t - x)p(x)\mathbf{1}(x < t) = (x - t)(1 - p(x))\mathbf{1}(x > t) = 0$ for a set of x with probability one under F, that is, iff p satisfies (11) with probability one under $x \sim F$. □

By symmetry, the design that minimizes $\mathbb{E}_p(xz)$ over all designs p with $\mathbb{E}_p(z) = \tilde{z}$ is $p_{1,0,s}$ where $s = F^{-1}((1+\tilde{z})/2)$. Notice that $\mathbb{E}_{p_{1,0,s}}(xz) = \widetilde{xz}_{\min}(\tilde{z}) \equiv 2\mathbb{E}(x\mathbf{1}(x < s)) < 0$. We impose a stricter lower bound of $\widetilde{xz} \geq 0$ in the context of problem (4). This is motivated by the fact that the running variable x has mean 0 (Assumption 1), meaning that $\mathbb{E}_p(xz) = 0$ whenever the design function p is constant, corresponding to an RCT. Designs with $\widetilde{xz} < 0$ exist for all $\tilde{z} \in (-1, 1)$ but would not be relevant in our motivating applications, as they represent scenarios where subjects with *smaller* x are more preferentially treated than in an RCT. We hence define the *feasible input space* $\mathcal{J}$ by

$$\mathcal{J} \equiv \big\{(\tilde{z}, \widetilde{xz}) \mid -1 < \tilde{z} < 1, 0 \leq \widetilde{xz} \leq \widetilde{xz}_{\max}(\tilde{z})\big\} \subseteq \mathbb{R}^2. \tag{13}$$

Any design function p for which the moments $(\mathbb{E}_p(z), \mathbb{E}_p(xz))$ lie within the feasible input space $\mathcal{J}$ is referred to as an *input-feasible* design function.

If the design p is input-feasible, we can write $\mathbb{E}_p(xz) = \delta \cdot \widetilde{xz}_{\max}(\mathbb{E}_p(z))$ for some $\delta \in [0, 1]$. The parameter δ corresponds to the amount of additional short-term gain attained by the design p over an RCT, relative to the amount of additional short-term gain attained by the generalized RDD $p_{\mathbb{E}_p(z)}$ that treats the same proportion of subjects as p. For instance, $\delta = 0.4$ means that the design p has a short-term gain that is 40% of the way from that of an RCT to the maximum attainable short-term gain under the treatment fraction constraint.

2.3. *Expected information matrix and equivalence of D-optimality and c-optimality.* We now explicitly compute the expected information matrix

$$\mathcal{I}(p) = \sigma^{-2}\mathbb{E}(n^{-1}\mathcal{X}^{\mathsf{T}}\mathcal{X}) = \sigma^{-2}\begin{pmatrix} 1 & 0 & \mathbb{E}(z) & \mathbb{E}(xz) \\ 0 & \mathbb{E}(x^2) & \mathbb{E}(xz) & \mathbb{E}(x^2 z) \\ \mathbb{E}(z) & \mathbb{E}(xz) & 1 & 0 \\ \mathbb{E}(xz) & \mathbb{E}(x^2 z) & 0 & \mathbb{E}(x^2) \end{pmatrix} = \sigma^{-2}\begin{pmatrix} D & C \\ C & D \end{pmatrix}, \tag{14}$$

where

$$C = \begin{pmatrix} \mathbb{E}(z) & \mathbb{E}(xz) \\ \mathbb{E}(xz) & \mathbb{E}(x^2 z) \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & \mathbb{E}(x^2) \end{pmatrix}.$$

We have omitted the dependence of the expectations on the design p for brevity. Recall that $\mathcal{I}$ depends on F as well, though the experimenter can only control p. When $F = (1/n)\sum_{i=1}^n \delta_{x_i}$ and the running variable values $x_1, \dots, x_n$ are mean-centered, the expected information matrix $\mathcal{I}(p)$ in (14) is precisely the fixed-x information matrix $\mathcal{I}_n(p_1, \dots, p_n)$, identifying $p_i \equiv p(x_i)$. This shows that the random-x problem (4) is indeed strictly more general than the fixed-x problem (3). Equation (14) also shows that any efficiency objective $\Psi(\mathcal{I}(p))$ only depends on the treatment indicators z through their marginal distributions conditional on x, and not on their joint distribution. In the fixed-x context, this means that, for instance, one can assign treatments to obey an exact budget constraint $n^{-1}\sum_{i=1}^n z_i = \widetilde{p}$ by stratification (instead of independently given x) without changing the information matrix.

While we will characterize solutions to the optimal design problem (4) for any continuous efficiency criterion $\Psi(\cdot)$, in Section 4 we will prove some additional results for the D-optimality criterion $\Psi_D(\cdot) = \log(\det(\cdot))$. We now show that D-optimality is of particular interest to us, as it corresponds exactly with c-optimality for $c = (0, 0, 0, 1)^{\top}$, the primary efficiency criterion considered by Owen and Varian (2020).

Following that paper, we note we can assume $\sigma^2 = 1$ WLOG (as D-optimality is scale invariant). Then for (x_i, z_i) independent, by the law of large numbers and standard block matrix inversion formulas,

$$n \operatorname{Var}(\hat{\beta}_3 | \mathcal{X}) = n\big[(\mathcal{X}^{\top}\mathcal{X})^{-1}\big]_{44} \overset{\text{a.s.}}{\to} (\mathcal{I}^{-1})_{44} = \frac{M_{11}(p)}{\det(M(p))}, \tag{15}$$

where $M = M(p)$ is the Schur complement

$$M = D - CD^{-1}C = \begin{pmatrix} 1 - \mathbb{E}(z)^2 - \dfrac{\mathbb{E}(xz)^2}{\mathbb{E}(x^2)} & -\mathbb{E}(xz)\cdot\mathbb{E}(z) - \dfrac{\mathbb{E}(x^2 z)\mathbb{E}(xz)}{\mathbb{E}(x^2)} \\ -\mathbb{E}(xz)\cdot\mathbb{E}(z) - \dfrac{\mathbb{E}(x^2 z)\mathbb{E}(xz)}{\mathbb{E}(x^2)} & \mathbb{E}(x^2) - \mathbb{E}(xz)^2 - \dfrac{\mathbb{E}(x^2 z)^2}{\mathbb{E}(x^2)} \end{pmatrix}. \tag{16}$$

Equation (15) then shows that maximizing $\operatorname{Eff}(p) := \det(M(p))/M_{11}(p)$ corresponds to c-optimality for $c = (0, 0, 0, 1)^{\top}$, so it suffices to show that maximizing $\operatorname{Eff}(p)$ is equivalent to D-optimality. Before proceeding further, we present the following result to show that $\operatorname{Eff}(p)$ is well-defined for any input-feasible design p. To maintain the interpretability of $\operatorname{Eff}(\cdot)$ as an inverse asymptotic variance, we do not normalize $\operatorname{Eff}(\cdot)$ to lie between 0 and 1 (some authors define the efficiency of a design as a ratio to the most efficient design).

COROLLARY 1. *For any* $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$, $M_{11} = 1 - \widetilde{z}^2 - (\widetilde{xz})^2/\mathbb{E}(x^2) > 0$.

PROOF. See Appendix A. □

Noting $\det(\mathcal{I}) = \det(D)\det(M) = \mathbb{E}(x^2)\det(M)$, we see that maximizing $\det(\mathcal{I})$ (D-optimality) is equivalent to maximizing $\det(M)$. But M_{11} only depends on p through $\mathbb{E}_p(z)$ and $\mathbb{E}_p(xz)$, so any two input-feasible designs p and p' satisfying the equality constraints in (4) must have $M_{11}(p) = M_{11}(p')$. We conclude that the D-optimality criterion $\Psi_D(\mathcal{I}(p))$ and the c-optimality criterion $\mathrm{Eff}(p)$ yield the same solutions to (4) for all $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$.

3. Optimal design characterizations. To solve the constrained optimization problem (4), we begin by observing that the expected information matrix $\mathcal{I}(p)$, computed in (14), only depends on the design function p through the quantities $\mathbb{E}_p(z)$, $\mathbb{E}_p(xz)$ and $\mathbb{E}_p(x^2z)$. Then the same is true for any efficiency objective $\Psi(\mathcal{I}(p))$. Consequently, for any continuous Ψ we can write $\Psi(\mathcal{I}(p))) = g_\Psi(\mathbb{E}_p(z), \mathbb{E}_p(xz), \mathbb{E}_p(x^2z))$ for some continuous $g_\Psi : \mathbb{R}^3 \to \mathbb{R}$ that may depend on the running variable distribution F.

Fixing $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ and the set $\mathcal{F}$ of permissible design functions, we say a *feasible* design $p \in \mathcal{F}$ is one that satisfies the equality constraints in (4), that is, $\mathbb{E}_p(z) = \widetilde{z}$ and $\mathbb{E}_p(xz) = \widetilde{xz}$. Thus, the efficiency criterion $\Psi(\mathcal{I}(p))$ can only vary among feasible designs p through the single quantity $\mathbb{E}_p(x^2z)$. Furthermore, any two feasible designs p and q with $\mathbb{E}_p(x^2z) = \mathbb{E}_q(x^2z)$ must have the same efficiency. Thus, we can break down the problem (4) into two steps. First, we find a solution

$$\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi) \in \underset{a \in I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})}{\arg\max}\, g_\Psi(\widetilde{z}, \widetilde{xz}, a), \tag{17}$$

where

$$I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz}) = \{\mathbb{E}_p(x^2z) | \mathbb{E}_p(z) = \widetilde{z}, \mathbb{E}_p(xz) = \widetilde{xz} \text{ for some } p \in \mathcal{F}\} \subseteq [-\mathbb{E}(x^2), \mathbb{E}(x^2)] \tag{18}$$

is the set of values of $\mathbb{E}_p(x^2z)$ attainable by some feasible design $p \in \mathcal{F}$. Then we must find a feasible design $p \in \mathcal{F}$ that satisfies $\mathbb{E}_p(x^2z) = \widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$.

The next result shows that when $\mathcal{F}$ is convex, $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$ is an interval. For our two choices of $\mathcal{F}$ of interest, Propositions 1 and 2 will show it is a closed interval, so (17) will always have a solution when $\Psi(\cdot)$ is continuous.

LEMMA 2. *Suppose feasible designs p, $p' \in \mathcal{F}$ satisfy $\mathbb{E}_p(x^2z) \le \mathbb{E}_{p'}(x^2z)$, where $\mathcal{F}$ is convex. Then if $\mathbb{E}_p(x^2z) \le \gamma \le \mathbb{E}_{p'}(x^2z)$, there exists feasible $p^{(\gamma)} \in \mathcal{F}$ with $\mathbb{E}_{p^{(\gamma)}}(x^2z) = \gamma$.*

PROOF. If $\mathbb{E}_{p'}(x^2z) = \mathbb{E}_p(x^2z)$, then either of them is a suitable $p^{(\gamma)}$. Otherwise, take $\lambda \in [0, 1]$ so that $\gamma = \lambda\mathbb{E}_p(x^2z) + (1-\lambda)\mathbb{E}_{p'}(x^2z)$. Then $p^{(\gamma)} = \lambda p + (1-\lambda)p'$ is in $\mathcal{F}$ by convexity and, by direct computation of the moments $\mathbb{E}_p(x^a z)$ for $a \in \{0, 1, 2\}$, feasible with $\mathbb{E}_p(x^2z) = \gamma$. □

The endpoints of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$ can be computed as the optimal values of the following constrained optimization problems:

$$\begin{array}{llcll} \text{maximize} & \mathbb{E}_p(x^2z) & \quad & \text{minimize} & \mathbb{E}_p(x^2z) \\ \text{over} & p \in \mathcal{F} & & \text{over} & p \in \mathcal{F} \\ \text{subject to} & \mathbb{E}_p(z) = \widetilde{z} & & \text{subject to} & \mathbb{E}_p(z) = \widetilde{z} \\ \text{and} & \mathbb{E}_p(xz) = \widetilde{xz} & & \text{and} & \mathbb{E}_p(xz) = \widetilde{xz}. \end{array} \tag{19}$$

Given solutions $p_{\max}$ and $p_{\min}$ to the problems (19), Lemma 2 shows that the design

$$p_{\text{opt}}(x) = \begin{cases} p_{\max}(x) & \mathbb{E}_{p_{\max}}(x^2 z) \le \widetilde{x^2 z}^*(\widetilde{z}, \widetilde{xz}; \Psi), \\ p_{\min}(x) & \mathbb{E}_{p_{\min}}(x^2 z) \ge \widetilde{x^2 z}^*(\widetilde{z}, \widetilde{xz}; \Psi), \\ \lambda p_{\min}(x) + (1-\lambda) p_{\max}(x) & \text{else}, \end{cases} \tag{20}$$

solves the problem (4) for $\lambda = (\mathbb{E}_{p_{\max}}(x^2 z) - \widetilde{x^2 z}^*(\widetilde{z}, \widetilde{xz}; \Psi))/(\mathbb{E}_{p_{\max}}(x^2 z) - \mathbb{E}_{p_{\min}}(x^2 z))$. If $\mathbb{E}_{p_{\max}}(x^2 z) = \mathbb{E}_{p_{\min}}(x^2 z)$, then all feasible designs p have the same efficiency, so any one of them is optimal.

The remainder of this section is concerned with characterizing the solutions to the problems (19) for two specific choices of design function classes $\mathcal{F}$: the set of all measurable functions into [0, 1], and the set of all such *monotone* functions. For these two choices of $\mathcal{F}$, solutions $p_{\max}$ and $p_{\min}$ exist for any $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ and are unique. Our argument uses extensions of the Neyman–Pearson lemma (Neyman and Pearson (1933)) in hypothesis testing. These extensions are in Dantzig and Wald (1951), whose two authors discovered the relevant results independently of each other. We use a modern formulation of their work, adapting the presentation by Lehmann and Romano (2005).

LEMMA 3. *Consider any measurable $h_1, \dots, h_{m+1} : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}(|h_i(x)|) < \infty$, $i = 1, \dots, m+1$. Define $S \subseteq \mathbb{R}^m$ to be the set of all points $c = (c_1, \dots, c_m)$ such that*

$$\mathbb{E}\big(p(x) h_i(x)\big) = c_i, \quad i = 1, \dots, m, \tag{21}$$

for some $p \in \mathcal{F}$, where $\mathcal{F}$ is some collection of measurable functions from $\mathbb{R}$ into [0, 1]. *For each $c \in S$, let $\mathcal{F}_c$ be the set of all $p \in \mathcal{F}$ satisfying* (21). *If $\mathcal{F}$ is such that*

$$S' = \big\{(c, c_{m+1}) \in \mathbb{R}^{m+1} | c \in S, c_{m+1} = \mathbb{E}\big(p(x) h_{m+1}(x)\big) \text{ for some } p \in \mathcal{F}_c\big\}$$

is closed and convex and $c \in \operatorname{int} S$, then:

1. *There exists $p \in \mathcal{F}_c$ and $k_1, \dots, k_m \in \mathbb{R}$ such that*

$$p \in \arg\max_{q \in \mathcal{F}} \mathbb{E}\left(q(x)\left(h_{m+1}(x) - \sum_{i=1}^m k_i h_i(x)\right)\right), \quad \text{and} \tag{22}$$

2. *$p \in \arg\max_{q \in \mathcal{F}_c} \mathbb{E}(q(x) h_{m+1}(x))$ if and only if $p \in \mathcal{F}_c$ satisfies* (22) *for some $k_1, \dots, k_m$.*

PROOF. Claim 1 and necessity of (22) in claim 2 follows from the proof of part (iv) of Theorem 3.6.1 in Lehmann and Romano (2005), which uses the fact that S' is closed and convex to construct a separating hyperplane in $\mathbb{R}^{m+1}$. Sufficiency of (22) in claim 2 follows from part (ii) of that theorem, and is often called the method of undetermined multipliers. □

Lemma 3 equates a constrained optimization problem (item 2) and a compound optimization problem (item 1). Unlike typical equivalence theorems, it does not require $\mathcal{F}$ to be the set of all measurable design functions, and uses an entirely different proof technique. Following Whittle (1973), equivalence theorems in optimal design are now popularly proven using the concept of Fréchet derivatives on the space of design functions (measures). However, such approaches often do not apply when $\mathcal{F}$ is restricted to be the set of all monotone design functions. Most relevant to our problem, the proof of Corollary 2.1 in the supplement of Metelkina and Pronzato (2017) involves Fréchet derivatives in the direction of design functions supported at a single value of x, which are not monotone. However, the use of Lemma 3 requires an objective linear in p, where typical equivalence theorems only require concavity.

3.1. *Globally optimal designs.* We now solve the design problem (4) in the case that $\mathcal{F}$ is the set of all measurable functions $p:\mathbb{R}\to[0,1]$. We first explain how the results of Metelkina and Pronzato (2017) do not adequately do so already. Identifying our design functions p with their design measures ξ, Corollary 2.1 of Metelkina and Pronzato (2017) does not provide any information about what an optimal solution $p_{\text{opt}}(x)$ to (4) would be for values of x where $G_1(p_{\text{opt}}(\cdot);x)=G_2(p_{\text{opt}}(\cdot);x)$. Here, G_1 and G_2 are quantities derived from the aforementioned Fréchet derivatives, depending on the constraints $\widetilde{z}$ and $\widetilde{xz}$. Unfortunately, this lack of information about p_{opt} holds for all x both in their Example 1 and in our setting. Their example skirts this limitation by noting some moment conditions on p_{opt} implied by the equality $G_1=G_2$ when the running variable is uniform and the efficiency criterion is D-optimality, and then manually searching for some parametric forms of p_{opt} for which it is possible to satisfy these conditions. By contrast, the results in this section apply Lemma 3 with $\mathcal{F}$ the set of all design functions, and show a simple stratified design function is always optimal for any running variable distribution F and continuous efficiency criterion. This enables optimal designs to be systematically and efficiently constructed (Section 5).

We will apply Lemma 3 with $m=2$ constraints pertaining to $h_1(x)=1$ and $h_2(x)=x$. Our objective function is based on $h_3(x)=x^2$. When the running variable distribution F is continuous, recalling the notation (6), the solutions to (19) take the forms $p_{\max}=p_{[a_1,a_2]^c}$ and $p_{\min}=p_{[b_1,b_2]}$ for some intervals $[a_1,a_2]$ and $[b_1,b_2]$ (here $[a_1,a_2]^c$ denotes $\mathbb{R}\setminus[a_1,a_2]$).

PROPOSITION 1. *Let $\mathcal{F}$ be the set of all measurable functions from $\mathbb{R}$ into $[0,1]$. For any $(\widetilde{z},\widetilde{xz})\in\mathcal{J}$, there exist unique solutions $p_{\max}$ and $p_{\min}$ to the optimization problems (19). These solutions are the unique feasible designs satisfying*

$$p_{\max}(x)=\begin{cases}1, & x\notin[a_1,a_2],\\ 0, & x\in(a_1,a_2),\end{cases}\quad\textit{and}\quad p_{\min}(x)=\begin{cases}1, & x\in(b_1,b_2),\\ 0, & x\notin[b_1,b_2],\end{cases}\tag{23}$$

for some $a_1\le a_2$ and $b_1\le b_2$, which depend on $(\widetilde{z},\widetilde{xz})$ and can be infinite if $\widetilde{xz}=\widetilde{xz}_{\max}(\widetilde{z})$.

PROOF. If $\widetilde{xz}=\widetilde{xz}_{\max}(\widetilde{z})$, then the proposition follows by Lemma 1 and taking $a_1=-\infty$, $a_2=t=b_1$ and $b_2=\infty$. Thus, we can assume that $\widetilde{xz}<\widetilde{xz}_{\max}(\widetilde{z})$. We give the proof for $p_{\max}$ in detail. The argument for $p_{\min}$ is completely symmetric.

As noted above, we are in the setting of Lemma 3 with $m=2$, $h_1(x)=1$, $h_2(x)=x$ and $h_3(x)=x^2$. The collection $\mathcal{F}$ here is the set of all measurable functions from $\mathbb{R}$ into $[0,1]$, so the corresponding S' is closed and convex, as shown in part (iv) of Theorem 3.6.1 in Lehmann and Romano (2005). By Lemma 1 and (5), we can write $\operatorname{int}S=\varphi_1(\mathcal{T})$ where φ_1 is defined in the discussion around (5) and

$$\mathcal{T}=\{(\widetilde{z},\widetilde{xz})\mid -1<\widetilde{z}<1,\widetilde{xz}_{\min}(\widetilde{z})<\widetilde{xz}<\widetilde{xz}_{\max}(\widetilde{z})\}.$$

Hence, our previous assumption $\widetilde{xz}<\widetilde{xz}_{\max}(\widetilde{z})$ ensures $c=\varphi_1(\widetilde{z},\widetilde{xz})\in\operatorname{int}S$.

With the conditions of Lemma 3 satisfied, we now show that (22) is equivalent to (23) for any feasible $p_{\max}$. A feasible design $p_{\max}$ satisfies (22) iff $p_{\max}\in\arg\max_{q\in\mathcal{F}}\mathbb{E}(q(x)(x^2-k_1x-k_2))$ for some k_1,k_2, or equivalently

$$p_{\max}(x)=\begin{cases}1, & x^2-k_1x-k_2>0,\\ 0, & x^2-k_1x-k_2<0,\end{cases}\tag{24}$$

(cf. part (ii) of Theorem 3.6.1 in Lehmann and Romano (2005)). If $x^2-k_1x-k_2$ has no real roots, then $p_{\max}(x)=1$ for all x, contradicting $\widetilde{z}<1$. Thus, we write $x^2-k_1x-k_2=(x-a_1)(x-a_2)$ for some (real) $a_1\le a_2$, showing that (24) is equivalent to (23). We can now conclude, by the second claim in Lemma 3, that the set of optimal solutions to (19) contains precisely those feasible designs satisfying (23). Furthermore, the first claim of Lemma 3 ensures that such a design must exist. It remains to show only one feasible design can satisfy (23); in Appendix B, we provide a direct argument, which does not rely on Lemma 3. □

REMARK 4. The necessity and sufficiency results of Proposition 1 do follow from Corollary 2.1 of Metelkina and Pronzato (2017). We again identify our design functions p with their design measures ξ and take $\psi(\xi) = \int x^2 \,\mathrm{d}\xi(x)$, which can be written as an affine function $\boldsymbol{\Psi}(\cdot)$ of the expected information matrix $\mathcal{I}(\xi)$. As discussed at the beginning of this section, however, the form of a solution to problem (4), cannot be constructed from their corollary without the reduction to (19) and applying (20), so that ψ is as above rather than something like D-optimality. We have also shown a stronger uniqueness result than Section 2.3.3 of Metelkina and Pronzato (2017), which only applies when the running variable distribution F has a density with respect to Lebesgue measure. Our Lemma 3 also provides an existence guarantee that does not rely on strict concavity of $\boldsymbol{\Psi}(\cdot)$ on the set of positive definite matrices; this is violated by the affine choice we need here.

As we will see in Section 5, when $\widetilde{z} < 0$ we frequently encounter $p_{\mathrm{opt}} = p_{\max}$ under D-optimality. In this case, it is intuitive that there is an efficiency advantage to strategically allocate the rare level $z = 1$ at both high and low x, compared to a three-level tie-breaker. But such a design is usually unacceptable in our motivating problems. We will thus constrain $\mathcal{F}$ to the set of monotone design functions in Section 3.2.

Before doing that, we present an alternative solution to (4) assuming the running variable distribution has an additional moment. This result shows that when the running variable is continuous, an optimal design with no randomization always exists. However, randomized assignment becomes essential once we restrict our attention to monotone designs in Section 3.2, as the only nonrandomized monotone designs are generalized RDDs (Remark 2).

THEOREM 1. *Suppose* $\mathbb{E}(|x|^3) < \infty$. *Then when* $\mathcal{F}$ *is the set of all measurable design functions, for any* $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ *there exists a solution to* (4) *with*

$$p_{\mathrm{opt}}(x) = \begin{cases} 1, & x < a_1, \\ 0, & a_1 < x < a_2, \\ 1, & a_2 < x < a_3, \\ 0, & x > a_3, \end{cases} \tag{25}$$

for some $a_1 \le a_2 \le a_3$, *which are finite unless* p_{opt} *is one of the designs* $p_{\max}$ *or* $p_{\min}$ *in* (23).

PROOF. The solutions to (4) are precisely the feasible design functions $p \in \mathcal{F}$ where $\mathbb{E}_p(x^2 z)$ is a solution to (17). Fix any such solution $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$ to (17); it suffices to find a feasible design p_{opt} with $\mathbb{E}_{p_{\mathrm{opt}}}(x^2 z) = \widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$. If $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$ is the lower (resp., upper) endpoint of the interval $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$, then by Proposition 1, the unique feasible design with $\mathbb{E}_p(x^2 z) = \widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$ is the design $p_{\min}$ (resp., $p_{\max}$). Then the result follows with $a_1 = -\infty$ (resp., $a_3 = \infty$).

Otherwise, $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$ is in the interior of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$, and we aim to apply Lemma 3 with $m = 3$, $h_i(x) = x^i$ for $i \in \{1, 2, 3\}$, and $c = \varphi_2(\widetilde{z}, \widetilde{xz}, \widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi))$. With S' closed and convex as shown in Proposition 1, we only need to show $c \in \operatorname{int} S$. With the interior of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$ being nonempty, there is more than one feasible design and the uniqueness result of Lemma 1 indicates that we must have $\widetilde{xz} < \widetilde{xz}_{\max}(\widetilde{z})$. With $\widetilde{z} \in (-1, 1)$ by assumption and $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi) \in \operatorname{int} I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$, indeed $c \in \operatorname{int} S$.

Applying Lemma 3, we know that there exists a feasible design p_{opt} with $\mathbb{E}_{p_{\mathrm{opt}}}(x^2z) = \widetilde{x^2z}^*(\tilde{z}, \widetilde{xz}; \Psi)$ and $p_{\mathrm{opt}}(x) \in \arg\max_{q\in\mathcal{F}} \mathbb{E}(q(x)(-x^3 - k_1x^2 - k_2x - k_3))$ for some k_1, k_2, k_3. If $f(x; k_1, k_2, k_3) \equiv -x^3 - k_1x^2 - k_2x - k_3$ had only one real root a_1, then the negative leading coefficient indicates $f(x; k_1, k_2, k_3) > 0$ when $x < a_1$ and $f(x; k_1, k_2, k_3) < 0$ when $x > a_1$. This would imply $p_{\mathrm{opt}}(x)$ is a design that always treats all subjects with $x < a_1$ and never treats any subject with $x > a_1$, which cannot be input-feasible. We conclude $f(x; k_1, k_2, k_3) = -(x - a_1)(x - a_2)(x - a_3)$ for some finite $a_1 \le a_2 \le a_3$, which are the roots of $f(x; k_1, k_2, k_3)$. This shows the existence of p_{opt} of the form (25). □

3.2. *Imposing a monotonicity constraint.* We now apply Lemma 3 to solve (4) in the case of principal interest, where $\mathcal{F}$ is the set of all *monotone* design functions. Note that the lower bound $\widetilde{xz} \ge 0$ that we imposed in Section 2.2 does not exclude any monotone designs. If $p(x)$ is monotone, then x and z necessarily have a nonnegative covariance $\mathbb{E}_p(xz)$.

Our argument follows the outline of Section 3.1. Suppose that $p^\dagger_{\max}$ and $p^\dagger_{\min}$ are solutions to (19) with $\mathcal{F}$ the set of monotone design functions, which we distinguish from the optimal designs $p_{\max}$ and $p_{\min}$ of Section 3.1. As $\mathcal{F}$ is convex, Lemma 2 applies, and thus a solution $p^\dagger_{\mathrm{opt}}$ to (4) is given by (20), replacing $p_{\max}$ and $p_{\min}$ by $p^\dagger_{\max}$ and $p^\dagger_{\min}$, respectively. Note that $\widetilde{x^2z}^*(\tilde{z}, \widetilde{xz})$ may differ from its value in Section 3.1 since $\mathcal{F}$ has changed.

We now characterize the designs $p^\dagger_{\max}$ and $p^\dagger_{\min}$. As in Proposition 1, these designs always exist and are unique for any $(\tilde{z}, \widetilde{xz}) \in \mathcal{J}$. When F is continuous, these are monotone two-level designs $p^\dagger_{\max} = p_{\ell,1,t}$ and $p^\dagger_{\min} = p_{0,u,s}$ as defined in (7). For general F, the designs $p^\dagger_{\max}$ and $p^\dagger_{\min}$ may differ from these designs at at the single discontinuity.

PROPOSITION 2. *For any $(\tilde{z}, \widetilde{xz}) \in \mathcal{J}$, there exist unique solutions $p^\dagger_{\max}$ and $p^\dagger_{\min}$ to the optimization problems (19), when $\mathcal{F}$ is the set of all monotone design functions. These solutions are the unique feasible designs satisfying*

$$(26)\qquad p^\dagger_{\max}(x) = \begin{cases} \ell, & x < t, \\ 1, & x > t, \end{cases} \quad \text{and} \quad p^\dagger_{\min}(x) = \begin{cases} 0, & x < s, \\ u, & x > s, \end{cases}$$

for some $\ell, u \in [0, 1]$ and constants s, t, which all depend on $(\tilde{z}, \widetilde{xz})$, where s and t may be infinite if $\widetilde{xz} = 0$.

PROOF. If $\widetilde{xz} = 0$, the only feasible monotone design is the fully randomized design $p(x) = (1 + \tilde{z})/2$, and the theorem holds trivially with $p^\dagger_{\max} = p^\dagger_{\min}$, $t = -s = \infty$ and $\ell = u = (1 + \tilde{z})/2$. Likewise, if $\widetilde{xz} = \widetilde{xz}_{\max}(\tilde{z})$ then the desired results follow by Lemma 1 (take $\ell = 0$, $u = 1$ and $s = t$ with t as in Lemma 1). Thus, we assume that $0 < \widetilde{xz} < \widetilde{xz}_{\max}(\tilde{z})$. Again, we only write out the argument for $p^\dagger_{\max}$; the proof for $p^\dagger_{\min}$ is completely analogous.

Once again, we are in the setting of Lemma 3 with $m = 2$, $h_1(x) = 1$, $h_2(x) = x$ and $h_3(x) = x^2$. The only difference from Proposition 1 is the definition of $\mathcal{F}$, so we must verify that the conditions on the corresponding S' and S are satisfied. Since $\mathbb{E}(x^a p(x))$ is linear in p for all a, and any convex combination of monotone functions is monotone (cf. Lemma 2), S' is convex. Now suppose that c_0 is a limit point of S'. Then there exists a sequence $p_1, p_2, \dots \in \mathcal{F}$ with $(\mathbb{E}(p_n(x)), \mathbb{E}(xp_n(x)), \mathbb{E}(x^2p_n(x))) \to c_0$ as $n \to \infty$. As $\mathcal{F}$ is sequentially compact, there exists a subsequence p_{n_i} and $p \in \mathcal{F}$ with $p_{n_i} \to p$ pointwise. But then $(\mathbb{E}(p(x)), \mathbb{E}(xp(x)), \mathbb{E}(x^2p(x))) = c_0$ by dominated convergence, so S' is closed. Finally, $\operatorname{int} S = \varphi_1(\mathcal{T}^\dagger)$ where

$$\mathcal{T}^\dagger \equiv \{(\tilde{z}, \widetilde{xz}) \mid -1 < \tilde{z} < 1, 0 < \widetilde{xz} < \widetilde{xz}_{\max}(\tilde{z})\} = \operatorname{int} \mathcal{J}.$$

Hence, the assumption that $0 < \widetilde{xz} < \widetilde{xz}_{\max}(\tilde{z})$ ensures that $c = \varphi_1(\tilde{z}, \widetilde{xz}) \in \operatorname{int} S$.

With the conditions of Lemma 3 once again satisfied, we now show that (22) is equivalent to (26) for any (monotone) feasible $p^{\dagger}_{\max}$. First, assume feasible $p^{\dagger}_{\max}$ satisfies (22), that is, $p^{\dagger}_{\max} \in \arg\max_{q\in\mathcal{F}} \mathbb{E}(q(x)(x^2 - k_1 x - k_2))$ for some k_1, k_2. The polynomial $x^2 - k_1 x - k_2$ having no real roots would mean this condition is equivalent to $p^{\dagger}_{\max}(x) = 1$, contradicting $\widetilde{z} < 1$. Hence, we can factor $x^2 - k_1 x - k_2 = (x - r)(x - t)$ for some $r \le t$. Considering the sign of $(x - r)(x - t)$ and monotonicity of any $q \in \mathcal{F}$, we see

$$\begin{aligned}
&\mathbb{E}\big(q(x)(x-r)(x-t)\big)\\
&\quad\le \mathbb{E}\big(q(r)(x-r)(x-t)\mathbf{1}(x<r)\big) + \mathbb{E}\big(q(r)(x-r)(x-t)\mathbf{1}(r\le x<t)\big)\\
&\qquad + \mathbb{E}\big((x-r)(x-t)\mathbf{1}(x\ge t)\big).
\end{aligned}$$

This inequality is strict unless $q(x) = 1$ for almost every $x > t$ and

$$\big(q(r) - q(x)\big)(x-r)(x-t)\mathbf{1}(x<r) = \big(q(x) - q(r)\big)(x-r)(x-t)\mathbf{1}(r\le x<t) = 0$$

with probability one under F, that is, $q(x) = q(r) = \ell$ for some $\ell \in [0, 1]$ and almost every $x < t$. Therefore, any design in $\arg\max_{q\in\mathcal{F}} \mathbb{E}(q(x)(x^2 - k_1 x - k_2))$ must satisfy the first condition in (26). Conversely, if a feasible, monotone $p^{\dagger}_{\max}$ satisfies (26) then let $r_1 \le t$ be such that $g(r_1) = \mathbb{E}((x - r_1)(x - t)\mathbf{1}(x < t)) = 0$. Such r_1 exists since assuming WLOG that $\Pr(x < t) > 0$, $g(\cdot)$ is continuous on $(-\infty, t]$ with $-\infty = \lim_{k_1\downarrow-\infty} g(k_1) < 0 \le g(t)$. Considering the signs of $(x - r_1)(x - t)$, we get for any $p \in \mathcal{F}$,

$$\begin{aligned}
&\mathbb{E}\big(\big(p^{\dagger}_{\max}(x) - p(x)\big)(x-r_1)(x-t)\big)\\
&\quad= \mathbb{E}\big(\big(1-p(x)\big)(x-r_1)(x-t)\mathbf{1}(x\ge t)\big)\\
&\qquad + \mathbb{E}\big(\big(\ell-p(x)\big)(x-r_1)(x-t)\mathbf{1}(x<t)\big)\\
&\quad\ge \mathbb{E}\big(\big(1-p(x)\big)(x-r_1)(x-t)\mathbf{1}(x\ge t)\big) + \big(\ell - p(r_1)\big)g(r_1)\\
&\quad= \mathbb{E}\big(\big(1-p(x)\big)(x-r_1)(x-t)\mathbf{1}(x\ge t)\big) \ge 0
\end{aligned}$$

and so $p^{\dagger}_{\max}$ satisfies (22) with $k_1 = r_1 + t$, and $k_2 = -tr_1$. The second claim in Lemma 3 then ensures that the set of optimal solutions to (19) consists of precisely those feasible, monotone designs satisfying (26). Such a design must exist by the first claim of Lemma 3. The remaining uniqueness claims are shown in Appendix C. □

Analogous to Theorem 1, if we assume the running variable has a third moment then we have a solution to (4) of a simpler form than (20). When the running variable x is continuous, this solution will be a two-level design $p_{\ell',u',t'}$ for some $0 \le \ell' \le u' \le 1$. In general, when x is not continuous, we may need a different treatment probability at the discontinuity t'.

THEOREM 2. *Suppose $\mathbb{E}(|x|^3) < \infty$. Then when $\mathcal{F}$ is the set of all monotone design functions, for any $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ there exists a solution to (4) with*

$$p^{\dagger}_{\mathrm{opt}}(x) = \begin{cases} \ell', & x < t', \\ u', & x > t', \end{cases} \tag{27}$$

for some $0 \le \ell' \le u' \le 1$ and $t' \in \mathbb{R}$.

PROOF. The proof structure is similar to that of Theorem 1. Fix any solution $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi)$ to (17). If it is an endpoint of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$, then the unique solution to (4) is $p^{\dagger}_{\max}$ or $p^{\dagger}_{\min}$ from Proposition 2, which takes the form (27) with $u' = 1$ or $\ell' = 0$, respectively. Otherwise, we apply Lemma 3 with $m = 3$, $h_i(x) = x^i$ for $i \in \{1, 2, 3\}$, and $c =$

$\varphi_2(\widetilde{z}, \widetilde{xz}, \widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi))$. The lemma applies since S' is closed and convex from the proof of Proposition 2, and $c \in \operatorname{int} S$ since our assumption that $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \Psi) \in \operatorname{int} I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$ indicates there is more than one feasible design, so $0 < \widetilde{xz} < \widetilde{xz}_{\max}(\widetilde{z})$.

Applying Lemma 3, we see that there exists a design $p^{\dagger}_{\text{opt}}$ that solves (4) with $p^{\dagger}_{\text{opt}} \in \arg\max_{q \in \mathcal{F}} \mathbb{E}(q(x) f(x; k_1, k_2, k_3))$ for some k_1, k_2, k_3. Here, as in the proof of Theorem 1, $f(x) = f(x; k_1, k_2, k_3) := -(x^3 + k_1 x^2 + k_2 x + k_3)$. We show this implies this particular solution $p^{\dagger}_{\text{opt}}$ is of the form (27) using the following claim.

Claim: Suppose $\operatorname{sign}(f(x)) = \mathbf{1}(x < a) - \mathbf{1}(x > a)$ w.p.1 for some $a \in \mathbb{R}$ and $\mathcal{F}$ is the set of monotone design functions. Then $p \in \arg\max_{q \in \mathcal{F}} \mathbb{E}(q(x) f(x))$ implies $p(x) = p(a)$ w.p.1.

Proof of claim: For any monotone design p, we can define $\tilde{p}(x) = \min(p(x), p(a))$ so that

$$\begin{aligned}\mathbb{E}((\tilde{p}(x) - p(x)) f(x) &= \mathbb{E}((\tilde{p}(x) - p(x)) f(x)\mathbf{1}(x \geq a)) \\ &= -\mathbb{E}((p(x) - p(a)) f(x)\mathbf{1}(x \geq a))\end{aligned}$$

is nonnegative, and zero iff $p(x) = p(a)$ for almost all $x \geq a$. Similarly, by considering $\tilde{p}(x) = \max(p(x), p(a))$, we conclude $p(x) = p(a)$ for almost all $x < a$.

We notice that f has either one real root a_1 or three real roots a_1, a_2, a_3. If a_1 is the only root, we know $f(x) < 0$ when $x > a_1$ and $f(x) > 0$ when $x < a_1$, since the leading coefficient of f is negative. Thus, we can apply the claim directly to show that $p^{\dagger}_{\text{opt}}$ is constant, in particular of the form (27) with $\ell' = u'$. If there are three real roots, we show $p^{\dagger}_{\text{opt}}$ is of this form with $t' = a_2$. Let $F_<$ ($F_>$) be the conditional distribution of x given $x < a_2$ ($x > a_2$), so

$$\mathbb{E}_F(q(x) f(x)) = \mathbb{E}_{F_<}(q(x) f(x)) \Pr(x < a_2) + \mathbb{E}_{F_>}(q(x) f(x)) \Pr(x > a_2).$$

We conclude the condition $p^{\dagger}_{\text{opt}} \in \arg\max_{q \in \mathcal{F}} \mathbb{E}_F(q(x) f(x; k_1, k_2, k_3))$ implies $p^{\dagger}_{\text{opt}}(x) = p^{\dagger}_{\text{opt}}(a_1)$ for almost all $x < a_2$ and $p^{\dagger}_{\text{opt}}(x) = p^{\dagger}_{\text{opt}}(a_3)$ for almost all $x > a_2$ by applying the claim twice (once for $F_<$, once for $F_>$). □

In general, the optimal designs derived in Theorems 1 and 2 are not unique when $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}, \Psi)$ is not on the boundary of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$. For example, in nondegenerate cases the solution $p^{\dagger}_{\text{opt}}$ in (27) typically has two levels, while the solution in (20) (with the monotonicity constraint) will have three levels. As another example, the three-level tie-breaker found by Owen and Varian (2020) to be optimal when F is uniform and $\widetilde{z} = 0$ does not take the form (25) whenever $\widetilde{xz} < \widetilde{xz}_{\max}(0)$. Conversely, Propositions 1 and 2 guarantee a unique optimal design when $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}, \Psi)$ is one of the endpoints of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$.

4. Exploration-exploitation trade-off. As discussed in Section 1, Owen and Varian (2020) showed that when $\widetilde{z} = 0$ and $F \sim \mathbb{U}(-1, 1)$, the efficiency (under their criterion $\operatorname{Eff}(\cdot)$) of the three-level tie-breaker (8) is monotonically increasing in the width Δ of the randomization window. As Δ is a strictly decreasing function of $\widetilde{xz}$, and the three-level tie-breaker solves (4) for all $\widetilde{xz}$, they conclude that there is a monotone trade-off between short-term gain and statistical efficiency. In other words, greater statistical efficiency from an optimal design requires giving up short-term gain.

We now extend these results to general $\widetilde{z}$ and other running variable distributions. Hereafter, $p_{\text{opt}} = p_{\text{opt};\widetilde{z},\widetilde{xz}}$ denotes an optimal design without the monotonicity constraint, to be contrasted with $p^{\dagger}_{\text{opt}}$ of Section 3.2. Note we have made the dependence of these designs on $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ explicit. We use the same efficiency criterion $\operatorname{Eff}(\cdot)$ as Owen and Varian (2020). Recall this is a c-optimality criterion corresponding to the scaled asymptotic variance of the OLS estimate for β_3 in (1), and equivalent to D-optimality for our problem (4) by Section 2.3.

THEOREM 3. *Suppose the distribution function F of the running variable has a positive derivative everywhere in I, the smallest open interval with $\int_I f(x)\,\mathrm{d}x = 1$. If additionally $F(x) = 1 - F(-x)$, $\forall x \in I$, then fixing any $\widetilde{z} \in (-1, 1)$, $\mathrm{Eff}(p_{\mathrm{opt};\widetilde{z},\widetilde{xz}})$ is decreasing in $\widetilde{xz}$.*

PROOF. See Appendix D. □

It turns out, however, that the gain versus efficiency trade-off is no longer monotone under the monotonicity constraint. Indeed, our next theorem shows that whenever $\widetilde{z} \neq 0$, if F is symmetric (or indeed, not extremely skewed), the fully randomized design $p_{\theta,\theta,0}$ is inadmissible for any $\theta \neq 1/2$, in the sense that there exists a different *monotone* design p with $\mathbb{E}_p(z) = \widetilde{z}$ but both $\mathrm{Eff}(p) > \mathrm{Eff}(p_{\theta,\theta,0})$ and $\mathbb{E}_p(xz) > \mathbb{E}_{p_{\theta,\theta,0}}(xz)$. In other words, the RCT is no longer admissible under $\mathrm{Eff}(\cdot)$ when $\widetilde{z} \neq 0$.

THEOREM 4. *Fix $\widetilde{z} \in (-1, 1) \setminus \{0\}$, and assume F satisfies the conditions of Theorem 3. If $\widetilde{z} < 0$, assume that $\mathbb{E}(x^2) < F^{-1}(1)^2$; otherwise, assume that $\mathbb{E}(x^2) < F^{-1}(0)^2$. Here, $F^{-1}(1) \equiv \sup I$ and $F^{-1}(0) \equiv \inf I$. Let $p_1 = p_{\theta,\theta,0}$ be the fully randomized monotone design with $\mathbb{E}_{p_1}(z) = \widetilde{z}$, so that $\theta = (1 + \widetilde{z})/2$. Then there exists a monotone design p_2 such that $\mathbb{E}_{p_2}(z) = \widetilde{z}$ yet both $\mathrm{Eff}(p_2) > \mathrm{Eff}(p_1)$ and $\mathbb{E}_{p_2}(xz) > 0 = \mathbb{E}_{p_1}(xz)$.*

PROOF. See Appendix E. □

5. Examples. In this section, we compute the optimal exploration-exploitation trade-off curves investigated in Section 4 for several specific running variable distributions F. We can obtain large gains in efficiency under the criterion $\mathrm{Eff}(\cdot)$ defined in Section 2.3 by moving away from the three-level tie-breaker design to $p^{\dagger}_{\mathrm{opt}}$, without sacrificing short-term gain. We see further (generally smaller) improvements when we remove the monotonicity constraint and move from $p^{\dagger}_{\mathrm{opt}}$ to p_{opt}.

To generate these curves, we compute optimal designs $p_{\mathrm{opt};\widetilde{z},\widetilde{xz}}$ and $p^{\dagger}_{\mathrm{opt};\widetilde{z},\widetilde{xz}}$ and evaluate their efficiency for various fixed $\widetilde{z} \in (-1, 1)$ as we vary the short-term gain constraint $\widetilde{xz}$ over a fine grid covering $[0, \widetilde{xz}_{\max}(\widetilde{z})]$. For interpretability, we write $\widetilde{xz} = \delta \cdot \widetilde{xz}_{\max}(\widetilde{z})$ and specify short-term gain with the normalized parameter $\delta \in [0, 1]$, as discussed in Section 2.2. When F is continuous, solutions p_{opt} and $p^{\dagger}_{\mathrm{opt}}$ to (4) are computed by noting that we can write $p_{\max} = p_{[a_1,a_2]^c}$, $p_{\min} = p_{[b_1,b_2]}$, $p^{\dagger}_{\max} = p_{\ell,1,t}$ and $p^{\dagger}_{\min} = p_{0,u,s}$ by Propositions 1 and 2. Each of these designs has two unknown parameters that must be the unique solutions to the two feasibility constraints $\mathbb{E}_p(z) = \widetilde{z}$ and $\mathbb{E}_p(xz) = \widetilde{xz}$. Given these parameters, we can apply (20) to compute p_{opt} and $p^{\dagger}_{\mathrm{opt}}$. We could also get an optimal design of the form (25) when $p_{\mathrm{opt}} \notin \{p_{\max}, p_{\min}\}$. First, we compute $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \mathrm{Eff})$ via (17), noting that the endpoints of $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$ are $\mathbb{E}_{p_{\min}(\widetilde{z},\widetilde{xz})}(x^2z)$ and $\mathbb{E}_{p_{\max}(\widetilde{z},\widetilde{xz})}(x^2z)$. Then (17) is simply maximizing a continuous function over a closed interval, so it can be handled by standard methods such as Brent's algorithm (Brent (1973)). Given $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \mathrm{Eff})$, we can then numerically search for a_1, a_2, a_3 such that $p_{\mathrm{opt}} = \mathbf{1}(x \le a_1) + \mathbf{1}(a_2 \le x \le a_3)$ is feasible with $\mathbb{E}_{p_{\mathrm{opt}}}(x^2z) = \widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \mathrm{Eff})$. By Theorem 1, such a solution will exist and be optimal. We can do a similar search for a two-level optimal monotone design $p^{\dagger}_{\mathrm{opt}}$ based on Theorem 2.

5.1. *Uniform running variable.* We begin with the case $F \sim \mathbb{U}(-1, 1)$. This is the distribution most extensively studied by Owen and Varian (2020), and allows closed form expressions for the parameters in $p_{\max}$, $p_{\min}$, $p^{\dagger}_{\max}$ and $p^{\dagger}_{\min}$, given in Table 1. Figure 1 shows

TABLE 1

A list of the parameters for the various designs considered in Section 5, for fixed $(\tilde{z}, \widetilde{xz}) \in \mathcal{J}$, when $F \sim \mathbb{U}(-1, 1)$. The values for Δ are only valid if they are between 0 and $\min((1-\tilde{z})/2, (1+\tilde{z})/2)$, inclusive. Otherwise, there is no feasible three-level tie-breaker design

Design	Parameter	Value
$p_{3;\tilde{z},\Delta}$	Δ	$2(1-\tilde{z}^2-2\widetilde{xz})^{1/2}$
$p_{\max;\tilde{z},\widetilde{xz}}$	a_1	$-\widetilde{xz}/(1-\tilde{z})-(1-\tilde{z})/2$
	a_2	$-\widetilde{xz}/(1-\tilde{z})+(1-\tilde{z})/2$
$p_{\min;\tilde{z},\widetilde{xz}}$	b_1	$\widetilde{xz}/(1+\tilde{z})-(1+\tilde{z})/2$
	b_2	$\widetilde{xz}/(1+\tilde{z})+(1+\tilde{z})/2$
$p^{\dagger}_{\max;\tilde{z},\widetilde{xz}}$	ℓ	$(1/2)(1-\tilde{z}^2-2\widetilde{xz})/(1-\tilde{z}-\widetilde{xz})$
	t	$1-2\widetilde{xz}/(1-\tilde{z})$
$p^{\dagger}_{\min;\tilde{z},\widetilde{xz}}$	u	$(1/2)(1+z)^2/(1+\tilde{z}-\widetilde{xz})$
	s	$2\widetilde{xz}/(1+\tilde{z})-1$

plots of $\mathrm{Eff}(p)^{-1}$ versus $\widetilde{xz}$ for $\tilde{z} \in \{0, -0.2, -0.5, -0.7\}$ under different designs: the three-level tie-breaker $p_{3;\tilde{z},\Delta}$, a globally optimal design $p_{\mathrm{opt};\tilde{z},\widetilde{xz}}$ and an optimal monotone design $p^{\dagger}_{\mathrm{opt};\tilde{z},\widetilde{xz}}$. Since F is symmetric, the curves would be identical if $\tilde{z}$ were replaced with $-\tilde{z}$.

As shown in Owen and Varian (2020), under the constraint $\tilde{z} = 0$ the three-level tie-breaker is optimal for all δ, and thus the three-level tie-breaker, p_{opt} and $p^{\dagger}_{\mathrm{opt}}$ all attain the optimal efficiency, as can be seen in the top left panel of Figure 1. The proof of Theorem 3 shows this would hold for any continuous, symmetric running variable distribution F. As $\tilde{z}$ moves away from 0, however, we see that the three-level tie-breaker becomes increasingly less efficient relative to both the optimal monotone design and the optimal design. At the same time, the range of short-term gain values $\widetilde{xz}$ attainable by three-level tie-breaker designs becomes smaller relative to the full range achievable by arbitrary designs. Accordingly, the inverse efficiency curves for short-term gain do not extend over the full range of δ values. Note that Figure 1 plots the *reciprocal* of the efficiency criterion $\mathrm{Eff}(\cdot)$, so that it can be interpreted as an asymptotic (conditional) variance for $\hat{\beta}_3$ via (15), and compared with Owen and Varian (2020).

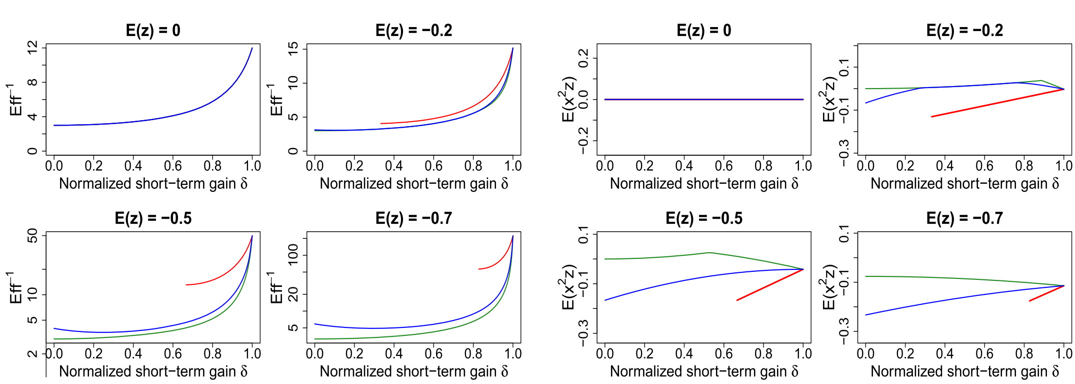

FIG. 1. *Left four plots: The inverse efficiencies of the three-level tie-breaker $p_{3;\tilde{z},\Delta}$, the optimal designs p_{opt} in Section 3.1 and the optimal monotone designs $p^{\dagger}_{\mathrm{opt}}$ in Section 3.2 as a function of the normalized short-term gain parameter δ for fixed $\tilde{z} \in \{0, -0.2, -0.5, -0.7\}$, when $F \sim \mathbb{U}(-1, 1)$. When $\tilde{z} = 0$, all three curves coincide. Note the logarithmic vertical spacing for the bottom two plots. Right four plots: The value of $\mathbb{E}_p(x^2 z)$ for the designs p in the left four plots.*

TABLE 2

An extension of Table 2 in Owen and Varian (2020), showing how p_{opt} and $p^{\dagger}_{\text{opt}}$ can greatly increase efficiency without sacrificing short-term gain, compared to a three-level tie-breaker. All designs p in this table satisfy $\mathbb{E}_p(z) = -0.7$ and assume $F \sim \mathbb{U}(-1, 1)$

Design p	Description	Normalized short-term gain δ	$\text{Eff}(p)^{-1}$
$p_{0,1,0.7}$	Sharp RDD	1.000	223.44
$p_{3;-0.7,\Delta}$	3-level tie-breaker	0.980	137.56
$p^{\dagger}_{\text{opt};-0.7,0.25}$	Optimal monotone design	0.980	54.90
$p_{\text{opt};-0.7,0.25}$	Optimal design	0.980	42.37

Table 2 extends Table 2 of Owen and Varian (2020), referring to a setting in which only 15% of subjects are to be treated ($\tilde{z} = -0.7$). That table shows the inverse efficiency of the sharp RDD $p_{0,1,0.7}$ is 223.44, while the three-level tie-breaker $p_{3;-0.7,0.05}$ reduces this by about 40% to 137.56, at the cost of around 2% of the short-term gain of the sharp RDD over the RCT. Then $\text{Eff}(p^{\dagger}_{\text{opt}})^{-1} = 54.90$ and $\text{Eff}(p_{\text{opt}})^{-1} = 42.37$, further improving efficiency for designs p_{opt} and $p^{\dagger}_{\text{opt}}$ achieving the same short-term gain as the three-level tie-breaker. For this example, we can directly compute with (31) that $p^{\dagger}_{\text{opt};-0.7,0.25} = p^{\dagger}_{\text{max};-0.7,-0.25} = p_{\ell,1,t}$ is the unique optimal monotone design, where by Table 1, $\ell = 0.0034$ and $t = 0.7059$. In other words, the unique optimal monotone tie-breaker design deterministically assigns treatment to the top 14.7%, and gives the other subjects an equal, small (0.34%) chance of treatment.

A limitation of this analysis is that in many practical settings, the two line regression model will not fit very well over the entire range of x values. In that case, the investigator might use a narrower data range, essentially fitting a less asymmetric two line model, as illustrated in Owen and Varian (2020). This is equivalent to using a local linear regression with a rectangular "boxcar" kernel. In this setting, we know from Figure 1 that when the treatment proportion is not exactly 50%, we can always do better than the three-level tie-breaker using monotone two-level design. Even with a small asymmetry, for example, 40% treatment ($\tilde{z} = -0.2$), we see a noticeable efficiency increase between the three-level tie-breaker and an optimal monotone design across all values of δ. Recalling that efficiency of a design p is a univariate function of $\mathbb{E}_p(x^2 z)$ under the treatment fraction and gain constraints, we see from the top right plot in Figure 1 that we lose no efficiency by imposing the monotonicity constraint when $0.3 < \delta < 0.8$.

Finally, consistent with the results of Section 4, we observe in Figure 1 that $\text{Eff}(p_{\text{opt}})$ decreases with the gain parameter δ for each $\tilde{z}$, while near $\delta = 0$, $\text{Eff}(p^{\dagger}_{\text{opt}})$ *increases* with δ for all $\tilde{z} \neq 0$. This clearly demonstrates the inadmissibility of the fully randomized design from Theorem 4. For example, if we fix $\tilde{z} = -0.5$ (so 25% of the subjects are to be treated), the fully randomized design $p_{0.25,0.25,0} = p^{\dagger}_{\text{opt};-0.5,0}$ has efficiency 0.25 and no short-term gain ($\delta = 0$), while $p^{\dagger}_{\text{opt};-0.5,0.1}$ has higher efficiency (0.28) with short-term gain $\delta = 0.27 > 0$. However, if we remove the monotonicity constraint, by Theorem 3 $p_{\text{opt};-0.5,0}$ is the most efficient design over all attainable gain values, attaining efficiency 0.33 with $\delta = 0$.

5.2. *Skewed running variable.* We now repeat the analysis of Section 5.1 for a skewed running variable distribution F:

$$F(x) = 1 - \exp(-\sqrt{x+2}) \tag{28}$$

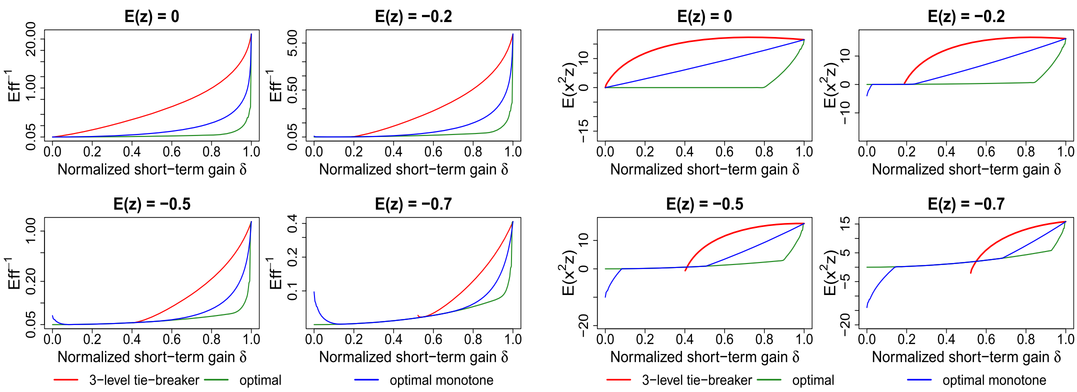

FIG. 2. *Same as Figure* 1, *except for the case where F is a centered Weibull distribution* (28).

for $x \in (-2, \infty) = I$. This corresponds to a mean-centered Weibull distribution with shape parameter 0.5 and scale parameter 1. Figure 2 shows the trade-off curves under this distribution F. We see, as expected by Theorem 4, that once again the fully randomized design is inadmissible, even within the class of monotone designs, when $\widetilde{z} \neq 0$.

Another notable feature when F is not symmetric is that the three-level tie-breaker is no longer optimal, even in the balanced case $\widetilde{z} = 0$. While the unconstrained optimal design attains the lower bound $\mathbb{E}(x^2 z) = \widetilde{x^2 z}^*(0, \widetilde{xz}; \mathrm{Eff}) = 0$ for a wide range of short-term gains, the right-hand side of Figure 2 shows the three-level tie-breaker does not, except in the case $\widetilde{z} = \widetilde{xz} = 0$ corresponding to the RCT. In Figure 2, we see the optimal design is over 100 times as efficient as the three-level tie-breaker for sufficiently large δ, even in the balanced setting $\widetilde{z} = 0$. In the unbalanced treatment cases, we also see a range of values for which optimal designs with and without the monotonicity constraint attain the same value of $\mathbb{E}(x^2 z)$. In those situations, there exists a globally optimal design that is also monotone.

5.3. *Fixed-x data example.* We now illustrate how to compute optimal designs for the original fixed-x problem (3) using a real data example. Ludwig and Miller (2007) used an RDD to analyze the impact of Head Start, a U.S. government program launched in 1965 that provides benefits such as preschool and health services to children in low-income families. When the program was launched, extra grant-writing assistance was provided to the 300 counties with the highest poverty rates in the country. This created a natural discontinuity in the amount of funding to counties as a function of x, a county poverty index based on the 1960 U.S. Census. The distribution of x over $n = 2{,}804$ counties is shown in Figure 3. The data is made freely available by Cattaneo, Titiunik and Vazquez-Bare (2017).

If the government had deemed it ethical to somewhat randomize the 300 counties receiving the grant-writing assistance, it could have more efficiently estimated the causal impact of this assistance using our $p^{\dagger}_{\mathrm{opt}}$, while still ensuring poorer counties are preferentially helped, and no county has a lower chance of getting the assistance than a more well-off county. As in the data example of Kluger and Owen (2023), we do not observe the potential outcomes, so we cannot actually implement such a design and compute any estimators. However, we can still study statistical efficiencies, which depend only on the expected information matrix $\mathcal{I}$.

We fix the treatment fraction at 300/2804, corresponding to $\widetilde{z} \approx -0.79$. Varying the short-term gain constraint $\widetilde{xz}$, we seek to compute $p^{\dagger}_{\max;\widetilde{z},\widetilde{xz}}$ and $p^{\dagger}_{\min;\widetilde{z},\widetilde{xz}}$. We describe how to compute the former. Because F is discrete, it suffices to only consider discontinuity points $t \in \{x_1, \dots, x_n\}$ where F places positive probability mass, as every design of the form of $p^{\dagger}_{\max}$ in (26) has a representation in that form with such t. Also, given the values of the discontinuity

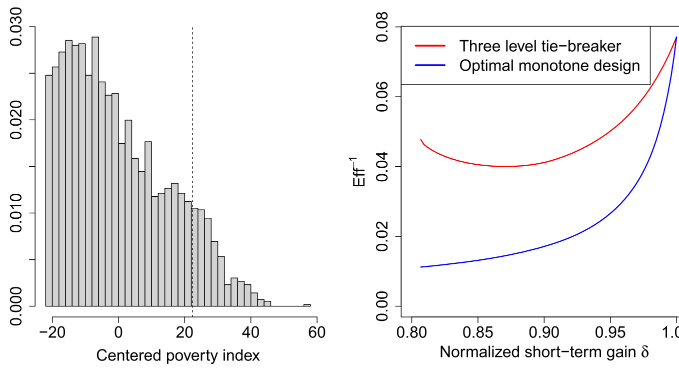

FIG. 3. *Left*: *A histogram of the mean-centered poverty index x for $n = 2{,}804$ counties used to determine eligibility for additional grant-writing assistance in the Head Start program. The dotted vertical line indicates the eligibility threshold. Right*: *The exploration-exploitation trade-offs for the Head Start data, comparing the three-level tie-breaker with the optimal monotone two-level design. The curves intersect at the value* Eff^{-1} *of the RDD.*

t and $p(t) = \epsilon$, there is at most one value $\ell = \ell(t, \epsilon) \in [0, 1]$ such that the resulting design p in the form of $p^{\dagger}_{\max}$ in (26) satisfies the treatment fraction constraint $\mathbb{E}_p(z) = \widetilde{z}$. When such an ℓ exists for some (t, ϵ), call the corresponding design $p^{(t,\epsilon)}$ (note we suppress the dependence on $\widetilde{z}$). From Appendix C, we deduce that $\mathbb{E}_{p^{(t,\epsilon)}}(xz) < \mathbb{E}_{p^{(t',\epsilon')}}(xz)$ and $\ell(t, \epsilon) > \ell(t', \epsilon')$ if $t > t'$, or if $t = t'$ and $p(t) < p'(t)$. This shows we can efficiently find the unique (t, ϵ) so that $p^{(t,\epsilon)}$ satisfies the desired short-term gain constraint $\mathbb{E}_{p^{(t,\epsilon)}}(xz) = \widetilde{xz}$. In particular, we compute $t = \max\{s \in \{x_1, \dots, x_n\} | \mathbb{E}_{p^{(s,1)}}(xz) \geq \widetilde{xz}\}$ via a binary search on $\{x_1, \dots, x_n\}$, then solve for ϵ to satisfy $\mathbb{E}_{p^{(t,\epsilon)}}(xz) = \widetilde{xz}$. Given sorted x, this entire procedure computes $p^{\dagger}_{\max}$ in $O(n)$ operations, as for each (t, ϵ), $\ell(t, \epsilon)$ and $\mathbb{E}_{p^{(t,\epsilon)}}(xz)$ can be computed in constant time using (5) given the partial sums $\{\sum_{i=1}^{m} x_i\}_{m=1}^{n}$.

After computing $p^{\dagger}_{\min}$ with a similar approach, we can apply (20) to compute an optimal design $p^{\dagger}_{\mathrm{opt}}$. As in the continuous case, we can alternately obtain a solution $p^{\dagger}_{\mathrm{opt}}$ of the form in Theorem 2 by finding ℓ', u', t' and ϵ' such that $\mathbb{E}_p(z) = \widetilde{z}$, $\mathbb{E}_p(xz) = \widetilde{xz}$, and $\mathbb{E}_p(x^2 z) = \widetilde{x^2 z}^{*}(\widetilde{z}, \widetilde{xz}; \mathrm{Eff})$ for $p(x) \equiv \ell' \mathbf{1}(x < t') + \epsilon' \mathbf{1}(x = t') + \mathbf{1}(x > t')$. Unlike the continuous case, we now have 4 unknown parameters instead of 3. We can search for an optimal set of these parameters by looping through the finite possible values of t' and then doing a univariate search for ϵ', noting that knowledge of t' and ϵ' determines ℓ' and u' by the equality constraint parameters $\widetilde{z}, \widetilde{xz}$). We implemented this search, along with the procedure to compute $p^{\dagger}_{\max}$ and $p^{\dagger}_{\min}$ described above, in the R language (R Core Team (2022)). See https://github.com/hli90722/optimal_tiebreaker_designs.

The right panel of Figure 3 shows the inverse efficiency for the three-level tie-breaker (9) versus the best two-level monotone design obtained by applying the above procedure to the x_i in the Head Start data. It turns out that for these x_i and our choice of $\widetilde{z}$, $p^{\dagger}_{\max}$ is optimal for all $\widetilde{xz}$ (and hence the unique optimal design, by Proposition 2). We note that with a normalized short-term gain $\delta \approx 0.958$, which corresponds to random assignment for about 150 counties in the three-level tie-breaker, the optimal monotone two-level design has inverse efficiency 0.030, compared to 0.050 for the three-level tie-breaker. That is, confidence intervals for β_3 using the three-level tie-breaker would be about 29% wider than for the optimal monotone two-level design, without additional short-term gain. The sharp RDD would give 62% wider intervals than the optimal monotone two-level design with only about 4.2% additional short-term gain.

6. Summary. Our results provide a thorough characterization of the solutions to a constrained optimal experiment design problem. Considering a linear regression model for a scalar outcome involving a binary treatment assignment indicator z, a scalar running variable x and their interaction, we seek to specify a randomized treatment assignment scheme based on x—a tie-breaker design—that optimizes a statistical efficiency criterion that is an arbitrary continuous function of the expected information matrix under this regression model. We have equality constraints on the proportion of subjects receiving treatment due to an external budget, and on the covariance between x and z due to a preference for treating subjects with higher values of x. Critically, our proof techniques, which deviate from those typically used to show equivalence theorems, enable an additional monotonicity constraint. This allows our results to handle the ethical or economic requirement that a subject cannot have a lower chance of receiving the treatment than another subject with a lower value of x.

In a setting where the running variable x is viewed as random from some distribution F—and thus part of the randomness in the expected information matrix defining the efficiency criterion—we prove the existence of constrained optimal designs that stratify x into a small number of intervals and assign treatment with the same probability to all individuals within each stratum. In particular, with the monotonicity constraint that is essential in our motivating applications, we only need three strata, one of which only contains a single running variable value. We also provide strong conditions on which the optimal tie-breaker design is unique. We emphasize the generality of our results, which apply for any continuous efficiency criterion, any running variable distribution F (subject only to weak moment existence conditions), and the full range of feasible equality constraints. The problem an investigator faces in practice, where there are a finite number of running variable values $x_1, \dots, x_n$ known (hence nonrandom) at the time of treatment assignment, is a special case of our more general problem where F is discrete and takes on values $x_1, \dots, x_n$ with equal probability. This enables optimal designs to be easily computed in practice, as described in Section 5.3.

We believe that this work provides a useful starting point to study optimal tie-breaker designs. For results on tie-breaker designs beyond the two line parametric regression, see Morrison and Owen (2022) for a multivariate regression context and Kluger and Owen (2023) for local linear regression models with a scalar running variable.

APPENDIX A: PROOF OF COROLLARY 1

For any $\widetilde{z} \in (-1, 1)$, we have $\widetilde{xz}_{\max}(\widetilde{z}) = \mathbb{E}_{p_{\widetilde{z}}}(xz) = 2\mathbb{E}(xp_{\widetilde{z}}(x))$ where $p_{\widetilde{z}}$ is as in Lemma 1. The desired condition $M_{11} > 0$ is equivalent to $(\widetilde{xz})^2 < \mathbb{E}(x^2)(1-\widetilde{z}^2)$ and so it suffices to show $\mathbb{E}(xp_{\widetilde{z}}(x))^2 < \mathbb{E}(x^2)((1+\widetilde{z})/2)((1-\widetilde{z})/2)$.

Applying Cauchy–Schwarz to $xp_{\widetilde{z}}(x)^{1/2} \times p_{\widetilde{z}}(x)^{1/2}$ and then $x(1-p_{\widetilde{z}}(x))^{1/2} \times (1-p_{\widetilde{z}}(x))^{1/2}$ yields the two equations

$$\mathbb{E}\big(xp_{\widetilde{z}}(x)\big)^2 < \mathbb{E}\big(x^2 p_{\widetilde{z}}(x)\big)\left(\frac{1+\widetilde{z}}{2}\right),$$

$$\mathbb{E}\big(xp_{\widetilde{z}}(x)\big)^2 = \mathbb{E}\big(x\big(1-p_{\widetilde{z}}(x)\big)\big)^2 < \mathbb{E}\big(x^2\big(1-p_{\widetilde{z}}(x)\big)\big)\left(\frac{1-\widetilde{z}}{2}\right),$$

where we have used the fact $\mathbb{E}(xp_{\widetilde{z}}(x)) = -\mathbb{E}(x(1-p_{\widetilde{z}}(x)))$ since $\mathbb{E}(x) = 0$. Note both inequalities are strict, since x cannot equal a scalar multiple of $p_{\widetilde{z}}(x)$ w.p.1. If it did, then $\mathbb{E}(kp_{\widetilde{z}}(x) - x) = k(1+\widetilde{z})/2 = 0$ for some k, implying $k = 0$, and hence $x = 0$ w.p.1, contradicting $\text{Var}(x) > 0$. As $\mathbb{E}(x^2) = \mathbb{E}(x^2 p_{\widetilde{z}}(x)) + \mathbb{E}(x^2(1-p_{\widetilde{z}}(x)))$, we know that either $\mathbb{E}(x^2 p_{\widetilde{z}}(x)) \le \mathbb{E}(x^2)\cdot(1-\widetilde{z})/2$ or $\mathbb{E}(x^2(1-p_{\widetilde{z}}(x))) \le \mathbb{E}(x^2)\cdot(1+\widetilde{z})/2$. □

APPENDIX B: PROOF OF UNIQUENESS IN PROPOSITION 1

We show uniqueness for $p_{\min}$. The same argument shows uniqueness for $1 - p_{\max}$, and hence uniqueness for $p_{\max}$. Suppose that $p(x) = \delta_1 \mathbf{1}(x = b_1) + \mathbf{1}(b_1 < x < b_2) + \delta_2 \mathbf{1}(x = b_2)$ and $p'(x) = \delta_1' \mathbf{1}(x = b_1') + \mathbf{1}(b_1' < x < b_2') + \delta_2' \mathbf{1}(x = b_2')$ are both solutions for $p_{\min}$. By symmetry, we can assume that either $b_1 < b_1'$, or both $b_1 = b_1'$ and $\delta_1 \le \delta_1'$. Since p and p' are feasible for (19), we must have $\mathbb{E}(p(x) - p'(x)) = 0$ and $\mathbb{E}(x(p(x) - p'(x))) = 0$, in view of (5). We show that $p(x) = p'(x)$ w.p.1. under $x \sim F$. Note that we can assume without loss of generality that $\Pr(x \in [b_1, b_1 + \epsilon)) > 0$ for any $\epsilon > 0$, because otherwise, we could increase b_1 to $b_1 + \sup\{\epsilon > 0 | \Pr(x \in [b_1, b_1 + \epsilon)) = 0\}$ without changing p on a set of positive probability. We can similarly assume that $\Pr(x \in (b_2 - \epsilon, b_2]) > 0$ for any $\epsilon > 0$. Finally, we impose these two canonicalizing conditions on b_1' and b_2' as well.

Assume first that $b_2 > b_1$. Then we cannot have $b_1' > b_1$ because we would then need either $b_2' > b_2$ or $b_2' = b_2$ with $\delta_2' > \delta_2$ and $\Pr(x = b_2) > 0$ to enforce $\mathbb{E}(p(x) - p'(x)) = 0$ and this would cause $\mathbb{E}(x(p(x) - p'(x))) < 0$. We similarly cannot have $b_1' = b_1$ with both $\delta_1' > \delta_1$ and $\Pr(x = b_1) > 0$. Therefore, after canonicalizing, we know that both p and p' are equivalent to designs of the form given with $b_1 = b_1'$ and $\delta_1 = \delta_1'$ along with the analogous conditions $b_2 = b_2'$ and $\delta_2 = \delta_2'$. Then our canonicalized p and p' satisfy $p(x) = p'(x)$ for all x and so in particular $\Pr(p(x) = p'(x)) = 1$.

It remains to handle the case where $b_1 = b_2$. We then have $\Pr(x = b_1) > 0$ since $\widetilde{z} > -1$. If $b_1' > b_1$, then the support of p' is completely to the right of that of p, which violates $\mathbb{E}(x(p(x) - p'(x))) = 0$. We can similarly rule out $b_2' < b_2$. As a result, p' must have $b_1' \le b_1 = b_2 \le b_2'$. Then we must have $\delta_1' \Pr(x = b_1') + \Pr(b_1' < x < b_1) = 0$ or else $\mathbb{E}(p'(x) - p(x)) > 0$. For the same reason, we must have $\Pr(b_2 < x < b_2') + \delta_2' \Pr(x = b_2') = 0$. It then follows that both p and p' have support $\{b_1\}$ and then $\mathbb{E}(p(x)) = \mathbb{E}(p'(x)) = (1 + \widetilde{z})/2$ forces $(1 + \widetilde{z})/(2 \Pr(x = b_1)) = \delta_1 = p(b_1) = p'(b_1) = \delta_1'$, so $\Pr(p(x) = p'(x)) = 1$.

APPENDIX C: PROOF OF UNIQUENESS IN PROPOSITION 2

We focus on $p^{\dagger}_{\max}$ and consider two monotone designs p and p' satisfying the feasibility constraints $\mathbb{E}_p(z) = \mathbb{E}_{p'}(z) = \widetilde{z}$ and $\mathbb{E}_p(xz) = \mathbb{E}_{p'}(xz) = \widetilde{xz}$ along with the characterization of $p^{\dagger}_{\max}$ in (26). Then $p(x) = \ell \mathbf{1}(x < t) + \delta \mathbf{1}(x = t) + \mathbf{1}(x > t)$ and $p'(x) = \ell' \mathbf{1}(x < t') + \delta' \mathbf{1}(x = t') + \mathbf{1}(x > t')$ for some $\ell, \ell' \in (0, 1)$ with $\ell \le \delta \le 1$ and $\ell' \le \delta' \le 1$. Note the cases $\ell \in \{0, 1\}$ (and the same for ℓ') are excluded by the assumptions that $\widetilde{xz} < \widetilde{xz}_{\max}(\widetilde{z})$ and $\widetilde{z} < 1$. Also, $\widetilde{z} < 1$ also guarantees $\min(\Pr(x \le t), \Pr(x \le t')) > 0$. Finally, we note that we only have to show $p(x) = p'(x)$ for almost all $x \ne t$, since then $\mathbb{E}(p(x) - p'(x)) = 0$ ensures either $p(t) = p'(t)$ or $\Pr(x = t) = 0$; in either case, this gives $p = p'$ w.p.1. By symmetry, we can assume that $t \le t'$ with $\delta \equiv p(t) \ge p(t') =: \delta'$ if $t = t'$. Then $p(x) = p'(x)$ for all $x > t'$.

Now we compute

$$
\begin{aligned}
&\mathbb{E}\bigl(x\bigl(p(x) - p'(x)\bigr)\bigr) \\
&\quad = \mathbb{E}\bigl((x - t)\bigl(p(x) - p'(x)\bigr)\bigr) \text{ since } \mathbb{E}\bigl(p(x)\bigr) = \mathbb{E}\bigl(p'(x)\bigr) \\
&\quad = \mathbb{E}\bigl((t - x)\bigl(p'(x) - p(x)\bigr)\mathbf{1}(x < t)\bigr) + \mathbb{E}\bigl((x - t)\bigl(p(x) - p'(x)\bigr)\mathbf{1}\bigl(t < x \le t'\bigr)\bigr) \\
&\quad = (\ell' - \ell)\mathbb{E}\bigl((t - x)\mathbf{1}(x < t)\bigr) + \mathbb{E}\bigl((x - t)\bigl(1 - p'(x)\bigr)\mathbf{1}\bigl(t < x \le t'\bigr)\bigr).
\end{aligned}
$$

If $t = t'$, then the right-hand side reduces to just $(\ell' - \ell)\mathbb{E}((t - x)\mathbf{1}(x < t))$. This is nonzero unless $\Pr(x < t) = 0$ or $\ell = \ell'$. In both cases, $p(x) = p'(x)$ for almost all $x \ne t$.

If $t < t'$, then we can assume $\Pr(t < x \le t') > 0$ (otherwise the problem reduces to the case $t = t'$). First, suppose $\ell \ge \ell'$. Then $p(x) \ge p'(x)$ for all x and so the treatment fraction constraint would require the identity

$$
\mathbf{1}\bigl(t < x \le t'\bigr) = p(x)\mathbf{1}\bigl(t < x \le t'\bigr) \ge p'(x)\mathbf{1}\bigl(t < x \le t'\bigr) = \delta'\mathbf{1}\bigl(x = t'\bigr) + \ell'\mathbf{1}\bigl(t < x < t'\bigr)
$$

to hold with equality w.p.1. But since $\ell' < 1$, equality w.p.1. can only occur if $\Pr(t < x < t') = 0$ and $\delta' = 1$. In that case, we immediately see $p(x) = p'(x)$ for almost all $x > t$, but $0 = \mathbb{E}(x(p(x) - p'(x))) = (\ell' - \ell)\mathbb{E}((t - x)\mathbf{1}(x < t))$ so $p(x) = p'(x)$ for almost all $x < t$ as well. Conversely, if we suppose $\ell < \ell'$, then $\mathbb{E}(x(p(x) - p'(x))) = 0$ requires $\Pr(x < t) = 0$ and $p'(x) = 1$ for almost all $x \in (t, t']$, so once again $p(x) = p'(x)$ for almost all $x \neq t$.

APPENDIX D: PROOF OF THEOREM 3

For any feasible design p, we have

$$\begin{aligned}\det(M) = &-\frac{(1-\widetilde{z}^2)\cdot(\mathbb{E}_p(x^2z))^2}{\mathbb{E}(x^2)} - \frac{2\widetilde{z}(\widetilde{xz})^2\cdot\mathbb{E}_p(x^2z)}{\mathbb{E}(x^2)} + \mathbb{E}(x^2)(1-\widetilde{z}^2)\\ &+ (\widetilde{xz})^2\left(\frac{(\widetilde{xz})^2}{\mathbb{E}(x^2)} - 2\right),\end{aligned} \tag{29}$$

where $M = M(p)$ as in (16). Thus, $\det(M(p))$ is a concave quadratic function of $\mathbb{E}_p(x^2z)$ globally maximized at

$$a^*(\widetilde{z}, \widetilde{xz}) \equiv -\frac{\widetilde{z}\cdot(\widetilde{xz})^2}{1-\widetilde{z}^2}. \tag{30}$$

It follows that $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \mathrm{Eff})$ is the point in $I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz}) = [I_{\mathcal{F};\min}(\widetilde{z}, \widetilde{xz}), I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz})]$ closest in absolute value to $a^*(\widetilde{z}, \widetilde{xz})$, that is,

$$\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \mathrm{Eff}) = \begin{cases} I_{\mathcal{F};\min}(\widetilde{z}, \widetilde{xz}), & a^*(\widetilde{z}, \widetilde{xz}) \le I_{\mathcal{F};\min}(\widetilde{z}, \widetilde{xz}),\\ a^*(\widetilde{z}, \widetilde{xz}), & I_{\mathcal{F};\min}(\widetilde{z}, \widetilde{xz}) < a^*(\widetilde{z}, \widetilde{xz}) < I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz}),\\ I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz}), & a^*(\widetilde{z}, \widetilde{xz}) \ge I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz}).\end{cases} \tag{31}$$

The above holds for any choice of $\mathcal{F}$; for the remainder of this proof we take $\mathcal{F}$ to be the set of all measurable design functions.

We first show the case where $\widetilde{z} = 0$. Note that $\mathbb{E}_p(x^2z) = 0 = a^*(0, \widetilde{xz})$ for any symmetric design p, by symmetry of the running variable distribution. By continuity, for any $\widetilde{xz} \in [0, \widetilde{xz}_{\max}(\widetilde{z})]$ there exists $\Delta \in [0, \infty]$ such that the three-level tie-breaker $p_{3;\widetilde{z},\Delta}$ (which is symmetric and always satisfies $\mathbb{E}_{p_{3;\widetilde{z},\Delta}}(z) = 0$) satisfies $\mathbb{E}_{p_{3;\widetilde{z},\Delta}}(xz) = \widetilde{xz}$, too. This shows that for all $\widetilde{xz} \in [0, \widetilde{xz}_{\max}(\widetilde{z})]$, $0 \in I_{\mathcal{F}}(\widetilde{z}, \widetilde{xz})$, and hence $\widetilde{x^2z}^*(0, \widetilde{xz}; \mathrm{Eff}) = 0$, meaning any feasible design p with $\mathbb{E}_p(x^2z) = 0$ is optimal. Then by (29),

$$\det(M(p_{\mathrm{opt};0,\widetilde{xz}})) = \mathbb{E}(x^2) + (\widetilde{xz})^2\left(\frac{(\widetilde{xz})^2}{\mathbb{E}(x^2)} - 2\right),$$

which is decreasing in $\widetilde{xz}$ on $[0, \widetilde{xz}_{\max}(\widetilde{z})]$, showing the theorem for $\widetilde{z} = 0$.

For the cases $\widetilde{z} < 0$ and $\widetilde{z} > 0$, we begin with the two following claims.

CLAIM 1. *For any* $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ *with* $\widetilde{z} < 0$, *we have* $I_{\mathcal{F};\min}(\widetilde{z}, \widetilde{xz}) \le 0 < a^*(\widetilde{z}, \widetilde{xz})$.

CLAIM 2. *For any* $(\widetilde{z}, \widetilde{xz}) \in \mathcal{J}$ *with* $\widetilde{z} > 0$, *we have* $I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz}) \ge 0 > a^*(\widetilde{z}, \widetilde{xz})$.

PROOF OF CLAIMS 1 AND 2. We write $I_{\mathcal{F};\min}(\widetilde{z}, \widetilde{xz}) = 2\mathbb{E}(x^2 p_{\min;\widetilde{z},\widetilde{xz}}(x)) - \mathbb{E}(x^2)$ and similarly rewrite $I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz})$. For claim 1, we proceed by writing $p_{\min} = p_{[b_1,b_2]}$ by Proposition 1 (suppressing the dependence of b_1 and b_2 on $(\widetilde{z}, \widetilde{xz})$ in our notation) and performing casework on the signs of b_1 and b_2 to show that $\mathbb{E}(x^2 p_{\min}(x)) \le \mathbb{E}(x^2)/2$ in each case. In the case $b_1 \le b_2 \le 0$, we have $\mathbb{E}(x^2 p_{\min}(x)) \le \mathbb{E}(x^2\mathbf{1}(x \le 0)) = \mathbb{E}(x^2)/2$ by symmetry; similarly, if $b_2 \ge b_1 \ge 0$ then $\mathbb{E}(x^2 p_{\min}(x)) \le \mathbb{E}(x^2\mathbf{1}(x \ge 0)) = \mathbb{E}(x^2)/2$. Next, if $b_1 \le 0 \le$

$-b_1 \le b_2$ then $F(b_2) + F(0) - F(b_1) = \Pr(b_1 < x \le b_2) + 1/2 = \mathbb{E}(p_{\min}(x)) + 1/2 < 1$ since $\widetilde{z} < 0$ implies $\mathbb{E}(p(x)) < 1/2$ by (5). Therefore,

$$\begin{aligned}\mathbb{E}\big(x^2 p_{\min}(x)\big) &= \mathbb{E}\big(x^2\mathbf{1}(b_1 \le x \le 0)\big) + \mathbb{E}\big(x^2\mathbf{1}(0 \le x \le b_2)\big)\\ &\le b_2^2 \Pr(b_1 \le x \le 0) + \mathbb{E}\big(x^2\mathbf{1}(0 \le x \le b_2)\big)\\ &= b_2^2 \Pr\big(b_2 \le x \le F^{-1}\big(F(b_2) + F(0) - F(b_1)\big)\big) + \mathbb{E}\big(x^2\mathbf{1}(0 \le x \le b_2)\big)\\ &\le \mathbb{E}\big(x^2\mathbf{1}\big(0 \le x \le F^{-1}\big(F(b_2) + F(0) - F(b_1)\big)\big)\big) \le \frac{\mathbb{E}(x^2)}{2},\end{aligned}$$

where the final inequality uses symmetry of F again. The final case $b_1 \le 0 \le b_2 \le -b_1$ follows by a symmetric argument. The proof of Claim 2 is completely analogous, with $p_{\max} = p_{[a_1,a_2]^c}$ by Proposition 1. □

We now proceed to prove the theorem. Given Claim 1, we have $\widetilde{x^2z}^*(\widetilde{z}, \widetilde{xz}; \mathrm{Eff}) = \min(I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz}), a^*(\widetilde{z}, \widetilde{xz}))$ by (31), and hence suppressing some $\widetilde{z}$ dependences

$$h(\widetilde{xz}) \equiv \det\big(M(p_{\mathrm{opt};\widetilde{z},\widetilde{xz}})\big) = \begin{cases} h^*(\widetilde{xz}), & g(\widetilde{xz}) \ge a^*(\widetilde{z}, \widetilde{xz}),\\ \det\big(M(p_{\max;\widetilde{z},\widetilde{xz}})\big), & g(\widetilde{xz}) \le a^*(\widetilde{z}, \widetilde{xz}),\end{cases}$$

where $g(\widetilde{xz}) \equiv I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz})$ and $h^*(\widetilde{xz})$ is defined by substituting $\mathbb{E}_p(x^2z) = a^*(\widetilde{z}, \widetilde{xz})$ into (29). We must show that $h(\widetilde{xz})$ is decreasing on $\widetilde{xz} > 0$.

First, we compute $h^*(\widetilde{xz}) = (\mathbb{E}(x^2))^2 M_{11}^2/(\mathbb{E}(x^2)(1 - \widetilde{z}^2))$ and note it is decreasing in $\widetilde{xz}$ since M_{11} is positive (Corollary 1) and decreasing in $\widetilde{xz}$ on $[0, \widetilde{xz}_{\max}(\widetilde{z})]$. Next, we show $\det(M(p_{\max;\widetilde{z},\widetilde{xz}}))$ is decreasing in $\widetilde{xz}$. Note (a_1, a_2) are the unique solutions to the system

$$F(a_1) + 1 - F(a_2) = (1 + \widetilde{z})/2,$$
$$\mathbb{E}\big(x\big(\mathbf{1}(x < a_1) + \mathbf{1}(x > a_2)\big)\big) = \widetilde{xz}/2.$$

By the implicit function theorem (e.g., de Oliveira (2018) since we do not require continuity of f)), it follows that $a_1 = a_1(\widetilde{xz})$ and $a_2 = a_2(\widetilde{xz})$ are differentiable and satisfy

(32) $$f(a_1)a_1'(\widetilde{xz}) - f(a_2)a_2'(\widetilde{xz}) = 0 \quad \text{and}$$

(33) $$a_1 f(a_1)a_1'(\widetilde{xz}) - a_2 f(a_2)a_2'(\widetilde{xz}) = 1/2.$$

Equations (32) and (33) imply that

$$g'(\widetilde{xz}) = \frac{\partial}{\partial \widetilde{xz}} \mathbb{E}_{p_{[a_1,a_2]^c}}(x^2 z) = 2a_1^2 f(a_1)a_1'(\widetilde{xz}) - 2a_2^2 f(a_2)a_2'(\widetilde{xz}) = a_1 + a_2 < 0.$$

The inequality follows by the assumption $\widetilde{z} < 0$, which ensures a_1 and a_2 must have different signs, and then noting that $\mathbb{E}_{p_{\max}}(xz) > 0$ requires $\mathbb{E}(xp_{\max}(x)\mathbf{1}(x > 0)) > -\mathbb{E}(xp_{\max}(x)\mathbf{1}(x < 0)) = \mathbb{E}(xp_{\max}(-x)\mathbf{1}(x > 0))$, the equality following by symmetry of F. Thus, for all $\widetilde{xz} > 0$ such that $g(\widetilde{xz}) \le a^*(\widetilde{z}, \widetilde{xz}) = -\widetilde{z}(\widetilde{xz})^2/(1 - \widetilde{z}^2)$ we have

$$\begin{aligned}\frac{\mathbb{E}(x^2)\partial \det(M(p_{\max}))}{\partial \widetilde{xz}} &= g(\widetilde{xz})\big(-2\big(1 - \widetilde{z}^2\big)g'(\widetilde{xz}) - 4(\widetilde{xz} \cdot \widetilde{z})\big)\\ &\quad - 2(\widetilde{xz})^2 \cdot \widetilde{z}g'(\widetilde{xz}) + 4(\widetilde{xz})\big((\widetilde{xz})^2 - \mathbb{E}(x^2)\big)\\ &\le \frac{4(\widetilde{xz})^3(\widetilde{z})^2}{1 - \widetilde{z}^2} + 4(\widetilde{xz})\big((\widetilde{xz})^2 - \mathbb{E}(x^2)\big) = -\frac{4M_{11} \cdot (\widetilde{xz})\mathbb{E}(x^2)}{1 - \widetilde{z}^2}.\end{aligned}$$

The RHS is negative (Corollary 1), so $\det(M(p_{\max;\widetilde{z},\widetilde{xz}}))$ is in fact decreasing in $\widetilde{xz} > 0$.

Finally, we fix $0 \le x_1 < x_2 \le \widetilde{xz}_{\max}(\widetilde{z})$ and show $h(x_1) > h(x_2)$. Note $\overline{g}(\widetilde{xz}) \equiv g(\widetilde{xz}) - a^*(\widetilde{z}, \widetilde{xz})$ is continuous in $\widetilde{xz}$, and $h^*(\widetilde{xz}) \ge \det(M(p_{\max;\widetilde{z},\widetilde{xz}}))$. We now carry out casework on the signs of $\bar{g}(x_1)$ and $\bar{g}(x_2)$.

$\overline{g}(x_1) \ge 0$*:* In this case,

$$h(x_1) = h^*(x_1) > h^*(x_2) \ge h(x_2). \tag{34}$$

$\overline{g}(x_1) < 0$ *and* $\overline{g}(x_2) \ge 0$*:* Define $S = \{x \in [x_1, x_2] | \overline{g}(x) \ge 0\}$, which contains x_2. Letting $x_3 = \inf S > x_1$, we have $\overline{g}(x_3) = 0$ and $\overline{g}(x) \le 0$ for $x \in [x_1, x_3]$, so

$$h(x_1) = \det\bigl(M(p_{\max;\widetilde{z},x_1})\bigr) > \det\bigl(M(p_{\max;\widetilde{z},x_3})\bigr) = h^*(x_3) \ge h^*(x_2) = h(x_2) \tag{35}$$

$\overline{g}(x_1) < 0$ *and* $\overline{g}(x_2) < 0$*:* In this case, either $\overline{g}(x) \le 0$ on $[x_1, x_2]$ (so $h(x_1) = \det(M(p_{\max;\widetilde{z},x_1})) > \det(M(p_{\max;\widetilde{z},x_2})) = h(x_2)$), or S as defined in the previous case is nonempty with $x_3 = \inf(S)$ and $x_4 = \sup(S)$ satisfying $x_1 < x_3 \le x_4 < x_2$ and $\overline{g}(x_3) = \overline{g}(x_4) = 0$. Then

$$h(x_1) \overset{(35)}{>} h(x_3) \overset{(34)}{\ge} h(x_4) \overset{(35)}{>} h(x_2),$$

which shows the theorem when $\widetilde{z} < 0$. The proof of the case $\widetilde{z} > 0$ is completely symmetric, and relies on Claim 2.

APPENDIX E: PROOF OF THEOREM 4

First, we fix $\widetilde{z} < 0$. It suffices to show that assuming $\mathbb{E}(x^2) < F^{-1}(1)^2$, there exists $\delta > 0$ such that $(\partial/\partial\widetilde{xz})\det(M(p^\dagger_{\text{opt};\widetilde{z},\widetilde{xz}})) > 0$ whenever $\widetilde{xz} \in (0, \delta)$, and that $\det(M(p^\dagger_{\text{opt};\widetilde{z},\widetilde{xz}}))$ is continuous in $\widetilde{xz}$ at $\widetilde{xz} = 0$.

From the assumed continuity of F and Proposition 2, we have $p^\dagger_{\max}(x) = p_{\ell,1,t}(x)$, with $\widetilde{z} > -1$ ensuring $F(t) > 0$. Again, we suppress the dependence of ℓ and t on $(\widetilde{z}, \widetilde{xz})$ in our notation for brevity. By the treatment fraction constraint $\mathbb{E}_{p_{\ell,1,t}}(z) = \widetilde{z}$, we must have $\ell = 1 - (1 - \widetilde{z})/(2F(t))$. From the short-term gain constraint $\mathbb{E}_{p_{\ell,1,t}}(xz) = \widetilde{xz}$, we see

$$\frac{\widetilde{xz}}{2} = (\ell - 1)\mathbb{E}\bigl(x\mathbf{1}(x < t)\bigr) = -\frac{1 - \widetilde{z}}{2F(t)}\mathbb{E}\bigl(x\mathbf{1}(x < t)\bigr).$$

We know by Proposition 2 and continuity of F that the two equations above have a unique solution $(\ell, t) = (\ell(\widetilde{xz}), t(\widetilde{xz}))$ for $\widetilde{xz} \in (0, \widetilde{xz}_{\max}(\widetilde{z}))$. Thus, we can differentiate both of the equations above with respect to $\widetilde{xz}$ to see that the derivatives of ℓ and t are given by

$$t' = t'(\widetilde{xz}) = \frac{F(t)^2}{(1 - \widetilde{z})f(t)\mathbb{E}((x - t)\mathbf{1}(x < t))} \quad \text{and} \quad \ell' = \ell'(\widetilde{xz}) = \frac{1}{2\mathbb{E}((x - t)\mathbf{1}(x < t))}.$$

Then $g(\widetilde{xz}) \equiv I_{\mathcal{F};\max}(\widetilde{z}, \widetilde{xz}) = 2(\ell - 1)\mathbb{E}(x^2\mathbf{1}(x < t)) + \mathbb{E}(x^2)$ is differentiable as well with

$$g'(\widetilde{xz}) = 2\ell'\mathbb{E}\bigl(x^2\mathbf{1}(x < t)\bigr) + 2(\ell - 1)t^2 f(t)t' = \frac{2(1 - \ell)t^2 F(t)^2 - (1 - \widetilde{z})\mathbb{E}(x^2\mathbf{1}(x < t))}{(1 - \widetilde{z})\mathbb{E}((t - x)\mathbf{1}(x < t))}.$$

Next, note that $g(0) = \widetilde{z}\mathbb{E}(x^2) < 0 = a^*(\widetilde{z}, 0)$, in the notation of (30). By differentiability (and thus continuity) of $a^*(\widetilde{z}, \cdot)$ and g (the latter due to differentiability of ℓ and t), we conclude that there exists $\epsilon > 0$ such that $a^*(\widetilde{z}, \widetilde{xz}) - g(\widetilde{xz}) \ge 0$ for all $\widetilde{xz} \in [0, \epsilon]$. By (31), this means $p^\dagger_{\text{opt};\widetilde{z},\widetilde{xz}} = p^\dagger_{\max,\widetilde{z},\widetilde{xz}}$ for all $\widetilde{xz} \in [0, \epsilon]$. Thus, it suffices to show $\frac{\partial}{\partial\widetilde{xz}}\det(M(p^\dagger_{\max;\widetilde{z},\widetilde{xz}})) > 0$ for all $\widetilde{xz} \in (0, \delta)$, for some $\delta \le \epsilon$. Continuity of $\det(M(p^\dagger_{\max;\widetilde{z},\widetilde{xz}}))$ at $\widetilde{xz} = 0$ follows immediately from continuity of g and (29).

As $\widetilde{xz} \downarrow 0$, we have $t(\widetilde{xz}) \uparrow F^{-1}(1)$ and $\ell(\widetilde{xz}) \uparrow (1 + \widetilde{z})/2$ and also $\mathbb{E}((t - x)\mathbf{1}(x < t)) = tF(t) - \mathbb{E}(x\mathbf{1}(x < t)) \uparrow F^{-1}(1)$. In the case $F^{-1}(1) < \infty$, we have $\lim_{\widetilde{xz}\downarrow 0} g'(\widetilde{xz}) =$

$F^{-1}(1) - \mathbb{E}(x^2)/F^{-1}(1) > 0$ by assumption. If $F^{-1}(1) = \infty$, then $g'(\widetilde{xz}) \to \infty$ as $\widetilde{xz} \downarrow 0$. Finally, we substitute into the formula (29) for $\det(M)$ getting

$$\frac{\partial \det(M(p^{\dagger}_{\max;\widetilde{z},\widetilde{xz}}))}{\partial \widetilde{xz}} = -\frac{2g'(\widetilde{xz})((1-\widetilde{z}^2)g(\widetilde{xz}) + \widetilde{z}(\widetilde{xz})^2)}{\mathbb{E}(x^2)} - \frac{4g(\widetilde{xz})\widetilde{z}(\widetilde{xz})}{\mathbb{E}(x^2)} + \frac{4(\widetilde{xz})^3}{\mathbb{E}(x^2)} - 4\widetilde{xz}.$$

Since $g(\widetilde{xz}) \to \widetilde{z} \cdot \mathbb{E}(x^2)$ as $\widetilde{xz} \downarrow 0$, we have $(1-\widetilde{z}^2)g(\widetilde{xz}) + (\widetilde{xz})^2\widetilde{z} \to \mathbb{E}(x^2)\widetilde{z}M_{11} < 0$ (Corollary 1). Our analysis of the limiting behavior on $g'(\widetilde{xz})$ then indicates that $\lim_{\widetilde{xz}\downarrow 0}(\partial/\partial\widetilde{xz})\det(M(p^{\dagger}_{\max;\widetilde{z},\widetilde{xz}})) = -2\widetilde{z}M_{11}(F^{-1}(1) - \mathbb{E}(x^2)/F^{-1}(1)) > 0$.

The proof for the case $\widetilde{z} > 0$ is completely analogous. We first show that $p^{\dagger}_{\mathrm{opt};\widetilde{z},\widetilde{xz}} = p^{\dagger}_{\min;\widetilde{z},\widetilde{xz}}$ whenever $\widetilde{xz}$ is sufficiently close to 0. Then we note (u, s) is the unique solution to the equations $u = (1+\widetilde{z})/(2(1-F(s)))$ and $\widetilde{xz}/2 = (1+\widetilde{z})\mathbb{E}(x\mathbf{1}(x \geq s))/(2(1-F(s)))$ to compute the derivatives $u'(\widetilde{xz})$ and $s'(\widetilde{xz})$. This enables us to show $\lim_{\widetilde{xz}\downarrow 0}(\partial/\partial\widetilde{xz})\det(M(p^{\dagger}_{\min;\widetilde{z},\widetilde{xz}})) > 0$ under the condition $\mathbb{E}(x^2) < F^{-1}(0)^2$.

Acknowledgments. The authors would like to thank Kevin Guo, Dan Kluger, Tim Morrison and several anonymous reviewers for helpful comments.

Funding. This work was supported by the U.S. National Science Foundation under grants IIS-1837931 and DMS-2152780.

REFERENCES

ABDULKADIROGLU, A., ANGRIST, J. D., NARITA, Y. and PATHAK, P. A. (2017). Impact evaluation in matching markets with general tie-breaking. Technical Report, National Bureau of Economic Research.

AIKEN, L. S., WEST, S. G., SCHWALM, D. E., CARROLL, J. L. and HSIUNG, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation: Efficacy of a university-level remedial writing program. *Eval. Rev.* **22** 207–244.

ANGRIST, J., AUTOR, D. and PALLAIS, A. (2020). Marginal effects of merit aid for low-income students. Technical Report, National Bureau of Economic Research.

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67. MR0655670 https://doi.org/10.1093/biomet/69.1.61

ATKINSON, A. C. (2014). Selecting a biased-coin design. *Statist. Sci.* **29** 144–163. MR3201860 https://doi.org/10.1214/13-STS449

ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS. Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

BANDYOPADHYAY, U. and BISWAS, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika* **88** 409–419. MR1844841 https://doi.org/10.1093/biomet/88.2.409

BISWAS, A. and BHATTACHARYA, R. (2018). A class of covariate-adjusted response-adaptive allocation designs for multitreatment binary response trials. *J. Biopharm. Statist.* **28** 809–823.

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

BRENT, R. P. (1973). *Algorithms for Minimization Without Derivatives. Prentice-Hall Series in Automatic Computation*. Prentice-Hall, Inc., Englewood Cliffs, NJ. MR0339493

CAMPBELL, D. T. (1969). Reforms as experiments. *Amer. Psychol.* **24** 409.

CATTANEO, M. D., TITIUNIK, R. and VAZQUEZ-BARE, G. (2017). Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality. *J. Policy Anal. Manage.* **36** 643–681.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Stat.* **24** 586–602. MR0058932 https://doi.org/10.1214/aoms/1177728915

CLYDE, M. and CHALONER, K. (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *J. Amer. Statist. Assoc.* **91** 1236–1244. MR1424621 https://doi.org/10.2307/2291742

COOK, R. D. and WONG, W. K. (1994). On the equivalence of constrained and compound optimal designs. *J. Amer. Statist. Assoc.* **89** 687–692. MR1294092

DANTZIG, G. B. and WALD, A. (1951). On the fundamental lemma of Neyman and Pearson. *Ann. Math. Stat.* **22** 87–93. MR0039962 https://doi.org/10.1214/aoms/1177729695

DE OLIVEIRA, O. (2018). The implicit function theorem for maps that are only differentiable: An elementary proof. *Real Anal. Exchange* **43** 429–443. MR3942588 https://doi.org/10.14321/realanalexch.43.2.0429

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660 https://doi.org/10.1093/biomet/58.3.403

GELMAN, A. and IMBENS, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *J. Bus. Econom. Statist.* **37** 447–456. MR3968384 https://doi.org/10.1080/07350015.2017.1366909

GOLDBERGER, A. S. (1972). Selection bias in evaluating treatment effects: Some formal illustrations. Technical Report No. Discussion paper 128–72, Institute for Research on Poverty, Univ. Wisconsin–Madison.

GOLDENSHLUGER, A. and ZEEVI, A. (2013). A linear response bandit problem. *Stoch. Syst.* **3** 230–261. MR3353472 https://doi.org/10.1214/11-SSY032

HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2245329 https://doi.org/10.1002/047005588X

HU, J., ZHU, H. and HU, F. (2015). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics. *J. Amer. Statist. Assoc.* **110** 357–367. MR3338508 https://doi.org/10.1080/01621459.2014.903846

JACOB, R., ZHU, P. and SOMERS, M. (2012). A practical guide to regression discontinuity. MDRC.

KLUGER, D. M. and OWEN, A. B. (2023). Kernel regression analysis of tie-breaker designs. *Electron. J. Stat.* **17** 243–290. MR4536118 https://doi.org/10.1214/23-ejs2102

LÄUTER, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist.* **5** 379–398. MR0440812 https://doi.org/10.1080/02331937408842203

LÄUTER, E. (1976). Optimal multipurpose designs for regression models. *Math. Operationsforsch. Statist.* **7** 51–68. MR0413385 https://doi.org/10.1080/02331937608842317

LEE, C. M.-S. (1987). Constrained optimal designs for regression models. *Comm. Statist. Theory Methods* **16** 765–783. MR0886553 https://doi.org/10.1080/03610928708829401

LEE, C. M.-S. (1988). Constrained optimal designs. *J. Statist. Plann. Inference* **18** 377–389. MR0926640 https://doi.org/10.1016/0378-3758(88)90114-0

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LIPSEY, M. W., CORDRAY, D. S. and BERGER, D. E. (1981). Evaluation of a juvenile diversion program: Using multiple lines of evidence. *Eval. Rev.* **5** 283–306.

LUDWIG, J. and MILLER, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Q. J. Econ.* **122** 159–208.

METELKINA, A. and PRONZATO, L. (2017). Information-regret compromise in covariate-adaptive treatment allocation. *Ann. Statist.* **45** 2046–2073. MR3718161 https://doi.org/10.1214/16-AOS1518

MORRISON, T. P. and OWEN, A. B. (2022). Optimality in multivariate tie-breaker designs. Technical Report, Stanford Univ. Available at arXiv:2202.10030.

NEYMAN, J. and PEARSON, E. S. (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A* **231** 289–337.

OWEN, A. B. and VARIAN, H. (2020). Optimizing the tie-breaker regression discontinuity design. *Electron. J. Stat.* **14** 4004–4027. MR4167610 https://doi.org/10.1214/20-EJS1765

R CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

ROSENBERGER, W. F. and SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statist. Sci.* **23** 404–419. MR2483911 https://doi.org/10.1214/08-STS269

STIGLER, S. M. (1971). Optimal experimental design for polynomial regression. *J. Amer. Statist. Assoc.* **66** 311–318.

SVERDLOV, O., ROSENBERGER, W. F. and RYEZNIK, Y. (2013). Utility of covariate-adjusted response-adaptive randomization in survival trials. *Stat. Biopharm. Res.* **5** 38–53.

THISTLETHWAITE, D. L. and CAMPBELL, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J. Educ. Psychol.* **51** 309.

TROCHIM, W. M. K. and CAPPELLERI, J. C. (1992). Cutoff assignment strategies for enhancing randomized clinical trials. *Control. Clin. Trials* **13** 190–212.

WHITTLE, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35** 123–130. MR0356388

ZHANG, L.-X. and HU, F. (2009). A new family of covariate-adjusted response adaptive designs and their properties. *Appl. Math. J. Chinese Univ. Ser. B* **24** 1–13. MR2486489 https://doi.org/10.1007/s11766-009-0001-6

ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *Ann. Statist.* **35** 1166–1182. MR2341702 https://doi.org/10.1214/009053606000001424