

Stochastic treatment choice with empirical welfare updating*

Toru Kitagawa[†]

Hugo Lopez[‡]

Jeff Rowley[§]

February 22, 2023

Abstract

This paper proposes a novel method to estimate individualised treatment assignment rules. The method is designed to find rules that are stochastic, reflecting uncertainty in estimation of an assignment rule and about its welfare performance. Our approach is to form a prior distribution over assignment rules, not over data generating processes, and to update this prior based upon an empirical welfare criterion, not likelihood. The social planner then assigns treatment by drawing a policy from the resulting posterior. We show analytically a welfare-optimal way of updating the prior using empirical welfare; this posterior is not feasible to compute, so we propose a variational Bayes approximation for the optimal posterior. We characterise the welfare regret convergence of the assignment rule based upon this variational Bayes approximation, showing that it converges to zero at a rate of $\ln(n)/\sqrt{n}$. We apply our methods to experimental data from the Job Training Partnership Act Study to illustrate the implementation of our methods.

Keywords: Empirical welfare maximisation, policy learning, PAC-Bayes learning, variational Bayes.

1 Introduction

The principal goal of programme evaluation is to inform the social planner as to which individuals within a target population should receive which treatment. When treatment effects are heterogeneous in individuals' observable characteristics, the social planner can improve social welfare by implementing an individualised treatment assignment rule based upon these characteristics. The literature on statistical treatment choice initiated by Manski (2004) studies how to estimate assignment rules based upon a finite sample and how to assess their welfare performance. Given an experimental or observational sample, existing approaches – including those proposed in Athey and Wager (2021), Hirano and Porter (2009), Kitagawa and Tetenov (2018), and Manski (2004) – yield deterministic assignment rules, which are functions mapping the individual's observable characteristics to a recommended treatment. That is, individuals that share the same observable characteristics are all assigned the same treatment. Such assignment rules are sharp and address the question of *who should be treated?* We adopt a broader view of the treatment choice problem by considering *stochastic assignment rules* that map individual observable characteristics to a probability distribution over the different treatment arms, instead addressing the question of *with what probability should an individual be treated?*

In static treatment choice problems with outcome distributions that exhibit the monotone likelihood ratio property, deterministic assignment rules form a class of admissible policies (Karlin and Rubin, 1956; Tetenov, 2012), such that restricting attention to this class is without loss (of welfare). Once we allow the

*The authors thank participants at the Bristol Econometrics Study Group and Midwest Econometrics Group conferences for beneficial comments. The authors also gratefully acknowledge financial support from the ERC (grant number 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP; grant number RES-589-28-0001).

[†]Department of Economics, Brown University. Email: toru_kitagawa@brown.edu

[‡]Department of Economics, University of Chicago.

[§]Department of Economics, University College London.

class of outcome distributions to be unconstrained though, there is little theoretical justification for focusing on deterministic rules. In comparison to stochastic assignment rules, deterministic assignment rules have the following three potentially undesirable features. First, deterministic assignment rules cannot incorporate confidence or uncertainty about which treatment is best for each individual, with individuals typically assigned treatment if conventional point estimates suggest that treatment has a positive effect on average. The strength of evidence in support of this conclusion, usually presented in the form of confidence intervals or p-values, is not generally acted upon; what matters is whether empirical evidence supports a positive point estimate, and not whether it is sufficient or insufficient to reject a non-positive effect. Such a sharp dichotomy of assignment is naturally overconfident in its prescription, and there is no theoretical basis for the incorporation of confidence intervals or p-values into frequentist-based decision-making. Stochastic assignment rules can represent such uncertainty, which arises due to the (finite sample) nature of experimental data or model misspecification, through their probability weighting of treatments. Second, stochastic assignment rules facilitate future evaluation, since implementing a stochastic assignment rule can generate a new experimental sample in which treatment is randomised conditional on individual observable characteristics. Third, unlike deterministic assignment rules for which the probability that a treatment is assigned changes discontinuously at some threshold, stochastic assignment rules feature assignment probabilities that smoothly change with respect to individual characteristics. Such a feature is desirable if a fairness criterion requiring that individuals with similar characteristics have similar probabilities of treatment (Dwork et al., 2012) is enforced.

This paper proposes novel and general methods for obtaining stochastic individualised assignment rules based on randomised control trial data. Assuming that the social planner assigns individuals to a binary treatment, with her goal being to maximise additive (utilitarian) social welfare as in Manski (2004), we exploit an empirical analogue of the social welfare criterion to generate individualised assignment probabilities. Specifically, we start with a prior distribution over a collection of deterministic assignment rules \mathcal{G} , each of which partitions the space of individual observable characteristics \mathcal{X} into a group of characteristics G and its complement G^c , thereby generating a deterministic assignment rule $g(x) = 1(x \in G)$. An individual is assigned treatment if their characteristics are such that $x \in G$, and is not assigned treatment if $x \in G^c$. We then update the prior distribution based upon an empirical analogue of the social welfare criterion to obtain a posterior distribution over \mathcal{G} . To generate a stochastic assignment, we draw a $g \in \mathcal{G}$ according to the posterior distribution over this collection, and implement the policy prescribed by g . In this way, the stochastic assignment rule applied to an individual with characteristics x treats her with the probability equal to the posterior probability that stochastic G contains x .

One of the main contributions of this paper is that we derive an optimal updating procedure for obtaining the posterior distribution over \mathcal{G} . This procedure minimises an upper bound on welfare regret and yields an exponential tilting of the prior over \mathcal{G} , where the tilting depends upon the empirical welfare criterion. This novel updating formula resembles the quasi-posterior distribution that appears in the Laplace-type estimation studied by Chernozhukov and Hong (2003), but differs in that the constant factor in the exponential tilting term is determined endogenously by the Lagrange multiplier of the optimisation.

Despite our analytical characterisation of the optimal posterior distribution, computation of this distribution or sampling of g from it is not straightforward. We therefore consider a variational approximation of the optimal posterior distribution by a parametric distribution. In particular, as a specification of \mathcal{G} , we consider the class of Linear Eligibility Score rules that assign treatment if $x^\top \gamma$ (the linear score) exceeds some threshold c (eligibility). Building upon the LES class, we exploit the invariance of the welfare criterion to the ratio of γ to c (scale invariance) and approximate the optimal posterior distribution using a multivariate von Mises-Fisher distribution.

For the practice of reporting and communicating individualised allocations of treatment, our approach of obtaining a posterior distribution over policies is useful for generating some quantities that existing methods yielding deterministic assignment rules cannot produce. First, the posterior probability that $g(x) = 1$ offers a personalised probabilistic assessment that individuals with characteristics x favour treatment over no treatment. Reporting such a probability offers a novel alternative to the common practice of using the p-values of hypothesis testing to express confidence in a positive treatment effect, something which does not easily translate to a recommendation about what the social planner should do. Second, viewing the posterior over g as an inferential tool for the welfare-optimality of (deterministic) assignment policies within \mathcal{G} , we can obtain a credible region for the optimal policy by, for instance, selecting its highest posterior

density region. This approach to obtaining confidence sets for the optimal treatment assignment policy is an alternative to the frequentist approach that is studied in Rai (2019). Third, analogous to the practice of using a Bayesian posterior with a noninformative prior as a visual summary of the likelihood function, we can use our variationally approximated posterior over g as a visual summary of the exponentiated empirical welfare criterion function.

To demonstrate how to implement our approach and what it delivers in practice, we apply our methods to the JTPA Study sample that is studied by Bloom et al. (1997). Given observations of prior earnings and years of education, we ask *with what probability should an individual be treated?* Restricting attention to linear assignment rules and a variational approximation of the optimal posterior distribution by a multivariate von Mises-Fisher distribution, we estimate a stochastic assignment rule that is more likely to allocate JTPA assistance to individuals with high prior earnings and fewer years of education. Kitagawa and Tetenov (2018), which similarly considers the JTPA Study sample, and estimates a deterministic assignment rule, serves as a useful benchmark for comparison. Aside from the obvious difference that we estimate a stochastic rule (i.e., every individual has a non-trivial probability to be allocated JTPA assistance under our rule), our estimated rule allocates JTPA assistance to a smaller fraction of the population than the deterministic rule of Kitagawa and Tetenov (2018), which targets individuals with low prior earnings and few years of education for treatment. This difference reflects the shape of the empirical welfare criterion, which can be captured by our approach but is missed by deterministic policies that are obtained as the mode of the empirical welfare criterion.

1.1 Literature review

This paper contributes to the growing literature on statistical treatment choice initiated by Manski (2004). Exact minimax regret assignment rules are studied in Ishihara and Kitagawa (2021), Schlag (2006), Stoye (2009), Stoye (2012), Tetenov (2012), and Yata (2021). Hirano and Porter (2009) analyses asymptotically-optimal assignment rules in limit experiments, and Bhattacharya and Dupas (2012) considers capacity constrained policies. Kitagawa and Tetenov (2018) proposes *Empirical Welfare Maximisation* (EWM) methods for individualised assignment, which maximise a sample analogue of the social welfare function over a constrained class of policies. Similar approaches have been studied in the literature on machine learning and personalised medicine, as in Beygelzimer and Langford (2009), Swaminathan and Joachims (2015), Zadrozny (2003), Zhang et al. (2012), and Zhao et al. (2012). Recent advances in learning individualised assignment policies include Adjaho and Christensen (2022), Athey and Wager (2021), D’Adamo (2021), Han (2022), Kido (2022), Kitagawa and Tetenov (2021), Kitagawa et al. (2021), Liu (2022), Mbakop and Tabord-Meehan (2021), Nie et al. (2021), Sakaguchi (2019), Sasaki and Ura (2020), Sun (2021), and Viviano (2021), to list but a few works. The assignment rules estimated in these works are all deterministic.

There are some earlier works that investigate the decision-theoretic justification for stochastic (fractional) assignment and the welfare performance of these rules, with Manski (2009) providing a detailed review of settings where stochastic rules are preferable. When the welfare criterion is only partially identified, minimax regret-optimal rules are stochastic given knowledge of the identified set (Manski, 2000a, 2005, 2007a,b), which remains true even after taking into account uncertainty of estimates of the bounds (Manski, 2022; Stoye, 2012; Yata, 2021). As shown by Manski and Tetenov (2007) and Manski (2009), stochastic assignment rules can also be justified by a nonlinear welfare criterion in a point-identified setting. Kitagawa et al. (2022) shows that, for a wide class of nonlinear welfare regret criteria, admissible assignment rules are stochastic (fractional). In particular, Kitagawa et al. (2022) shows that the minimax squared-regret rule is stochastic, with the probability of assignment equal to the posterior probability of a positive treatment effect under the least-favorable null. Kitagawa et al. (2022) proposes using this probability as a measure of the strength of evidence for a positive treatment effect, replacing the commonly used p-value of a hypothesis test. In contrast, this paper obtains the probability of assignment from a posterior probability distribution over assignment rules, rather than over the treatment effect parameters, with a quasi-likelihood built upon the empirical welfare criterion. Kock et al. (2022) obtains a stochastic assignment rule in a setting where the oracle optimal rule is fractional due to nonlinearity in the social planner’s chosen welfare criterion. In contrast to the static treatment assignment problem, dynamic treatment assignment problems analysed in the multi-arm bandit literature often consider stochastic assignments that balance the exploitation versus exploration trade-off, such as the posterior probability matching algorithm of Thompson (1933) does. Thompson

sampling algorithms build upon the standard Bayesian posterior distribution for treatment effects such that the allocation algorithm crucially relies on a parametric specification of the data generating process, which our approach does not require.

Chamberlain (2011) and Dehejia (2005) approach the treatment choice problem from a Bayesian perspective. In their framework, the potential outcome distributions are parametric, and it is over the parameters of these distributions that a prior is imposed. For the standard mean welfare criterion, the Bayes optimal allocation rule is deterministic. Our approach differs from these works in that we do not assume a prior distribution over the data generating process. We instead impose few restrictions on the data generating process, and form prior and posterior distributions over the parameters that index assignment rules. Our approach can be advantageous when compared to Chamberlain (2011) if the social planner is concerned about potential misspecification of the likelihood. If the likelihood is misspecified, the resulting Bayes-optimal assignment rule can be suboptimal even for large samples. In contrast, our approach is guaranteed to yield a distribution over policies that is guaranteed to concentrate on welfare-optimal policies without requiring a specification for the data generating process.

Our approach is also related to that of Bissiri et al. (2016) and Csaba and Szoke (2020), where loss function-driven (quasi-Bayes) updating rules are proposed. Rather than follow their approach by adopting exponentiated loss as a quasi-likelihood and solving the quasi-Bayesian decision problem, we obtain an optimal learning rule by minimising a high probability upper bound on welfare regret. This way of establishing optimality is similar to the structural risk minimisation approach of Vapnik (1998) and the *Probably Approximately Correct* (henceforth, PAC) analysis proposed by Valiant (1984), which was extended to the study of randomised predictors in McAllester (1999), and Shawe-Taylor and Williamson (1997), constituting the development of PAC-Bayes theory. For classification and regression problems, various PAC-Bayes bounds on prediction generalisation errors are obtained in Bégin et al. (2014, 2016), Catoni (2007), Derbeko et al. (2004), Germain et al. (2009), McAllester (2003), Pentina and Lampert (2015), and Seeger (2002), and can accommodate quasi-Bayesian procedures similar to ours. See Guedj (2019) for a recent review of this literature. To our knowledge, the PAC-Bayes bounds that we derive for treatment choice are new to the literature and offer a contribution of independent interest. We also note that Pellatt (2022) makes use of PAC-Bayes theory to analyse the treatment allocation problem with stochastic assignment rules under a budget (or resource) constraint.

Although the treatment choice problem is distinct from prediction problems – as is discussed in Kitagawa and Tetenov (2018) – the EWM approach for treatment choice is closely related to the cost-sensitive binary *classification* problem, as first pointed out by Zadrozny (2003). The PAC-Bayes classification analysis with variational posterior approximation that is proposed by Alquier et al. (2016) is, therefore, closely related to our analysis. There are, however, important differences with Alquier et al. (2016). First, we make use of the approach proposed by Bégin et al. (2016), which allows for the construction of a variety of different bounds via a general convex function. This introduces the complication that classification is not standard (i.e., cost is homogeneous) and is instead cost-sensitive. Introducing heterogeneity in the cost of misclassification leads to a non-trivial challenge in deriving the PAC bounds. To address these complications we leverage results in Maurer (2004) for continuous loss functions over the unit interval. Second, Alquier et al. (2016) considers approximating the optimal posterior distribution by a Gaussian distribution. We exploit the scale invariance property of the welfare criterion and approximate the optimal posterior distribution by a multivariate von Mises-Fisher distribution over the hypersphere.

2 Model and Setup

2.1 Notation and setting

We let $D \in \{1, 0\}$ be a binary treatment; we let $Y_1 \in \mathbb{R}$ and $Y_0 \in \mathbb{R}$ be the potential outcomes associated with the two treatment states; and we let $X \in \mathcal{X}$ be a vector of observed characteristics, which we refer to as covariates, where $\mathcal{X} \subset \mathbb{R}^{m-1}$. We suppose that experimental data comprising n independent and identically distributed observations concatenated as (Y_i, D_i, X_i) are available, where $Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$ is the post-treatment observed outcome of individual i . We denote the joint distribution of the experimental sample, which we reiterate is a probability distribution over (Y_i, D_i, X_i) , by P^n . The joint distribution of (Y_1, Y_0, D, X) that induces the independent and identical distribution of the observations (Y_i, D_i, X_i) ,

$i = 1, \dots, n$ is referred to as the data generating process and denoted by P . We assume that the covariates consist only of those characteristics that the planner can use to discriminate between individuals in the target population, with budgetary, ethical or legal considerations precluding the use of other characteristics.

Throughout our analysis, we maintain several assumptions. We follow Manski (2004) and the subsequent literature in supposing that the social welfare criterion is that of a utilitarian social planner who aims to maximise the average level of individual outcomes. We note that other criteria could also be implemented, such as inequality-averse social welfare and Gini social welfare.¹

Assumption 1 (External validity). *The population to which policy is to be applied – the target population – has the same distribution over (Y_1, Y_0, X) as the marginal distribution of (Y_1, Y_0, X) that is obtained from the data generating process.*

Assumption 2 (Unconfoundedness). *The data generating process satisfies $(Y_1, Y_0) \perp\!\!\!\perp D|X$.*

Assumptions 1 and 2 are satisfied, for instance, if the experimental data is extracted directly from the target population and the treatment is, conditional on the covariates, randomly assigned,² independently of the potential outcomes (Rosenbaum and Rubin, 1983).

Assumption 3 (Bounded outcomes). *There exists a constant $0 < M < \infty$ such that the support of Y is contained in $[0, M]$.*

Assumption 4 (Strict overlap). *There exists a $\psi \in (0, 1/2)$ such that the propensity score satisfies, for all $x \in \mathcal{X}$, $e(x) \in [\psi, 1 - \psi]$, where $e(x) = E_P(D|X)$.*

We normalise the outcome variable to the interval $[0, M]$. As is discussed in Kitagawa and Tetenov (2018) and Swaminathan and Joachims (2015), policies that maximise an empirical welfare criterion are not invariant to positive affine transformations of outcomes, which is the case for the empirical welfare criterion that we consider in this paper. Given Assumptions 3 and 4, we can define

$$H \equiv \frac{\psi \cdot Y/M}{e(X) \cdot D + (1 - e(X)) \cdot (1 - D)}, \quad (2.1)$$

which is confined to the unit interval, and which we interpret as weights and that are motivated by an unbiased estimator of the (scaled) expected potential outcomes. We solve the planner's problem using these transformed outcomes. This transformation of outcomes does not affect the welfare ranking over assignment policies, both in the population and according to in-sample welfare criteria, yet facilitates the proof of our PAC-Bayes bounds shown in Theorem 1 below.

Let $G \subset \mathcal{X}$ specify a set of individuals – identified by their individual characteristics – to be treated deterministically. We refer to G as a policy. Adopting an additive utilitarian perspective, the average level of social welfare attained by G is proportional to

$$W(G) \equiv E_P(Y_1 \cdot 1(X \in G) + Y_0 \cdot 1(X \notin G)). \quad (2.2)$$

Given Assumptions 1, 2 and 4 and that $Y = Y_0 + D \cdot (Y_1 - Y_0)$, we can re-write Equation (2.2) as

$$W(G) = E_P \left(\frac{Y \cdot D}{e(X)} \cdot 1(X \in G) + \frac{Y \cdot (1 - D)}{1 - e(X)} \cdot 1(X \notin G) \right), \quad (2.3)$$

$$= E_P \left(\frac{Y \cdot (1 - D)}{1 - e(X)} \right) + E_P \left(\left(\frac{Y \cdot D}{e(X)} - \frac{Y \cdot (1 - D)}{1 - e(X)} \right) \cdot 1(X \in G) \right). \quad (2.4)$$

¹For example, Kasy (2016) and Kitagawa and Tetenov (2021) study a setting where the social welfare function is a weighted average of the outcomes with rank-dependent weights, which includes the Gini social welfare function as a special case.

²Kitagawa and Tetenov (2018) considers a setting where the marginal distribution of X differs between the population of interest and the data generating process. Adjaho and Christensen (2022) and Kido (2022) study settings that differ also in terms of the distribution of potential outcomes.

Accordingly, the sample analogue of Equation (2.2) can be written as

$$W_n(G) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i \cdot d_i}{e(x_i)} \cdot 1(x_i \in G) + \frac{y_i \cdot (1-d_i)}{1-e(x_i)} \cdot 1(x_i \notin G) \right), \quad (2.5)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{y_i \cdot (1-d_i)}{1-e(x_i)} + \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i \cdot d_i}{e(x_i)} - \frac{y_i \cdot (1-d_i)}{1-e(x_i)} \right) \cdot 1(x_i \in G), \quad (2.6)$$

where $W_n(G)$ is an unbiased estimator for the true level of welfare that arises from the implementation of a particular G . Given the additive social welfare criterion, the maximal welfare level can be attained by a deterministic policy. Hence, as far as the population welfare maximisation problem is concerned, the social planner wants to select the G that maximises $W(G)$.

Each G can be associated with a binary function g that indicates membership in G . We refer to g as a deterministic assignment rule, or simply as an assignment rule, with \mathcal{G} constituting the class of assignment rules. With this notation to hand, we can write Equation (2.6), with some abuse of notation, as

$$W_n(g) = \frac{1}{n} \sum_{i=1}^n \frac{y_i \cdot (1-d_i)}{1-e(x_i)} + \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i \cdot d_i}{e(x_i)} - \frac{y_i \cdot (1-d_i)}{1-e(x_i)} \right) \cdot g(x_i), \quad (2.7)$$

$$= \frac{1}{n} \sum_{i=1}^n h_i \cdot \frac{M}{\psi} \cdot (e(x_i) \cdot d_i + (1-e(x_i)) \cdot (1-d_i)) \cdot \left[\frac{d_i}{e(x_i)} \cdot g(x_i) + \frac{1-d_i}{1-e(x_i)} \cdot (1-g(x_i)) \right], \quad (2.8)$$

$$= \frac{1}{n} \sum_{i=1}^n h_i \cdot \frac{M}{\psi} \cdot [d_i \cdot g(x_i) + (1-d_i) \cdot (1-g(x_i))], \quad (2.9)$$

$$= \frac{1}{n} \sum_{i=1}^n h_i \cdot \frac{M}{\psi} \cdot 1(g(x_i) = d_i), \quad (2.10)$$

$$= \frac{1}{n} \sum_{i=1}^n h_i \cdot \frac{M}{\psi} - \frac{1}{n} \sum_{i=1}^n h_i \cdot \frac{M}{\psi} \cdot 1(g(x_i) \neq d_i), \quad (2.11)$$

where h_i is the realisation of H , as defined in Equation (2.1), for observation (y_i, d_i, x_i) . We observe from Equation (2.11) that $W_n(g)$ depends upon g only through its second term, such that

$$\underset{g}{\operatorname{argmax}} W_n(g) = \underset{g}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n h_i \cdot 1(g(x_i) \neq d_i). \quad (2.12)$$

Accordingly, we define

$$R(g) \equiv E_P(H \cdot 1(g(X) \neq D)), \quad (2.13)$$

which we term the *welfare risk* of g , and its empirical analogue

$$R_S(g) \equiv \frac{1}{n} \sum_{i=1}^n h_i \cdot 1(g(x_i) \neq d_i), \quad (2.14)$$

which we term the *empirical welfare risk* of g . In view of Equation (2.12), the social planner's objective is to minimise Equation (2.13) via Equation (2.14) in g , following the empirical risk minimisation principle of Vapnik (1998).

One special set of policies that we draw particular attention to is the *Linear Eligibility Score* (LES) class that is defined in Kitagawa and Tetenov (2018), and that we denote by \mathcal{F} . Assignment rules in this class are indexed by a finite-dimensional parameter vector γ and a threshold c , and are associated with a binary function f_β that satisfies, for all $x \in \mathcal{X}$,

$$f_\beta(x) \equiv 1(x^\top \gamma \geq c), \quad (2.15)$$

where we take β to include both γ and c (i.e., β is an m -dimensional vector). Each LES rule induces a partitioning of the covariate space into two half-spaces, such that individuals in the upper contour set receive treatment and individuals in the lower contour set do not. By restricting β to the unit hypersphere (i.e., the Euclidean length of β is one), we guarantee that each policy is associated with a unique β . In what follows, we exploit the interchangeability of β and the LES rule that it indexes, adopting β as the argument of the loss functions that we consider. For instance, and again with some abuse of notation, whenever we focus on the LES class of decision rules we write

$$R(\beta) = \mathbb{E}_P(H \cdot 1(f_\beta(X) \neq D)), \quad (2.16)$$

and

$$R_S(\beta) = \frac{1}{n} \sum_{i=1}^n h_i \cdot 1\{f_\beta(x_i) \neq d_i\}, \quad (2.17)$$

respectively, in place of Equations (2.13) and (2.14).

2.2 Posterior over policies as a stochastic assignment rule

We now adapt Equations (2.13) and (2.14) to handle stochastic assignment rules. We let Π denote a probability distribution over \mathcal{G} that we interpret as a posterior distribution, assuming that \mathcal{G} can be embedded in a measurable space.³ We let \mathcal{M} denote the collection of all posterior distributions.

Definition 1 (Posterior assignment rule). *Let Π be a probability distribution over \mathcal{G} that is constructed upon observing the sample. The posterior assignment rule under Π is a stochastic assignment rule that assigns individuals with $x \in \mathcal{X}$ to treatment with probability $Q^\Pi(x) \equiv \int_{\mathcal{G}} g(x) \cdot d\Pi$.*

To implement posterior assignment rules in practice, we randomly draw a g from \mathcal{G} according to Π for each individual in the target population. In this way, similar individuals, who can have similar assignment probabilities, can be assigned to different treatment arms. Moreover, this approach does not require computation of the probability of treatment.

Definition 2 (Expected welfare risk under Π). *We define the expected welfare risk under Π as*

$$\begin{aligned} R^\Pi &\equiv \int_{\mathcal{G}} R(g) \cdot d\Pi(g), \\ &= \mathbb{E}_P(H \cdot (D \cdot (1 - Q^\Pi(X)) + (1 - D) \cdot Q^\Pi(X))), \end{aligned} \quad (2.18)$$

with its empirical analogue taking the form

$$\begin{aligned} R_S^\Pi &\equiv \int_{\mathcal{G}} R_S(g) \cdot d\Pi(g), \\ &= \frac{1}{n} \sum_{i=1}^n h_i \cdot (d_i \cdot (1 - Q^\Pi(x_i)) + (1 - d_i) \cdot Q^\Pi(x_i)). \end{aligned} \quad (2.19)$$

The interpretation of R^Π is the average welfare loss that the social planner expects from stochastic implementation of g in \mathcal{G} in the target population when g is distributed according to Π .

We reiterate that stochastic assignment is achieved by randomly drawing g according to Π . This way of selecting assignment rules is reminiscent of the *Gibbs classifier* in statistical learning theory (Germain et al., 2009) and might offer a computational advantage over other methods if drawing g according to Π is easier than maximising empirical welfare or finding the mode of Π , say. An advantage of stochastic assignment is the possibility for sequential treatment evaluation: the induced assignment of treatment and non-treatment to individuals in the target population by Π is random conditional on X , which allows for estimation of the causal effect of treatment in future studies. In this sense, stochastic assignment is well suited to balancing

³For \mathcal{G} to be embedded in a measurable space, \mathcal{G} cannot be too rich. We defer to Molchanov (2005) and Gunsilius (2019) for further discussion of this point.

existing evidence about what constitutes the optimal assignment for each individual against the benefit of further exploration of the treatment effect (Manski, 2000b). We do not, however, study this channel and focus on a purely static problem in this paper.

Our framework allows the social planner to hold some prior as to what constitutes the best policy. We differentiate the subjective beliefs that the social planner holds, which we encode using the prior distribution π_0 , and their updated beliefs following their observation of sample data, which we encode using Π . In the analysis that follows, we provide finite sample regret guarantees in the form of PAC-Bayes bounds for stochastic assignment. The approach discussed here differs significantly from other Bayesian treatment choice settings, such as those discussed in Chamberlain (2011), because we do not impose any kind of prior belief on P . We instead choose to model the beliefs that the social planner has regarding the optimal policy. Using a decision procedure that is free from specification of the likelihood comes with a desirable robustness property, as we discuss in due course.

Aside from its more conventional role as a means of expressing existing information about what constitutes the best policy, π_0 can also play several other roles within our framework. For instance, π_0 can also embed any constraints that are imposed upon the set of policies through truncation of its support. Such a zero density condition can be imposed in lieu of an explicit restriction on \mathcal{G} and is easy to implement in practice via rejection sampling, with only those policies that satisfy any budgetary, ethical or legal constraints being retained under the sampling procedure. Moreover, π_0 can also be used to describe the status quo, with restrictions on the shape of π_0 governing how much policy can deviate. These interpretations of π_0 naturally extend to Π .

3 Optimal stochastic assignment and convergence of welfare

3.1 Bounding expected welfare risk

Seeing as experimental data provides only an insight into the welfare performance of any policy in the target population, a very natural question to ask is how much we can expect R_S^Π to differ from R^Π for any given Π . We provide an answer to this question here.

Theorem 1. *Suppose that Assumptions 1 to 4 are satisfied, and $n \geq 8$. Then, for any $0 < \epsilon < 1$ and $\Pi \in \mathcal{M}$ that is absolutely continuous with respect to π_0 , the following inequality holds with probability at least $1 - \epsilon$ in terms of P^n :*

$$R^\Pi \leq R_S^\Pi + \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi \| \pi_0) + \ln\left(\frac{2\sqrt{n}}{\epsilon}\right) \right)}. \quad (3.1)$$

We present a proof of this theorem in Appendix A.1. The proof builds upon Bégin et al. (2016, §Lemma 3), which offers a flexible approach that allows for the recovery of many different PAC-Bayes bounds. The approach centres around a convex function of R^Π and R_S^Π , which we specify so as to recover the form that is presented in McAllester (2003, see §(8) in Bégin et al., 2016). We leverage results in Maurer (2004, §Lemma 3 and §Theorem 1), exploiting the properties of Bernoulli random variables and convex functions, to adapt Bégin et al. (2016) and the bound that is presented therein to allow for heterogeneous cost (in lieu of the standard binary loss function that is prevalent in the classification literature and that is studied in Bégin et al., 2016).

We note that Theorem 1 holds for any rules that update π_0 and deliver Π , and that we have not committed to any particular updating rule to obtain Equation (3.1). A feature of Equation (3.1) is that the regularisation term – the square root term containing the Kullback-Leibler divergence of Π from π_0 – enters additively, which we find is a convenient feature for establishing convergence of our variational approximation (regularisation can otherwise be effected multiplicatively).⁴ The regularisation term, by design, prevents overfitting. To illustrate this point, suppose that π_0 is uniform over \mathcal{G} : the best response of the social planner absent regularisation is to concentrate probability mass on the optimal (in-sample) g as suggested by data, such that Π is degenerate. In the presence of regularisation, however, this is no longer a best response, since the Kullback-Leibler divergence infinitely penalises degeneracy *vis-à-vis* uniformity. Rather, the best response of

⁴See Bégin et al. (2016) for the implications of different convex functions.

the social planner is to allocate probability mass on all g in \mathcal{G} , albeit concentrating more mass on those g that are associated with low empirical welfare risk. Put differently, the regularisation term controls how far away from π_0 a stochastic assignment rule can be, with this difference governed by the number of observations in the sample.

The Vapnik-Chervonenkis (VC) dimension is the standard measure of complexity in the statistical learning literature. We instead associate complexity with the Kullback-Leibler divergence. If one is willing to impose distributional constraints on a posterior assignment rule, an advantage of the PAC-Bayes approach and of using the Kullback-Leibler divergence is that complexity is then purely in terms of the selected (stochastic) assignment rule Π rather than in terms of the class of possible stochastic assignment rules \mathcal{M} or the class of underlying deterministic rules \mathcal{G} . As such, Theorem 2 does not explicitly require any assumption about the VC dimension of \mathcal{G} . The influence of VC dimension for \mathcal{G} is implicit in our setting: the upper bound on the difference between R_S^Π and R^Π implied by Equation (3.1) is governed by the Kullback-Leibler divergence, which is increasing in the dimension of the support of Π and π_0 , and so is non-decreasing in the complexity of \mathcal{G} .

3.2 Optimal updating rule

In a standard Bayesian setting, unknown parameters index the distribution of data and inference on parameters is conducted with respect to the posterior distribution. Typically, the posterior distribution is constructed from a well-defined likelihood function via Bayes' theorem. We leverage Theorem 1 to construct Π from π_0 and R_S^Π . This approach is valid since Theorem 1 holds for all $\Pi \in \mathcal{M}$. We emphasise that the posterior distribution that we construct is over assignment rules rather than over the data generating processes, which is distinct from Bissiri et al. (2016) and Csaba and Szoke (2020).

Following McAllester (2003) and Germain et al. (2009), we define an optimal posterior distribution, which we denote by Π^* , as a distribution over \mathcal{G} that minimises the right hand side of Equation (3.1). That is, Π^* minimises

$$R_S^\Pi + \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi \| \pi_0) + \ln\left(\frac{2\sqrt{n}}{\epsilon}\right) \right)}. \quad (3.2)$$

Theorem 2. *The optimal posterior Π^* over \mathcal{G} satisfies*

$$d\Pi^*(g) \equiv \frac{\exp(-\chi \cdot R_S(g))}{\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g)} \cdot d\pi_0(g), \quad (3.3)$$

where

$$\chi \equiv 4n \cdot \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi^* \| \pi_0) + \ln\left(\frac{2\sqrt{n}}{\epsilon}\right) \right)}. \quad (3.4)$$

We present a proof of this theorem in Appendix A.2.

The posterior distribution that we derive is analogous to the optimal posterior in McAllester (2003) with the difference that our observations are mapped to the unit interval rather than to $\{-1, 1\}$, which is the standard support in classification. This particular distribution is common in the statistical mechanics literature and is a Boltzmann (or Gibbs) distribution, and has the form of exponential tilting of the prior, where the exponential tilting term involves the negative empirical welfare risk. The degree of tilting depends upon the magnitude of $\chi > 0$, which is the inverse of the Lagrange multiplier of the associated minimisation problem and that corresponds to the root of Equation (3.4). The Lagrange multiplier controls the extent to which π_0 is updated by empirical welfare risk in minimising the upper bound for R^Π .

Although Theorem 2 offers an analytical characterisation of the optimal posterior, we are unable to obtain a closed-form expression for the optimal posterior density. Whilst Equation (3.4) does suggest a means to compute this density, it is likely that, in practice, this computation is difficult to perform with any precision. Given that we have, however, established that this density exists, we can consider approximating it using the *variational approximation* method, as is considered in Alquier et al. (2016).

3.3 Variational approximation of the optimal stochastic assignment rule

We develop a variational approximation of the optimal posterior density. Variational approximation is useful in situations where Gibbs distributions are difficult to sample from directly, such as is the case for graphical models where Markov Chain Monte Carlo (MCMC) sampling is costly (Wainwright and Jordan, 2008). In variational approximation, we choose to approximate the optimal posterior distribution via a family of distributions of our choice, $\mathcal{V} \subset \mathcal{M}$. This allows us to develop an analytically tractable upper bound for the welfare regret attained by the resulting stochastic assignment rule.

Aside from guaranteeing tractability in estimation, variational approximation can also be motivated as a convenient way to impose constraints on the set of policies. For instance, a fairness criterion requiring that individuals with similar characteristics have similar probabilities of treatment (Dwork et al., 2012) can be enforced by specifying that \mathcal{V} is a continuous family, and by limiting the concentration of the density. A similar approach can be used if \mathcal{V} is a parametric family to limit how much policy can deviate from the status quo, by fixing the parameters of the posterior distribution or restricting them to some set, say.

We approximate Π^* (implicitly defined in Theorem 2) by minimising the right-hand side of Equation (3.1) with respect to posterior distributions in \mathcal{V} , defining

$$\tilde{\Pi} \equiv \underset{\Pi \in \mathcal{V}}{\operatorname{argmin}} \left\{ R_S^\Pi + \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi \| \pi_0) + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)} \right\}. \quad (3.5)$$

We then use this optimal variational posterior distribution to define a new bound for welfare regret from which we can characterise its convergence rate.

Lemma 1. *With probability $1 - \epsilon$ in terms of P^n , the expected welfare risk under the optimal variational posterior satisfies*

$$R^{\tilde{\Pi}} \leq \inf_{\Pi \in \mathcal{V}} \left\{ R^\Pi + \frac{a(\lambda, n)}{\lambda} + \frac{\text{KL}(\Pi \| \pi_0)}{\lambda} + \frac{\ln(\frac{2}{\epsilon})}{\lambda} + \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi \| \pi_0) + \ln \left(\frac{4\sqrt{n}}{\epsilon} \right) \right)} \right\}, \quad (3.6)$$

where $\epsilon \in (0, 1)$ and $\lambda > 0$ are arbitrary constants and a is a function that depends upon a positive constant λ and the sample size n .

In what follows, we let \mathcal{V} be a variational family of distributions that assigns positive density to assignment rules in \mathcal{F} only. Equivalently, we let \mathcal{V} be a directional family that assigns positive density to unit vectors on the hypersphere, recalling that every policy in the LES class can be uniquely associated with a unit vector (see Equation (2.15) and surrounding discussion). A particularly tractable directional family, and one that we use, is the von Mises-Fisher family of distributions, which is characterised by a probability density function satisfying, for all $\kappa > 0$ and m -dimensional unit vectors μ ,

$$d\Pi(\beta; \kappa, \mu) \equiv \frac{\kappa^{m/2-1} \cdot \exp(\kappa \cdot \mu^\top \beta)}{(2\pi)^{m/2} \cdot I_{m/2-1}(\kappa)} \cdot d\beta, \quad (3.7)$$

where $I_\nu(z)$ is a modified Bessel function of the first kind with order ν and argument $z > 0$. A von Mises-Fisher distribution is the analogue of a multivariate Gaussian distribution on the unit hypersphere. We refer to κ as the concentration parameter and to μ as the mean direction (the location parameter), noting that the distribution becomes degenerate as $\kappa \rightarrow \infty$ and uniform as $\kappa \rightarrow 0$.

We now introduce some further notation that facilitates our analysis. First, we let $\bar{R} = \inf_{\Pi \in \mathcal{M}} R^\Pi$ denote the minimum expected welfare risk amongst assignment rules in \mathcal{F} and $\bar{\beta}$ denote the vector that parametrises the assignment rule that induces \bar{R} . We also add the following assumption that restricts the marginal distribution of X .

Assumption 5 (Margin assumption). *There exists a constant $c > 0$ such that, for any m -dimensional unit vectors β^\dagger and β^\ddagger , $P(\langle X, \beta^\dagger \rangle \cdot \langle X, \beta^\ddagger \rangle \leq 0) \leq c \cdot \|\beta^\dagger - \beta^\ddagger\|_2$.*

This assumption is satisfied whenever X has bounded density on the unit hypersphere and is also present in the analysis of Alquier et al. (2016). An interpretation of this assumption is that the proportion of indi-

viduals whose treatment status switches is continuous with respect to the linear eligibility score coefficients, with an implication being that $R(\beta) - \bar{R} \leq 2c \cdot \|\beta - \bar{\beta}\|_2$.

We now present a high-probability uniform upper bound for the welfare regret of the stochastic assignment rule obtained by variational approximation via the von Mises-Fisher family of distributions.

Theorem 3. *Suppose that Assumptions 1 to 5 are satisfied, that π_0 is a uniform distribution over the unit hypersphere, and that Π is a von Mises-Fisher distribution. Then for $n \geq 8$, with probability at least $1 - \epsilon$ in terms of P^n ,*

$$R^{\tilde{\Pi}} - \bar{R} \leq M \cdot \frac{\ln(n)}{\sqrt{n}}, \quad (3.8)$$

where M is a universal constant.

We present proof of this theorem and provide an analytical expression for the universal constant in Appendix A.4.

The uniform upper bound on welfare regret that is defined by Equation (3.8) decays at a rate of $\ln(n)/\sqrt{n}$. This rate is slightly slower than the welfare regret convergence rate of the EWM (deterministic) assignment rule studied in Kitagawa and Tetenov (2018). A simple comparison of these rates, however, is not quite meaningful for the following reason: we do not know if the convergence rate of $\ln(n)/\sqrt{n}$ that is obtained in Theorem 3 is sharp or not. Theorem 3 requires Assumption 5, whilst Kitagawa and Tetenov (2018) does not impose this assumption in showing that $1/\sqrt{n}$ is the minimax optimal rate of welfare regret convergence.⁵ We do not know what the minimax optimal rate of welfare regret convergence is when Assumption 5 is additionally imposed and, hence, cannot rule out the possibility that the convergence rate of Theorem 3 can be improved upon and made faster than $1/\sqrt{n}$. We leave further investigation of this matter for future research.

It is worth emphasising that the regret convergence result of Theorem 3 imposes weak restrictions on the distribution of data (Assumptions 3 - 5) and does not require the specification of a likelihood function or of regression equations. This contrasts with other approaches such as a Bayesian approach, where misspecification of likelihood can lead to non-convergence of the welfare regret even when Bayes optimal policies are constrained to deterministic ones in \mathcal{F} .

Our approach is similar to Alquier et al. (2016) in which the families of distributions for variational approximation are multivariate Gaussian distributions on Euclidean space with flexible covariance matrices. In our approach, we stipulate a class of von Mises-Fisher distributions, which are the hyperspherical analogues of multivariate Gaussian distribution with diagonal covariance matrices featuring a constant variance element. Since the empirical welfare criterion for LES rules is invariant to the scale of β , it is natural to consider von Mises-Fisher distributions rather than Gaussian ones. The scale invariance of von Mises-Fisher distributions can simplify optimisation of the variational approximation by reducing the set of optima to a singleton.

4 Implementation

To implement our procedure, we restrict attention to \mathcal{F} and let \mathcal{V} be the von Mises-Fisher family of distributions. Our goal is then to minimise the objective function,

$$R_S^{\Pi} + \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi|\pi_0) + \ln\left(\frac{2\sqrt{n}}{\epsilon}\right) \right)}, \quad (4.1)$$

with respect to κ and μ , which are the parameters of our chosen variational family (the concentration parameter and the mean direction, respectively). Here, we reiterate that ϵ relates to the probability with

⁵Kitagawa and Tetenov (2018, §Theorem 2.3 and §Theorem 2.4) establishes that the minimax-optimal rate under a stronger condition than our Assumption 5 is $1/n^{2/3}$. This stronger condition embeds a margin assumption that implies our Assumption 5 but also embeds the requirement that the first-best treatment rule is contained in the set of admissible decision rules – that \mathcal{F} contains the deterministic assignment rule that minimises expected welfare risk amongst all assignment rules. We do not make any assumption about whether \mathcal{F} (or indeed \mathcal{G} in earlier parts of our analysis) contains the first-best assignment rule, or whether this rule is deterministic or stochastic.

which our high probability bounds hold. We assume that π_0 is the uniform distribution over the sphere and set ϵ equal to the 5% level throughout.

We propose numerically minimising the objective function, approximating R_S^{Π} using Monte Carlo draws of β from the von Mises-Fisher distribution for a given realisation of data and for fixed values of the parameters of the von Mises-Fisher distribution. We let $\{\beta^j : j = 1, \dots, J\}$ be the pseudo-random draws that we obtain, and compute

$$\hat{R}_S^{\Pi} \equiv \frac{1}{n} \sum_{i=1}^n h_i \cdot \left(D_i \cdot \left(1 - \hat{\Pi}_i \right) + (1 - D_i) \cdot \hat{\Pi}_i \right), \quad (4.2)$$

where, for all $i = 1, \dots, n$,

$$\hat{\Pi}_i \equiv \frac{1}{J} \sum_{j=1}^J \mathbf{1} \left(x_i^\top \beta^j \geq 0 \right). \quad (4.3)$$

Fast pseudo-random sampling of von Mises-Fisher random vectors is possible using the rejection sampling method of Wood (1994) or the inversion method of Kurz and Hanebeck (2015). The analogue of Equation (4.1) that we then minimise is

$$\hat{R}_S^{\Pi} + \sqrt{\frac{1}{2n} \cdot \left(\nu \cdot \ln \left(\frac{\kappa}{2} \right) - \ln(I_\nu(\kappa)) - \ln(\Gamma(\nu + 1)) + \frac{I_{\nu+1}(\kappa)}{I_\nu(\kappa)} \cdot \kappa + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)}, \quad (4.4)$$

which follows from our assumption of a uniform prior and the form that the Kullback-Leibler divergence takes under that assumption (Kitagawa and Rowley, 2022).

As is shown in Kitagawa and Rowley (2022), the Kullback-Leibler divergence of the von Mises-Fisher distribution from the uniform distribution over the hypersphere does not depend upon μ and increases at the logarithmic rate in κ . In contrast, although R_S^{Π} is a function of μ and κ , the nature of this relationship is not clear. That several combinations of μ and κ induce local minima of R_S^{Π} cannot be ruled out for instance. Moreover, since we construct \hat{R}_S^{Π} based upon random draws of β from Π , our objective function is not a smooth function of μ and κ . Given this, we suggest that a grid-based search for the minimum of the objective function is appropriate, with this search limited to values of κ between zero and some specified upper limit. Further insight about the behaviour of the objective function is provided in an online appendix.

5 Empirical illustration

We illustrate our procedure using data from the National Job Training Partnership Act (JTPA) Study. Applicants to the Study were randomly allocated to one of two groups. Applicants allocated to the treatment group were extended training, job search assistance and other services provided by the JTPA over a period of 18 months. Applicants allocated to the control group were excluded from JTPA assistance. Along with information collected prior to the commencement of the intervention, the Study also collected administrative and survey data relating to applicants' earnings in the 30 months following its start. Further details about the data and the Study can be found elsewhere (see, for instance, Bloom et al., 1997). We restrict attention to a sample of 9,223 observations for which data on years of education and pre-programme earnings amongst the sample of adults (aged 22 years and older) used in the original evaluation of the programme and in subsequent studies (Abadie et al., 2002; Bloom et al., 1997; Heckman et al., 1997) is available. Applicants in this sample were assigned to the treatment group with a probability of two thirds. Like Kitagawa and Tetenov (2018), we define D to be the initial assignment of treatment, rather than the actual take-up due to the presence of non-compliance in the experiment. We study stochastic assignment rules.

We follow Kitagawa and Tetenov (2018) in considering total individual earnings in the 30 months after programme assignment as our principal welfare measure. Moreover, we focus exclusively on the class of linear rules,

$$\begin{aligned} \mathcal{F} &= \{x : \beta_0 + \beta_1 \cdot \text{prior earnings} + \beta_2 \cdot \text{education} \geq 0 \mid \|(\beta_0, \beta_1, \beta_2)\|_2 = 1\}, \\ &= \{x : \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 \geq 0 \mid \|(\beta_0, \beta_1, \beta_2)\|_2 = 1\}, \end{aligned} \quad (5.1)$$

that are studied in that paper.

[Figure 1 about here.]

To implement our procedure, we map prior earnings and education to the unit interval,⁶ and calculate H as outlined in Equation (2.1). We perform this calculation without adjusting post-programme earnings by the average cost of JTPA assistance (\$774 per individual) for treated individuals.⁷ We then utilise a grid search approach over the parameters of the von Mises-Fisher distribution, specifying a reasonably fine grid over the unit sphere and over a finite subset of the reals.⁸

For each point on our grid, we draw 1,000 values of β from the corresponding von Mises-Fisher distribution and approximate empirical welfare risk as per Equation (4.2). We then substitute these values into Equation (4.4) to provide an estimate of the objective function.⁹

We find that the objective function is minimised (amongst the class of von Mises-Fisher distributed linear assignment rules) by the stochastic assignment rule with $\kappa = 1.550$ and $\mu = (0.883, 0.442, 0.158)$,¹⁰ which we label as κ^* and μ^* , respectively.

[Figure 2 about here.]

The optimal stochastic assignment rule, on average, assigns treatment to individuals in the JTPA Study sample around 83% of the time. This probability is not, however, uniform, and there is some variation in the probability with which distinct individuals are assigned treatment. This variation in assignment propensity can be seen in Figure 1, which plots the individual characteristics of all individuals in the sample. The propensity with which individuals with distinct characteristics are assigned treatment is represented by the color of each point, and the weight given to individuals in the sample with particular characteristics is represented by the size of each point. The weight attached to a given point is proportional to the sum of post-programme earnings over all individuals with those characteristics.¹¹ Figure 1 shows that the optimal stochastic assignment rule is more likely to assign treatment to an individual with few years of education and high prior earnings than an individual with more years of education and lower prior earnings, with the assignment probability ranging from 78% to 84%. The deterministic assignment rule of Kitagawa and Tetenov (2018), in contrast, assigns only individuals with few years of education and low prior earnings to treatment, with around 93% of individuals assigned treatment. We plot this deterministic rule as a useful benchmark for comparison in Figure 1.

It is important to emphasise that it is the regularisation term and, in particular, the Kullback-Leibler divergence that limits the value of κ at the optimum, and leads to an interior probability of assignment for all individuals. This can be seen in Figure 2, which plots the USD equivalent of the objective function (left-hand axis, solid line) and of empirical welfare risk (right-hand axis, dashed line) for a range of κ , holding fixed μ at μ^* . We observe that empirical welfare risk decreases as κ increases, remaining low and constant once its value is sufficiently large.¹² The intuition here is that large values of κ lead to stochastic assignment rules that mimic deterministic ones; we expect μ to eventually coincide with the deterministic assignment

⁶We map each variable to the unit interval by dividing through by its maximum in the sample. Kitagawa and Tetenov (2018) also does this. Such a change of units is useful when the domain of one variable is much larger than the domain of another and the respective coefficients on the two variables reflect this. For instance, in our sample, every individual has between seven and 18 years of education, and no individual earned more than \$63,000 prior to the start of the intervention.

⁷We adjust post-programme earnings by the average cost of JTPA assistance in an online appendix.

⁸We design our grid so as to place an upper limit on the great-circle distance between any point on the sphere and its closest point on the grid. Our grid comprises a total of 10,116 directional vectors combined with a sequence of evenly-spaced concentrations on the $[0, 5]$ interval. For reference, the surface area of the sphere is 4π , which means that our grid has a density of approximately 0.001.

⁹In Equation (4.4), given Equation (5.1) and its restriction of β to the unit sphere, $\nu = 1/2$. More generally, maintaining our convention of defining $\mathcal{X} \in \mathbb{R}^{m-1}$ and adding to this vector a constant (i.e., an intercept), $\nu = m/2 - 1$.

¹⁰This directional vector can be represented by an azimuth of 27° and an inclination of 81° using spherical coordinates.

¹¹To simplify Figure 1, we scale the weights such that they sum to one.

¹²That empirical welfare risk decreases for small to moderate values of κ is specific to the data and chosen μ , and is arguably also attributable to the lack of consideration given to the cost of treatment.

rule of Kitagawa and Tetenov (2018) as the value of the concentration parameter approaches infinity, since there does not exist a (linear) deterministic rule that can improve upon this. Tempering this preference towards large values of κ is the Kullback-Leibler divergence of the von Mises-Fisher distribution from the uniform distribution, which is increasing at the logarithmic rate in κ and generates the difference between the objective function and empirical welfare risk in Figure 2. As κ increases, the regularisation term begins to dominate.

[Figure 3 about here.]

[Figure 4 about here.]

To better understand why μ^* does not coincide with the deterministic assignment rule of Kitagawa and Tetenov (2018) holding fixed κ at κ^* , we refer to Figure 3, which plots empirical welfare risk for all vectors on the unit sphere – i.e., the empirical welfare risk associated for each deterministic assignment rule in \mathcal{F} . Figure 3 utilises the spherical coordinate system

$$(\beta_0, \beta_1, \beta_2) = (\cos(\theta) \sin(\phi), \sin(\theta) \sin(\phi), \cos(\phi)), \quad (5.2)$$

where $\theta \in [-180^\circ, 180^\circ]$ is the azimuth and $\phi \in [0, 180^\circ]$ is the inclination. It is perhaps convenient to think of the azimuth as related to longitude and the inclination as related to latitude. For non-trivial κ , the von Mises-Fisher distribution allocates probability mass to the sphere in such a way that its density contours are concentric about μ , with points closer to μ more likely to occur. The deterministic assignment rule of Kitagawa and Tetenov (2018) can be seen from Figure 3 to be located on the boundary between a high risk region (no-one treated) and a moderate risk region (everyone treated). As such, a stochastic assignment rule with non-trivial κ and μ located at this point would approximately allocate probability mass to each of these regions in equal amounts. By shifting μ towards the centre of the moderate risk region, we allocate relatively more mass to rules that induce moderate risk and less mass to rules that induce high risk, which reduces empirical welfare risk overall.

This pattern underlies what we observe in Figure 4, which plots empirical welfare risk for all μ on the sphere holding fixed κ at κ^* . Despite the apparent discontinuity of risk over deterministic assignment rules, empirical welfare risk (and the objective function) appear to vary smoothly.

6 Conclusion

Central to our analysis of the treatment choice problem is the question of *how should the social planner allocate individuals to treatment with a given probability rather than with certainty?* To answer this question, we focus on stochastic assignment rules that we formulate as posterior distributions obtained from well-defined prior distributions via a PAC-Bayes approach. These distributions are able to accommodate any initial belief that the social planner holds about what constitutes the best treatment, as well as any budgetary, ethical or legal constraints that are imposed and that can arise due to concerns about fairness or about maintaining the status quo. We establish that it is possible to obtain the minimum expected welfare risk under a variational approximation of the optimal posterior distribution, which we also characterise. In keeping with the notion that variational approximation replaces a general class of assignment rules with a simpler set of policies, we focus on stochastic assignment rules that can be expressed as density functions over the LES class of (deterministic) assignment rules. We exploit the scale invariance of the LES class to restrict attention to distributional families on the unit hypersphere, selecting the von Mises-Fisher family of distributions as our chosen variational family. We demonstrate how our methods can be used in an empirical setting by estimating which individuals should be entered onto a job training programme, using data from the well-known JTPA Study.

Our research suggests several further questions that remain unanswered. *How does the choice of variational family affect the rate of convergence?* The von Mises-Fisher family of distributions is not the only directional family – other families such as the Matrix Langevin and Kent families allow for richer covariance structures. Whilst more general distributions admit more complex stochastic assignment rules, does this additional generality come at an additional cost? *Do other parametrisations of the class of assignment rules exist that can achieve faster decay rates?* We establish that the optimal stochastic assignment rule in our

chosen variational family yields a convergence rate for welfare regret of $\ln(n)/\sqrt{n}$. Ignoring the logarithm in the numerator, this rate coincides with the minimax optimal rate for deterministic assignment rules in the absence of the margin assumption, as shown in Kitagawa and Tetenov (2018). Our convergence rate result, however, relies upon Assumption 5 that constrains the marginal distribution of X . We do not know whether this assumption leads to the minimax optimal rate being faster than $1/\sqrt{n}$ and whether the proposed method can attain a faster convergence rate than $\ln(n)/\sqrt{n}$.

A Proofs of theorems and associated lemmata

A.1 Proof of Theorem 1

We let $c : [0, 1]^2 \rightarrow \mathbb{R}$ denote a convex function. Then

$$\begin{aligned} n \cdot c(R_S^\Pi, R^\Pi) &= n \cdot c[\mathbb{E}_\Pi(R_S(g)), \mathbb{E}_\Pi(R(g))], \\ &\leq \mathbb{E}_\Pi[n \cdot c(R_S(g), R(g))], \end{aligned} \quad (\text{A.1})$$

by convexity.

Bégin et al. (2016, §Lemma 3) (Kullback-Leibler change of measure). *Let $\phi : \mathcal{G} \rightarrow \mathbb{R}$ be a measurable function. For any $g \in \mathcal{G}$ and any distributions Π and π_0 on \mathcal{G} such that Π is absolutely continuous with respect to π_0 ,*

$$\mathbb{E}_\Pi(\phi(g)) \leq \text{KL}(\Pi\|\pi_0) + \ln[\mathbb{E}_\Pi(\exp(\phi(g)))]. \quad (\text{A.2})$$

We apply the lemma above, with $\phi(g) = n \cdot c(R_S(g), R(g))$, to obtain

$$\mathbb{E}_\Pi[n \cdot c(R_S(g), R(g))] \leq \text{KL}(\Pi\|\pi_0) + \ln(\mathbb{E}_\Pi(\exp[n \cdot c(R_S(g), R(g))])). \quad (\text{A.3})$$

We then apply Markov's inequality to $\mathbb{E}_\Pi(\exp(n \cdot c(R_S(g), R(g))))$ to obtain

$$P^n \left(\mathbb{E}_\Pi(\exp(n \cdot c(R_S(g), R(g)))) \geq \frac{1}{\epsilon} \cdot \mathbb{E}_{P^n}(\mathbb{E}_\Pi(\exp[n \cdot c(R_S(g), R(g))])) \right) \leq \epsilon. \quad (\text{A.4})$$

Thus, with probability at least $1 - \epsilon$ we have, for all $\Pi \in \mathcal{M}$,

$$\mathbb{E}_\Pi(\exp[n \cdot c(R_S(g), R(g))]) \leq \frac{1}{\epsilon} \cdot \mathbb{E}_{P^n}(\mathbb{E}_\Pi(\exp[n \cdot c(R_S(g), R(g))])). \quad (\text{A.5})$$

Combining Equations (A.1), (A.3) and (A.5), we obtain

$$n \cdot c(R_S^\Pi, R^\Pi) \leq \text{KL}(\Pi\|\pi_0) + \ln \left[\frac{1}{\epsilon} \cdot \mathbb{E}_{P^n}(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))])) \right]. \quad (\text{A.6})$$

We now look to simplify $\mathbb{E}_{P^n}(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))]))$.

Maurer (2004, §Lemma 3). *Let U denote a vector of n independent and identically distributed random variables, each with values in the unit interval and $\mathbb{E}(U) = \mu$. In addition, let U^\dagger be a Bernoulli random variable satisfying $\mathbb{E}(U^\dagger) = \mu$. Suppose that $b : \{0, 1\}^n \rightarrow \mathbb{R}$ is a convex function. Then*

$$\mathbb{E}(b(U)) \leq \mathbb{E}(b(U^\dagger)). \quad (\text{A.7})$$

If t is permutation-symmetric in its arguments and $\theta(k)$ denotes the n -dimensional binary vector whose first k coordinates are one and whose remaining $n - k$ coordinates are zero, we also have

$$\mathbb{E}(b(U^\dagger)) = \sum_{k=0}^n \binom{n}{k} \cdot (1 - \mu)^{n-k} \cdot \mu^k \cdot b(\theta(k)). \quad (\text{A.8})$$

We then note that, by defining $b(U^\dagger) = \frac{1}{n} \sum_{i=1}^n U_i^\dagger$ such that $\mathbb{E}_{P^n}(b(U^\dagger)) = \mathbb{E}_{P^n}(R_S(g)) = R(g)$,

since $c(\cdot, R(g))$ is convex, and since the exponential function is convex and non-decreasing, $\exp[c(\cdot, R(g))]$ is convex. Applying Equation (A.7), we obtain

$$\mathbb{E}_{P^n} (\mathbb{E}_{\pi_0} (\exp [n \cdot c(R_S(g), R(g))])) \leq \mathbb{E}_{P^n} (\mathbb{E}_{\pi_0} (\exp [n \cdot c(b(U^\dagger), R(g))])), \quad (\text{A.9})$$

into which we substitute the definition of the expectation to write

$$\mathbb{E}_{P^n} (\mathbb{E}_{\pi_0} (\exp [n \cdot c(R_S(g), R(g))])) \leq \mathbb{E}_{\pi_0} \left(\sum_k P^n \left(b(U^\dagger) = \frac{k}{n} \right) \cdot \exp \left[n \cdot c \left(\frac{k}{n}, R(g) \right) \right] \right), \quad (\text{A.10})$$

where we omit the limits of summation to avoid confusion. We then apply Equation (A.8) and the binomial theorem to obtain

$$\mathbb{E}_{\pi_0} \left(\sum_k P^n \left(b(U^\dagger) = \frac{k}{n} \right) \cdot \exp \left[n \cdot c \left(\frac{k}{n}, R(g) \right) \right] \right) \quad (\text{A.11})$$

$$= \mathbb{E}_{\pi_0} \left(\sum_{k=0}^n \binom{n}{k} \cdot R(g)^k \cdot (1 - R(g))^{n-k} \cdot \exp \left[n \cdot c \left(\frac{k}{n}, R(g) \right) \right] \right), \quad (\text{A.12})$$

$$= \mathbb{E}_{\pi_0} \left(\sum_{k=0}^n \text{Bin}_k^n [R(g)] \cdot \exp \left(n \cdot c \left(\frac{k}{n}, R(g) \right) \right) \right), \quad (\text{A.13})$$

$$\leq \sup_{r \in [0, 1]} \sum_{k=0}^n \text{Bin}_k^n (r) \cdot \exp \left(n \cdot c \left(\frac{k}{n}, r \right) \right). \quad (\text{A.14})$$

If

$$\begin{aligned} c(q, p) &= \text{KL}(q \| p), \\ &= q \cdot \ln \left(\frac{q}{p} \right) + (1 - q) \cdot \ln \left(\frac{1 - q}{1 - p} \right), \end{aligned} \quad (\text{A.15})$$

then

$$\sup_{r \in [0, 1]} \sum_{k=0}^n \text{Bin}_k^n (r) \cdot \exp \left(n \cdot c \left(\frac{k}{n}, r \right) \right) = \sup_{r \in [0, 1]} \sum_{k=0}^n \binom{n}{k} \cdot \left(\frac{k}{n} \right)^k \cdot \left(1 - \frac{k}{n} \right)^{n-k}, \quad (\text{A.16})$$

$$= \sum_{k=0}^n \binom{n}{k} \cdot \left(\frac{k}{n} \right)^k \cdot \left(1 - \frac{k}{n} \right)^{n-k}, \quad (\text{A.17})$$

$$\equiv \mathcal{I}_{KL}, \quad (\text{A.18})$$

which is a function of n only.

Maurer (2004, §Theorem 1). *For all $n \geq 2$,*

$$\mathbb{E} (\exp [n \cdot \text{KL}(R_S(g) \| R(g))]) \leq \exp \left(\frac{1}{12n} \right) \cdot \sqrt{\frac{\pi n}{2}} + 2. \quad (\text{A.19})$$

We emphasise that, here, π signifies the mathematical constant rather than a probability distribution. As is discussed in Maurer (2004), for all $n \geq 8$,

$$\exp \left(\frac{1}{12n} \right) \cdot \sqrt{\frac{\pi n}{2}} + 2 \leq 2\sqrt{n}. \quad (\text{A.20})$$

Hence, for all $n \geq 8$ we have $\mathcal{I}_{KL} \leq 2\sqrt{n}$. We are of course interested in the maximum distance between the expected welfare risk and its empirical analogue and so define $c_{V^2}(q, p) = 2(q - p)^2$. Finally, we substitute $c_{V^2}(q, p)$ into Equation (A.6) and utilise Pinsker's Inequality, which states that $c_{V^2}(q, p) \leq$

$\text{KL}(q\|p)$, together with Maurer (2004, §Theorem 1) to obtain

$$R^\Pi - R_S^\Pi \leq \sqrt{\frac{1}{2n} \cdot \text{KL}(\Pi\|\pi_0) + \frac{1}{2n} \cdot \ln \left[\frac{1}{\epsilon} \cdot \mathbb{E}_{P^n} (\mathbb{E}_{\pi_0} (\exp [n \cdot c_{V^2} (R_S(g), R(g))])) \right]}, \quad (\text{A.21})$$

$$\leq \sqrt{\frac{1}{2n} \cdot \text{KL}(\Pi\|\pi_0) + \frac{1}{2n} \cdot \ln \left[\frac{1}{\epsilon} \cdot \mathbb{E}_{P^n} (\mathbb{E}_{\pi_0} (\exp [n \cdot \text{KL}(R_S(g)\|R(g))])) \right]}, \quad (\text{A.22})$$

$$\leq \sqrt{\frac{1}{2n} \cdot \text{KL}(\Pi\|\pi_0) + \frac{1}{2n} \cdot \ln \left(\frac{1}{\epsilon} \cdot \mathcal{J}_{KL} \right)}, \quad (\text{A.23})$$

$$\leq \sqrt{\frac{1}{2n} \left(\text{KL}(\Pi\|\pi_0) + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)}. \quad (\text{A.24})$$

The desired result follows by rearrangement.

A.2 Proof of Theorem 2

An optimal posterior minimises

$$\mathbb{E}_\Pi (R_S(g)) + \sqrt{\frac{1}{2n} \cdot \left(\mathbb{E}_\Pi \left(\ln \left(\frac{d\Pi(g)}{d\pi_0(g)} \right) \right) + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)} \text{ subject to } \int_{\mathcal{G}} d\Pi(g) = 1, \quad (\text{A.25})$$

$$\inf_{g \in \mathcal{G}} d\Pi(g) \geq 0.$$

Provided that the solution satisfies the non-negativity constraints (the second constraint of Equation (A.25)), this is equivalent to minimising

$$\int_{\mathcal{G}} R_S(g) \cdot d\Pi(g) + \sqrt{\frac{1}{2n} \cdot \left(\int_{\mathcal{G}} \ln \left(\frac{d\Pi(g)}{d\pi_0(g)} \right) \cdot d\Pi(g) + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)} + \xi \cdot \left(\int_{\mathcal{G}} d\Pi(g) - 1 \right), \quad (\text{A.26})$$

where ξ is the Lagrange multiplier. We separate this minimisation into two parts, by minimising Equation (A.26) over Π subject to its Kullback-Leibler divergence from π_0 being equal to $c \geq 0$ and, subsequently, by searching for the c that minimises the objective function.

We can write the constrained minimisation that forms the first part of the problem as

$$\int_{\mathcal{G}} R_S(g) \cdot d\Pi(g) + \sqrt{\frac{1}{2n} \cdot \left(c + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)} + \xi \cdot \left(\int_{\mathcal{G}} d\Pi(g) - 1 \right) + \frac{1}{\chi} \cdot \left(\int_{\mathcal{G}} \ln \left(\frac{d\Pi(g)}{d\pi_0(g)} \right) \cdot d\Pi(g) - c \right), \quad (\text{A.27})$$

where ξ and $1/\chi$ are Lagrange multipliers, and where $c \geq 0$ is a constant, provided that the omitted non-negativity constraints are satisfied at the solution. The associated first order condition of the minimand with respect to $d\Pi(g)$ is

$$R_S(g) + \xi + \frac{1}{\chi} \cdot \left(\ln \left(\frac{d\Pi(g)}{d\pi_0(g)} \right) + 1 \right) = 0. \quad (\text{A.28})$$

Rearranging, we obtain

$$d\Pi^*(g) = \frac{\exp(-\chi \cdot R_S(g))}{\exp(1 + \xi \cdot \chi)} \cdot d\pi_0(g), \quad (\text{A.29})$$

where we emphasise that Π^* is a function of χ . In view of the first constraint of Equation (A.26),

$$\exp(1 + \xi \cdot \chi) = \int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g), \quad (\text{A.30})$$

and so, for all $g \in \mathcal{G}$,

$$d\Pi^*(g) = \frac{\exp(-\chi \cdot R_S(g))}{\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g)} \cdot d\pi_0(g), \quad (\text{A.31})$$

which integrates to one as is required. We reiterate that Equation (A.31) is derived for an arbitrary $c \geq 0$, and so holds for any feasible values of χ .

For Π^* to satisfy $\text{KL}(\Pi^* \parallel \pi_0) = c$, we require that

$$\begin{aligned} c &= \int_{\mathcal{G}} \ln \left(\frac{d\Pi^*(g)}{d\pi_0(g)} \right) \cdot d\Pi^*(g), \\ &= - \int_{\mathcal{G}} \chi \cdot R_S(g) \cdot d\Pi^*(g) - \ln \left(\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g) \right). \end{aligned} \quad (\text{A.32})$$

This relationship between the radius of the Kullback-Leibler ball and the inverse of the Lagrange multiplier shows that c is strictly monotonically increasing in $\chi > 0$ provided that $R_S(g)$ is not constant over those g supported by π_0 , since

$$\frac{dc}{d\chi} = - \int_{\mathcal{G}} R_S(g) \cdot d\Pi^*(g) - \int_{\mathcal{G}} \chi \cdot R_S(g) \cdot \frac{d}{d\chi} \left(\frac{d\Pi^*(g)}{d\pi_0(g)} \right) \cdot d\pi_0 + \int_{\mathcal{G}} R_S(g) \cdot d\Pi^*(g) \quad (\text{A.33})$$

$$= \chi \cdot \int_{\mathcal{G}} \left(R_S(g) - R_S^{\Pi^*} \right)^2 \cdot d\Pi^*(g) \quad (\text{A.34})$$

$$\geq 0, \quad (\text{A.35})$$

where Equation (A.35) is strict if $\chi > 0$ and $R_S(g) \neq R_S^{\Pi^*}$ for at least some g supported by π_0 . Accordingly, in the second part of the optimisation, we substitute Π^* into Equation (A.27) in place of Π and solve the unconstrained minimisation problem with respect to χ . That is, we solve

$$\min_{\chi} \left\{ R_S^{\Pi^*} + \underbrace{\sqrt{\frac{1}{2n} \cdot \left(-\chi \cdot R_S^{\Pi^*} - \ln \left(\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g) \right) + \ln \left(\frac{2\sqrt{n}}{\epsilon} \right) \right)}}}_{\text{Penalty}} \right\}. \quad (\text{A.36})$$

We note that

$$\begin{aligned} \frac{d}{d\chi} R_S^{\Pi^*} &= \frac{d}{d\chi} \int_{\mathcal{G}} R_S(g) \cdot \frac{\exp(-\chi \cdot R_S(g))}{\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g)} \cdot d\pi_0(g), \\ &= \int_{\mathcal{G}} R_S(g) \cdot \frac{d}{d\chi} \frac{\exp(-\chi \cdot R_S(g))}{\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g)} \cdot d\pi_0(g), \\ &= - \int_{\mathcal{G}} \left(R_S(g) - R_S^{\Pi^*} \right)^2 \cdot d\Pi^*(g), \end{aligned} \quad (\text{A.37})$$

which we interpret as the variance of empirical welfare risk under Π^* . Using this result, we further note that

$$\begin{aligned} \frac{d}{d\chi} \text{Penalty} &= \frac{1}{4n \cdot \text{Penalty}} \cdot \left(-R_S^{\Pi^*} - \chi \cdot \frac{d}{d\chi} R_S^{\Pi^*} + \frac{\int_{\mathcal{G}} R_S(g) \cdot \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g)}{\int_{\mathcal{G}} \exp(-\chi \cdot R_S(g)) \cdot d\pi_0(g)} \right), \\ &= \frac{1}{4n \cdot \text{Penalty}} \cdot -\chi \cdot \frac{d}{d\chi} R_S^{\Pi^*}, \\ &= \frac{1}{4n \cdot \text{Penalty}} \cdot \chi \cdot \int_{\mathcal{G}} \left(R_S(g) - R_S^{\Pi^*} \right)^2 \cdot d\Pi^*(g). \end{aligned} \quad (\text{A.38})$$

Together Equations (A.37) and (A.38) imply that the associated first order condition of Equation (A.36) with respect to χ is

$$\frac{1}{4n \cdot \text{Penalty}} \cdot \chi \cdot \int_{\mathcal{G}} \left(R_S(g) - R_S^{\Pi^*} \right)^2 \cdot d\Pi^*(g) = \int_{\mathcal{G}} \left(R_S(g) - R_S^{\Pi^*} \right)^2 \cdot d\Pi^*(g). \quad (\text{A.39})$$

Provided – trivially, we might add – that Π^* is not degenerate and there is variation in empirical welfare risk (the conditions under which Equation (A.35) is strict), then this condition reduces to

$$\chi = 4n \cdot \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi^* \parallel \pi_0) + \ln\left(\frac{2\sqrt{n}}{\epsilon}\right) \right)}, \quad (\text{A.40})$$

which is exactly the condition that appears in Theorem 2.

Although Equations (A.31) and (A.40) characterise a possible interior solution of the optimisation, we have yet to guarantee that this proposed solution is a global optimum. To address this, we show that the two possible corner solutions are sub-optimal, such that the first order conditions from which Equation (A.40) are derived are applicable. Continuity and differentiability of Equation (A.27) are then sufficient¹³ to guarantee that a fixed point satisfying Equation (A.40) exists.

Neither π_0 nor a degenerate distribution are optimal. To establish that π_0 is not optimal, we note that Π^* coincides with π_0 when $\chi = 0$. By marginally increasing χ , such that we move in the direction of Π^* , we obtain a probability distribution that is in the interior. It suffices to show that such a probability distribution reduces the value of the objective function relative to π_0 . Evaluating Equations (A.37) and (A.38) at $\chi = 0$, we obtain

$$\begin{aligned} \frac{d}{d\chi} \left(R_S^{\Pi^*} + \text{Penalty} \right) \Big|_{\chi=0} &= \frac{d}{d\chi} R_S^{\pi_0}, \\ &< 0, \end{aligned} \quad (\text{A.41})$$

as we require, with any χ satisfying $0 < \chi < 2n \cdot R^{\pi_0}$ yielding a strictly lower value than $\chi = 0$. To establish that a degenerate distribution is not optimal, we note that such a distribution implies infinite divergence from π_0 (if π_0 were itself degenerate then our analysis would be meaningless since the prior and posterior distributions would always coincide; and if π_0 is atomic then the discrete case would apply). Given that π_0 attains a finite value of the objective function, however, π_0 is always preferred to a degenerate distribution, and so a degenerate distribution cannot be optimal.

A.3 Proof of Lemma 1

In the course of this proof, we rely on several results in Alquier et al. (2016) and follow the standard PAC-Bayesian approach (see Catoni, 2007). We define

$$d\Pi^*(g) = \frac{\exp[-\lambda \cdot (R(g) - R_S(g))]}{\int_{\mathcal{G}} \exp[-\lambda \cdot (R(g) - R_S(g))] \cdot d\pi_0(g)} \cdot d\pi_0(g), \quad (\text{A.42})$$

which differs from Equation (A.31) in its use of the notation λ rather than χ and in terms of the measures of risk that it uses, such that

$$\ln \left(\int_{\mathcal{G}} \exp[-\lambda \cdot (R(g) - R_S(g))] \cdot d\pi_0(g) \right) = - \inf_{\Pi \in \mathcal{M}} \left\{ \int_{\mathcal{G}} \lambda \cdot (R(g) - R_S(g)) \cdot d\Pi(g) + \text{KL}(\Pi \parallel \pi_0) \right\}, \quad (\text{A.43})$$

by definition of the Kullback-Leibler divergence and of Π^* .

As in Alquier et al. (2016), we define a Hoeffding assumption as

$$E_{\pi_0}(E_{P^n}(\exp[\lambda \cdot (R(g) - R_S(g))])) \leq \exp(a(\lambda, n)), \quad (\text{A.44})$$

where a is a function that depends upon $\lambda > 0$. This follows directly from applying Hoeffding's lemma (Boucheron et al., 2013) to the left-hand side of Equation (A.44) and then taking the expectation over \mathcal{G} with respect to π_0 . As we have bounded the welfare criterion such that it is contained in the unit interval, this assumption holds directly without any further restrictions. Applying Fubini's Theorem to Equation (A.44),

¹³It is possible to show that the difference between the left- and right-hand sides of Equation (A.40) is negative at $\chi = 0$ and positive if χ is sufficiently large, with existence then established using extensions of the intermediate value theorem and its corollary, Bolzano's theorem.

we obtain

$$\mathbb{E}_{P^n} \left(\int_{\mathcal{G}} \exp(\lambda \cdot (R_S(g) - R(g)) - a(\lambda, n)) \cdot d\pi_0 \right) \leq 1. \quad (\text{A.45})$$

Given that (A.43) can be re-stated as

$$\int_{\mathcal{G}} \exp[-\lambda \cdot (R(g) - R_S(g))] \cdot d\pi_0(g) = \exp \left(\sup_{\Pi \in \mathcal{M}} \int_{\mathcal{G}} \lambda \cdot (R_S(g) - R(g)) \cdot d\Pi(g) - \text{KL}(\Pi \| \pi_0) \right), \quad (\text{A.46})$$

it follows from Equations (A.43) and (A.45) that

$$\mathbb{E}_{P^n} \left[\exp \left(\sup_{\Pi \in \mathcal{M}} \lambda \cdot \int_{\mathcal{G}} (R_S(g) - R(g)) \cdot d\Pi(g) - \text{KL}(\Pi \| \pi_0) - a(\lambda, n) \right) \right] \leq 1. \quad (\text{A.47})$$

Multiplying both sides of Equation (A.47) by $\epsilon/2$ and using the fact that $\mathbb{E}(\exp(U)) \geq \Pr(U > 0)$ for any random variable U , we obtain

$$P^n \left(\sup_{\Pi \in \mathcal{M}} \lambda \cdot \int_{\mathcal{G}} (R_S(g) - R(g)) \cdot d\Pi(g) - \text{KL}(\Pi \| \pi_0) - a(\lambda, n) + \ln \left(\frac{\epsilon}{2} \right) > 0 \right) \leq \frac{\epsilon}{2}. \quad (\text{A.48})$$

Keeping this bound in mind, we now turn to the problem of bounding expected welfare regret.

We let c denote a convex function that maps from the unit square to the reals. Then, by Jensen's Inequality,

$$\begin{aligned} n \cdot c(R_S^\Pi, R^\Pi) &= n \cdot c[\mathbb{E}_\Pi(R_S(g)), \mathbb{E}_\Pi(R(g))], \\ &\leq \mathbb{E}_\Pi[n \cdot c(R_S(g), R(g))]. \end{aligned} \quad (\text{A.49})$$

Letting $\phi(g) = n \cdot c(R_S(g), R(g))$, we apply Bégin et al. (2016, §Lemma 3) to obtain

$$\mathbb{E}_\Pi[n \cdot c(R_S(g), R(g))] \leq \text{KL}(\Pi \| \pi_0) + \ln(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))])), \quad (\text{A.50})$$

which, in conjunction with Equation (A.49), yields

$$n \cdot c(R_S^\Pi, R^\Pi) \leq \text{KL}(\Pi \| \pi_0) + \ln(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))])). \quad (\text{A.51})$$

We then apply Markov's inequality to the term inside the logarithm, which yields

$$P^n \left[\mathbb{E}_{\pi_0}(\exp(n \cdot c[R_S(g), R(g)])) \geq \frac{2}{\epsilon} \cdot \mathbb{E}_{P^n}(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))])) \right] \leq \frac{\epsilon}{2}. \quad (\text{A.52})$$

Combining Equations (A.48) and (A.52) by means of Bonferroni's inequalities, we determine that, for all $\Pi \in \mathcal{M}$,

$$\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))]) \leq \frac{2}{\epsilon} \cdot \mathbb{E}_{P^n}(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))])), \quad (\text{A.53})$$

and

$$\lambda \cdot \mathbb{E}_\Pi(R_S(g)) \leq \lambda \cdot \mathbb{E}_\Pi(R(g)) + \text{KL}(\Pi \| \pi_0) + a(\lambda, n) + \ln \left(\frac{2}{\epsilon} \right), \quad (\text{A.54})$$

which both hold simultaneously with probability at least $1 - \epsilon$. We then take the exponential of both sides of Equation (A.51) and rearrange, relating the result to the left-hand side of Equation (A.53) to obtain

$$n \cdot c(R_S^\Pi, R^\Pi) \leq \text{KL}(\Pi \| \pi_0) + \ln \left[\frac{2}{\epsilon} \cdot \mathbb{E}_{P^n}(\mathbb{E}_{\pi_0}(\exp[n \cdot c(R_S(g), R(g))])) \right]. \quad (\text{A.55})$$

We note that Equation (A.55) is identical to Equation (A.6) except for the logarithm on the right hand-side, the argument of which differs only in the constant that multiplies the expectation. As such, and following

the steps of Appendix A.1 that follow Equation (A.6), we are able to show that, for all $\Pi \in \mathcal{M}$,

$$R^\Pi \leq R_S^\Pi + \sqrt{\frac{1}{2n} \cdot \left(\text{KL}(\Pi || \pi_0) + \ln\left(\frac{4\sqrt{n}}{\epsilon}\right) \right)}, \quad (\text{A.56})$$

and

$$\lambda R_S^\Pi \leq \lambda \cdot R^\Pi + \text{KL}(\Pi || \pi_0) + a(\lambda, n) + \ln\left(\frac{2}{\epsilon}\right), \quad (\text{A.57})$$

which both hold simultaneously with probability at least $1 - \epsilon$. We now denote by $\mathcal{V} \subset \mathcal{M}$ a family of distributions over the class of feasible assignment rules and restrict attention to this subset. Since $\mathcal{V} \subset \mathcal{M}$, Equations (A.56) and (A.57) hold with probability at least $1 - \epsilon$ for all $\Pi \in \mathcal{V}$ and, in particular, for the variational approximation of Π^* , which we define as

$$\tilde{\Pi} = \underset{\Pi \in \mathcal{G}}{\operatorname{argmin}} \text{KL}(\Pi || \Pi^*). \quad (\text{A.58})$$

Hence, from Equation (A.56),

$$R^{\tilde{\Pi}} \leq \inf_{\Pi \in \mathcal{V}} R_S^\Pi + \sqrt{\frac{1}{2n} \left(\text{KL}(\Pi || \pi_0) + \ln\left(\frac{4\sqrt{n}}{\epsilon}\right) \right)}. \quad (\text{A.59})$$

Finally, dividing both sides of Equation (A.57) by λ and replacing R_S^Π inside the infimum of Equation (A.59) with the right-hand side of Equation (A.57), we obtain the required result, which we reiterate holds with probability at least $1 - \epsilon$.

A.4 Proof of Theorem 3

We rely on an intermediate result that we introduce now, in advance of the proof.

Lemma 2. *The circular variance of a von Mises-Fisher random vector with concentration parameter $\kappa \geq 0$ is bounded from above by*

$$\frac{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2} - \kappa}{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}, \quad (\text{A.60})$$

where $\nu = m/2 - 1$ for convenience, with $\underline{c}_\nu = \nu + 1/2$ and $\bar{c}_\nu = \nu + 3/2$.

As is shown in Kitagawa and Rowley (2022), the circular variance is an $O(\kappa^{-1})$ function. Proof of Lemma 2 is deferred until the end of this proof.

Lemma 3. *The Kullback-Leibler divergence of a von Mises-Fisher random vector with concentration parameter $\kappa \geq 0$ is bounded from above by*

$$\underline{c}_\nu \cdot \ln\left(\frac{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}{\underline{c}_\nu + \bar{c}_\nu}\right) + \sqrt{\kappa^2 + \underline{c}_\nu^2} - \sqrt{\kappa^2 + \bar{c}_\nu^2} + 1, \quad (\text{A.61})$$

where $\nu = m/2 - 1$ for convenience, with $\underline{c}_\nu = \nu + 1/2$ and $\bar{c}_\nu = \nu + 3/2$.

As is shown in Kitagawa and Rowley (2022), the Kullback-Leibler divergence is an $O(\ln(\kappa))$ function. Proof of Lemma 3 is deferred until the end of this proof.

With Lemmata 2 and 3 to hand, we begin by noting that, together, Assumption 5 and Lemma 1 imply that

$$R^{\tilde{\Pi}} \leq \bar{R} + \inf_{\mu, \kappa} a_1 + a_2 + \sqrt{a_3}, \quad (\text{A.62})$$

where

$$\begin{aligned} a_1 &= \int_{\mathbb{S}^{m-1}} 2c \cdot \|\beta - \bar{\beta}\|_2 \cdot d\Pi(\beta; \kappa, \mu), \\ a_2 &= \frac{1}{\lambda} \cdot \left(a(\lambda, n) + \text{KL}(\Pi || \pi_0) + \ln\left(\frac{2}{\epsilon}\right) \right), \\ a_3 &= \frac{1}{2n} \cdot \left(\text{KL}(\Pi || \pi_0) + \ln\left(\frac{4\sqrt{n}}{\epsilon}\right) \right), \end{aligned} \quad (\text{A.63})$$

and where \mathbb{S}^m is the m -sphere (i.e., the hypersphere in \mathbb{R}^{m+1}).

Focusing on a_1 , we use Jensen's inequality to show that

$$2c \cdot \int_{\mathbb{S}^{m-1}} \|\beta - \bar{\beta}\|_2 \cdot d\Pi(\beta; \kappa, \mu) \leq 2c \cdot \sqrt{\int_{\mathbb{S}^{m-1}} \sum_{l=0}^m (\beta_l - \bar{\beta}_l)^2 \cdot d\Pi(\beta; \kappa, \mu)} \quad (\text{A.64})$$

$$= 2c \cdot \sqrt{\text{tr}(\mathbf{E}_{\Pi}((\beta - \bar{\beta}) \cdot (\beta - \bar{\beta})^\top))}. \quad (\text{A.65})$$

Rearranging, we obtain

$$\mathbf{E}_{\Pi}((\beta - \bar{\beta}) \cdot (\beta - \bar{\beta})^\top) = \mathbf{E}_{\Pi}((\beta - \bar{\beta} + \mathbf{E}_{\Pi}(\beta) - \mathbf{E}_{\Pi}(\beta)) \cdot (\beta - \bar{\beta} + \mathbf{E}_{\Pi}(\beta) - \mathbf{E}_{\Pi}(\beta))^\top) \quad (\text{A.66})$$

$$= \mathbf{E}_{\Pi}((\beta - \mathbf{E}_{\Pi}(\beta)) \cdot (\beta - \mathbf{E}_{\Pi}(\beta))^\top) + (\mathbf{E}_{\Pi}(\beta) - \bar{\beta}) \cdot (\mathbf{E}_{\Pi}(\beta) - \bar{\beta})^\top \quad (\text{A.67})$$

$$= \text{Variance}_{\Pi}(\beta) + (A_m(\kappa) \cdot \mu - \bar{\beta}) \cdot (A_m(\kappa) \cdot \mu - \bar{\beta})^\top, \quad (\text{A.68})$$

where

$$A_m(\kappa) = \frac{I_{m/2}(\kappa)}{I_{m/2-1}(\kappa)}, \quad (\text{A.69})$$

which relies on two results from Mardia and Jupp (2009) and Kitagawa and Rowley (2022). These results relate to the first two moments of the von Mises-Fisher distribution and are

$$\mathbf{E}_{\Pi}(\beta) = A_m(\kappa) \cdot \mu, \quad (\text{A.70})$$

$$\text{Variance}_{\Pi}(\beta) = \frac{1}{\kappa} \cdot A_m(\kappa) \cdot \mathbb{I}_m + \left(1 - \frac{m}{\kappa} \cdot A_m(\kappa) - A_m^2(\kappa)\right) \cdot \mu \cdot \mu^\top, \quad (\text{A.71})$$

where \mathbb{I}_m is the m -dimensional identity matrix. Restricting the set of von Mises-Fisher distributions to those satisfying $\mu = \bar{\beta}$, a_1 is bounded by

$$\begin{aligned} 2c \cdot \sqrt{\text{tr}(\mathbf{E}((\beta - \bar{\beta}) \cdot (\beta - \bar{\beta})^\top))} &= 2c \cdot \sqrt{\text{tr}(\text{Variance}(\beta) + (A_m(\kappa) - 1)^2 \cdot \mu \cdot \mu^\top)} \\ &= \sqrt{8c^2} \cdot \sqrt{1 - A_m(\kappa)}, \end{aligned} \quad (\text{A.72})$$

which is proportional to the square root of the circular variance.¹⁴ Lemma 2 details the behaviour of the circular variance. Using Lemma 2, Equation (A.72) can be bounded from above by

$$\sqrt{8c^2 \cdot \frac{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2} - \kappa}{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}} \leq 4c \cdot \sqrt{\frac{\nu + 1}{\kappa}}, \quad (\text{A.73})$$

where the right-hand side follows from the definitions of \underline{c}_ν and \bar{c}_ν and concavity of the square root, and by decreasing the denominator via elimination of \underline{c}_ν and \bar{c}_ν (both are non-negative constants). We substitute the right-hand side of Equation (A.73) into Equation (A.62) in place of a_1 .

¹⁴Recall that the right-hand side of Equation (A.62) is preceded by an infimum over μ ; it is not necessary for $\mu = \bar{\beta}$ to minimise Equation (A.72) for this choice to deliver a useful upper bound.

We now focus on the Kullback-Leibler divergence, which appears in both a_2 and a_3 . We observe that the difference between the square roots in Equation (A.61) is increasing in κ (i.e., the difference becomes less negative as κ increases and as κ increases in importance relative to ν). As such, it suffices to omit the difference between the square roots in Equation (A.61) and to simply bound the Kullback-Leibler divergence from above by

$$\underline{c}_\nu \cdot \ln \left(\frac{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}{\underline{c}_\nu + \bar{c}_\nu} \right) + 1 \leq \underline{c}_\nu \cdot \ln(\kappa + 1) + 1, \quad (\text{A.74})$$

which is at least one due to the non-negativity of κ .

Substituting Equations (A.73) and (A.74) into the infimand of Equation (A.62), we obtain

$$4c \cdot \sqrt{\frac{\nu + 1}{\kappa}} + \frac{a(\lambda, n)}{\lambda} + \frac{\underline{c}_\nu}{\lambda} \cdot \ln(\kappa + 1) + \frac{1}{\lambda} \cdot \ln \left(\frac{2e}{\epsilon} \right) + \sqrt{\frac{1}{2n} \cdot \left(\underline{c}_\nu \cdot \ln(\kappa + 1) + \ln \left(\frac{4e \cdot \sqrt{n}}{\epsilon} \right) \right)}, \quad (\text{A.75})$$

which we emphasise is an upper bound on the infimum. Our objective is to minimise Equation (A.75) by appropriately choosing κ and λ alongside the functional form of a . Accordingly, we let $\kappa = n$ and $\lambda = \sqrt{n}$ alongside $a(\lambda, n)/\lambda = 1/\lambda$. Given these choices, it then follows that Equation (A.75) is itself bounded from above by

$$4c \cdot \sqrt{\frac{\nu + 1}{n}} + \frac{1}{\sqrt{n}} + \frac{\underline{c}_\nu}{\sqrt{n}} \cdot \ln(n + 1) + \frac{1}{\sqrt{n}} \cdot \ln \left(\frac{2e}{\epsilon} \right) + \sqrt{\frac{\underline{c}_\nu \cdot \ln(n + 1)}{2n}} + \sqrt{\frac{1}{2n} \ln \left(\frac{4e \cdot \sqrt{n}}{\epsilon} \right)} \quad (\text{A.76})$$

which exploits the concavity of the square root. Defining

$$M \equiv 4c \cdot \sqrt{\nu + 1} + 1 + \underline{c}_\nu \cdot \frac{\ln(9)}{\ln(8)} + \ln \left(\frac{2e}{\epsilon} \right) + \sqrt{\frac{\underline{c}_\nu \cdot \ln(9)}{2 \cdot \ln(8)}} + \sqrt{\frac{1}{2} \cdot \ln \left(\frac{4e}{\epsilon} \right)} + \frac{\sqrt{\ln(8)}}{\sqrt{2} \cdot \ln(8)}, \quad (\text{A.77})$$

which relies on the maintained assumption of Theorem 3 that $n \geq 8$, we then determine that

$$R^{\tilde{\Pi}} \leq \bar{R} + M \cdot \frac{\ln(n)}{\sqrt{n}}, \quad (\text{A.78})$$

such that M is a universal constant.

A.5 Proof of Lemma 2

The circular variance of a von Mises-Fisher random vector is one minus the ratio of consecutive modified Bessel functions of the first kind (i.e., one minus the mean resultant length). Bounds on these ratios are derived in Amos (1974; restated for immediate application to hyperspherical problems in Kitagawa and Rowley, 2022). The result then immediately follows upon subtraction of the lower bound from one.

A.6 Proof of Lemma 3

Kitagawa and Rowley (2022) shows that, when Π is a von Mises-Fisher distribution and π_0 is the uniform distribution over the hypersphere,

$$\text{KL}(\Pi \| \pi_0) = \nu \cdot \ln \left(\frac{\kappa}{2} \right) - \underbrace{\ln(I_\nu(\kappa))}_{\text{Bessel fn.}} - \ln(\Gamma(\nu + 1)) + \underbrace{A_m(\kappa) \cdot \kappa}_{\text{Ratio fn.}}, \quad (\text{A.79})$$

where we recall that $\nu = m/2 - 1$, with $\underline{c}_\nu = \nu + 1/2$ and $\bar{c}_\nu = \nu + 3/2$. To derive an upper bound on the Kullback-Leibler divergence that does not involve modified Bessel functions or their ratios, we replace the terms labelled *Bessel fn.* and *Ratio fn.* in Equation (A.79) with appropriate lower and upper bounds, respectively. To do so, we rely on results (and notation) in Kitagawa and Rowley (2022), which are themselves adapted and restated from Amos (1974).

First, the term labelled *Bessel fn.* is bounded from below (recall that the term enters negatively) by

$$\frac{1}{2} \cdot \ln\left(\frac{2}{\kappa}\right) - \ln(\Gamma(\nu + 1)) + \underline{c}_\nu \cdot \ln\left(\frac{\kappa \cdot (\nu + 1)}{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}\right) + \frac{\kappa^2}{\bar{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}. \quad (\text{A.80})$$

Second, the term labelled *Ratio fn.* is bounded from above by

$$\frac{\kappa^2}{\underline{c}_\nu + \sqrt{\kappa^2 + \underline{c}_\nu^2}}. \quad (\text{A.81})$$

Substituting Equations (A.80) and (A.81) into Equation (A.79), cancelling terms and noting that

$$\nu \cdot \ln\left(\frac{\kappa}{2}\right) - \frac{1}{2} \cdot \ln\left(\frac{2}{\kappa}\right) = \left(\nu + \frac{1}{2}\right) \cdot \ln\left(\frac{\kappa}{2}\right), \quad (\text{A.82})$$

we obtain

$$\underline{c}_\nu \cdot \ln\left(\frac{\kappa}{2}\right) - \underline{c}_\nu \cdot \ln\left(\frac{\kappa \cdot (\nu + 1)}{\underline{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}}\right) - \frac{\kappa^2}{\bar{c}_\nu + \sqrt{\kappa^2 + \bar{c}_\nu^2}} + \frac{\kappa^2}{\underline{c}_\nu + \sqrt{\kappa^2 + \underline{c}_\nu^2}}. \quad (\text{A.83})$$

To obtain the required result, we simplify the first two terms of Equation (A.83) and use the fact that, for any real c (in this case, equal to either \underline{c}_ν or \bar{c}_ν , which are positive reals),

$$\frac{\kappa^2}{c + \sqrt{\kappa^2 + c^2}} = \sqrt{\kappa^2 + c^2} - c, \quad (\text{A.84})$$

which makes use of the formula for the difference of two squares.

References

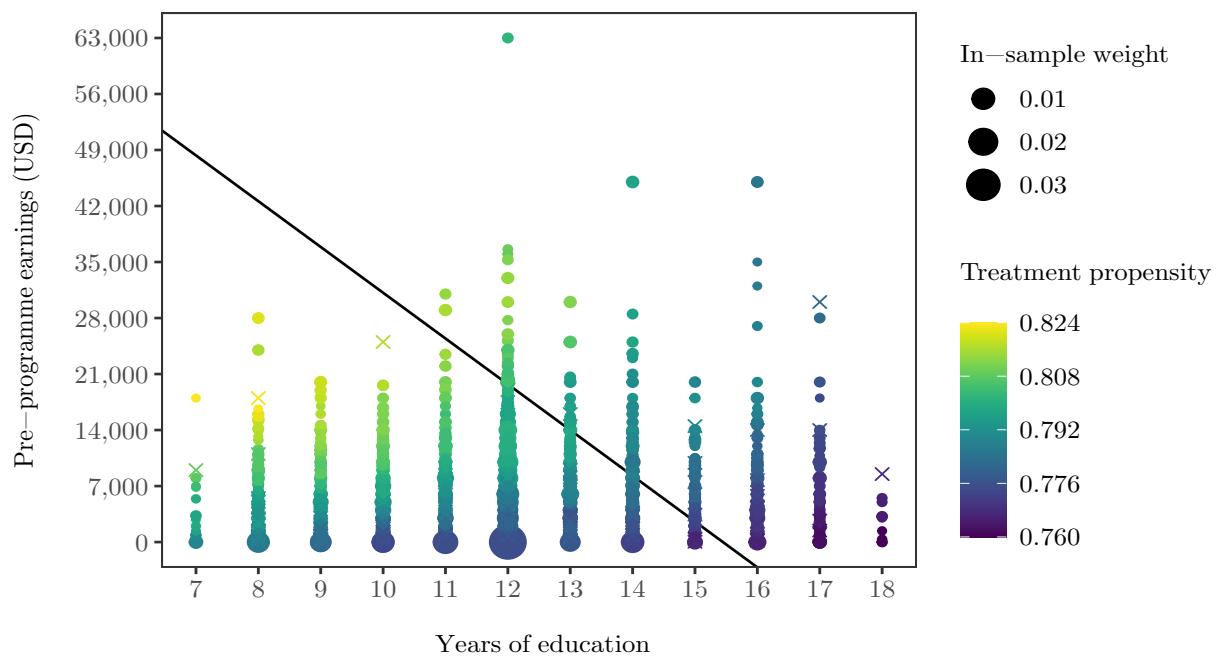
- ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): “Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings,” *Econometrica*, 70, 91–117.
- ADJAHO, C. AND T. CHRISTENSEN (2022): “Externally Valid Treatment Choice,” *arXiv preprint arXiv:2205.05561*.
- ALQUIER, P., J. RIDGWAY, AND N. CHOPIN (2016): “On the properties of variational approximations of Gibbs posteriors,” *The Journal of Machine Learning Research*, 17, 8374–8414.
- AMOS, D. E. (1974): “Computation of modified Bessel functions and their ratios,” *Mathematics of computation*, 28, 239–251.
- ATHEY, S. AND S. WAGER (2021): “Policy learning with observational data,” *Econometrica*, 89, 133–161.
- BÉGIN, L., P. GERMAIN, F. LAVIOLETTE, AND J.-F. ROY (2014): “PAC-Bayesian theory for transductive learning,” in *Artificial Intelligence and Statistics*, 105–113.
- (2016): “PAC-Bayesian bounds based on the Rényi divergence,” in *Artificial Intelligence and Statistics*, 435–444.
- BEYGELZIMER, A. AND J. LANGFORD (2009): “The offset tree for learning with partial labels,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 129–138.
- BHATTACHARYA, D. AND P. DUPAS (2012): “Inferring welfare maximizing treatment assignment under budget constraints,” *Journal of Econometrics*, 167, 168–196.
- BISSIRI, P. G., C. C. HOLMES, AND S. G. WALKER (2016): “A general framework for updating belief distributions,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 1103–1130.
- BLOOM, H. S., L. L. ORR, S. H. BELL, G. CAVE, F. DOOLITTLE, W. LIN, AND J. M. BOS (1997): “The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study,” *Journal of human resources*, 549–576.

- BOUCHERON, S., G. LUGOSI, AND P. MASSART (2013): *Concentration inequalities: A nonasymptotic theory of independence*, Oxford university press.
- CATONI, O. (2007): “PAC-Bayesian supervised classification: the thermodynamics of statistical learning,” *arXiv preprint arXiv:0712.0248*.
- CHAMBERLAIN, G. (2011): “Bayesian aspects of treatment choice,” *The Oxford Handbook of Bayesian Econometrics*, 11–39.
- CHERNOZHUKOV, V. AND H. HONG (2003): “An MCMC approach to classical estimation,” *Journal of Econometrics*, 115, 293–346.
- CSABA, D. AND B. SZOKE (2020): “Learning with misspecified models,” Tech. rep., mimeo.
- D’ADAMO, R. (2021): “Policy Learning Under Ambiguity,” *arXiv preprint arXiv:2111.10904*.
- DEHEJIA, R. H. (2005): “Program evaluation as a decision problem,” *Journal of Econometrics*, 125, 141–173.
- DERBEKO, P., R. EL-YANIV, AND R. MEIR (2004): “Explicit learning curves for transduction and application to clustering and compression algorithms,” *Journal of Artificial Intelligence Research*, 22, 117–142.
- DWORK, C., M. HARDT, T. PITASSI, O. REINGOLD, AND R. ZEMEL (2012): “Fairness through awareness,” in *Proceedings of the 3rd innovations in theoretical computer science conference*, 214–226.
- GERMAIN, P., A. LACASSE, F. LAVIOLETTE, AND M. MARCHAND (2009): “PAC-Bayesian learning of linear classifiers,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, 353–360.
- GUEDJ, B. (2019): “A primer on PAC-Bayesian learning,” *arXiv preprint arXiv:1901.05353*.
- GUNSILIUS, F. (2019): “A path-sampling method to partially identify causal effects in instrumental variable models,” *arXiv preprint arXiv:1910.09502*.
- HAN, S. (2022): “Optimal dynamic treatment regimes and partial welfare ordering,” *unpublished manuscript*.
- HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): “Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme,” *The review of economic studies*, 64, 605–654.
- HIRANO, K. AND J. R. PORTER (2009): “Asymptotics for statistical treatment rules,” *Econometrica*, 77, 1683–1701.
- ISHIHARA, T. AND T. KITAGAWA (2021): “Evidence Aggregation for Treatment Choice,” *arXiv preprint arXiv:2108.06473*.
- KARLIN, S. AND H. RUBIN (1956): “The theory of decision procedures for distributions with monotone likelihood ratio,” *The Annals of Mathematical Statistics*, 272–299.
- KASY, M. (2016): “Partial identification, distributional preferences, and the welfare ranking of policies,” *Review of Economics and Statistics*, 98, 111–131.
- KIDO, D. (2022): “Distributionally Robust Policy Learning with Wasserstein Distance,” *arXiv preprint arXiv:2205.04637*.
- KITAGAWA, T., S. LEE, AND C. QIU (2022): “Treatment Choice with Nonlinear Regret,” *arXiv preprint arXiv:2205.08586*.
- KITAGAWA, T. AND J. ROWLEY (2022): “Statistical divergences of von Mises-Fisher distributions,” *arXiv preprint arXiv:2202.05192*.
- KITAGAWA, T., S. SAKAGUCHI, AND A. TETENOV (2021): “Constrained classification and policy learning,” *arXiv preprint arXiv:2106.12886*.
- KITAGAWA, T. AND A. TETENOV (2018): “Who should be treated? empirical welfare maximization methods for treatment choice,” *Econometrica*, 86, 591–616.
- (2021): “Equality-Minded Treatment Choice,” *Journal of Business Economics and Statistics*, 39, 561–574.
- KOCK, A. B., D. PREINERSTORFER, AND B. VELYEV (2022): “Treatment recommendation with distributional targets,” *Journal of Econometrics*.
- KURZ, G. AND U. D. HANEBECK (2015): “Stochastic sampling of the hyperspherical von Mises–Fisher distribution without rejection methods,” in *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, IEEE, 1–6.
- LIU, Y. (2022): “Policy Learning under Endogeneity Using Instrumental Variables,” *arXiv preprint arXiv:2206.09883*.
- MANSKI, C. F. (2000a): “Identification problems and decisions under ambiguity: empirical analysis of treatment response and normative analysis of treatment choice,” *Journal of Econometrics*, 95, 415–442.

- (2000b): “Using Studies of Treatment Response to Inform Treatment Choice in Heterogeneous Populations,” .
- (2004): “Statistical treatment rules for heterogeneous populations,” *Econometrica*, 72, 1221–1246.
- (2005): *Social choice with partial knowledge of treatment response*, Princeton University Press.
- (2007a): *Identification for prediction and decision*, Harvard University Press.
- (2007b): “Minimax-regret treatment choice with missing outcome data,” *Journal of Econometrics*, 139, 105–115.
- (2009): “The 2009 Lawrence R. Klein Lecture: Diversified treatment under ambiguity,” *International Economic Review*, 50, 1013–1041.
- (2022): “Identification and Statistical Decision Theory,” *arXiv preprint arXiv:2204.11318*.
- MANSKI, C. F. AND A. TETENOV (2007): “Admissible treatment rules for a risk-averse planner with experimental data on an innovation,” *Journal of Statistical Planning and Inference*, 137, 1998–2010.
- MARDIA, K. V. AND P. E. JUPP (2009): *Directional statistics*, vol. 494, John Wiley & Sons.
- MAURER, A. (2004): “A note on the PAC Bayesian theorem,” *arXiv preprint cs/0411099*.
- MBAKOP, E. AND M. TABORD-MEEHAN (2021): “Model selection for treatment choice: Penalized welfare maximization,” *Econometrica*, 89, 825–848.
- MCALLESTER, D. A. (1999): “Some PAC-Bayesian Theorems,” *Machine Learning*, 37, 355–363.
- (2003): “PAC-Bayesian stochastic model selection,” *Machine Learning*, 51, 5–21.
- MOLCHANOV, I. (2005): *Theory of Random Sets*, Springer verlag.
- NIE, X., E. BRUNSKILL, AND S. WAGER (2021): “Learning When-to-Treat Policies,” *Journal of the American Statistical Association*, 116, 392–409.
- PELLATT, D. F. (2022): “PAC-Bayesian Treatment Allocation Under Budget Constraints,” *arXiv preprint arXiv:2212.09007*.
- PENTINA, A. AND C. H. LAMPERT (2015): “Lifelong learning with non-iid tasks,” in *Advances in Neural Information Processing Systems*, 1540–1548.
- RAI, Y. (2019): “Statistical Inference for Treatment Assignment Policies,” *unpublished manuscript*.
- ROSENBAUM, P. R. AND D. B. RUBIN (1983): “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, 70, 41–55.
- SAKAGUCHI, S. (2019): “Estimating Optimal Dynamic Treatment Assignment Rules under Intertemporal Budget Constraints,” Tech. rep., cemmap working paper.
- SASAKI, Y. AND T. URA (2020): “Welfare analysis via marginal treatment effects,” *arXiv preprint arXiv:2012.07624*.
- SCHLAG, K. H. (2006): “ELEVEN-Tests needed for a Recommendation,” *Economics Working Papers*, ECO2006/2.
- SEEGER, M. (2002): “Pac-bayesian generalisation error bounds for gaussian process classification,” *Journal of machine learning research*, 3, 233–269.
- SHAWE-TAYLOR, J. AND R. C. WILLIAMSON (1997): “A PAC analysis of a Bayesian estimator,” in *Annual Workshop on Computational Learning Theory: Proceedings of the tenth annual conference on Computational learning theory*, vol. 6, 2–9.
- STOYE, J. (2009): “Minimax regret treatment choice with finite samples,” *Journal of Econometrics*, 151, 70–81.
- (2012): “Minimax regret treatment choice with covariates or with limited validity of experiments,” *Journal of Econometrics*, 166, 138–156.
- SUN, L. (2021): “Empirical welfare maximization with constraints,” *arXiv preprint*.
- SWAMINATHAN, A. AND T. JOACHIMS (2015): “Counterfactual risk minimization: Learning from logged bandit feedback,” in *International Conference on Machine Learning*, 814–823.
- TETENOV, A. (2012): “Statistical treatment choice based on asymmetric minimax regret criteria,” *Journal of Econometrics*, 166, 157–165.
- THOMPSON, W. R. (1933): “On the likelihood that one unknown probability exceeds another in view of the evidence of two samples,” *Biometrika*, 25, 285–294.
- VALIANT, L. G. (1984): “A theory of the learnable,” in *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, ACM, 436–445.
- VAPNIK, V. N. (1998): *Statistical Learning Theory*, Wiley.
- VIVIANO, D. (2021): “Policy targeting under network interference,” *arXiv preprint*.

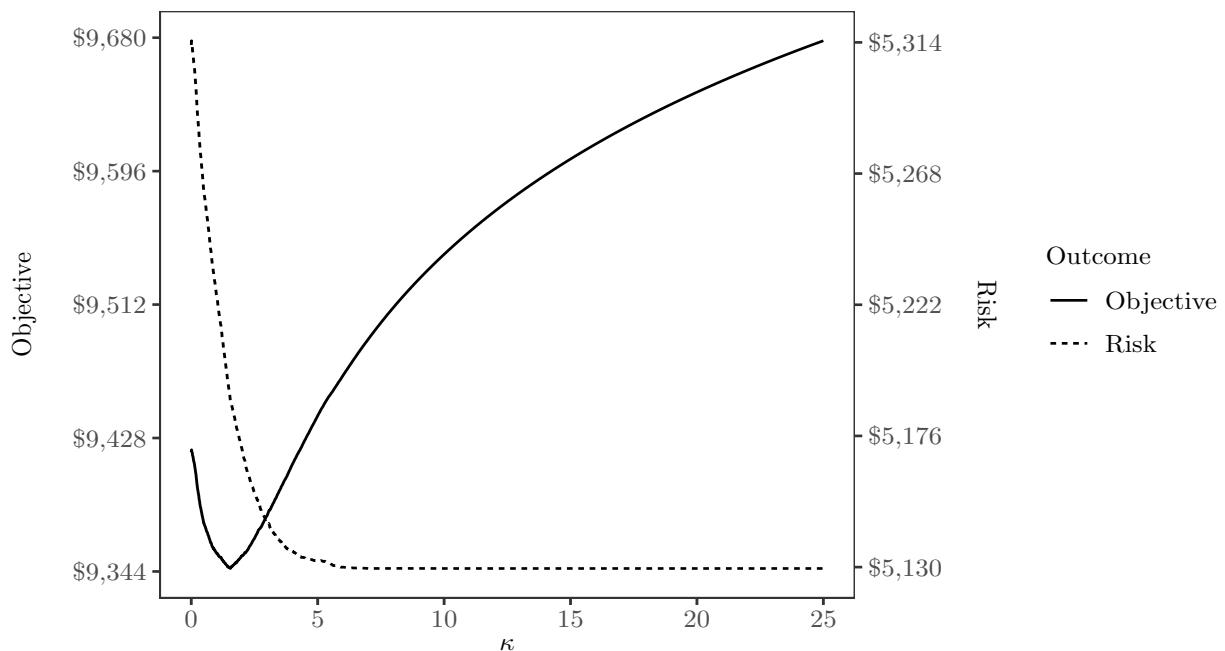
- WAINWRIGHT, M. J. AND M. I. JORDAN (2008): “Graphical Models, Exponential Families, and Variational Inference,” *Foundations and Trends in Machine Learning*, 1, 1 – 305.
- WOOD, A. T. (1994): “Simulation of the von Mises Fisher distribution,” *Communications in statistics-simulation and computation*, 23, 157–164.
- YATA, K. (2021): “Optimal Decision Rules Under Partial Identification,” *arXiv preprint arXiv:2111.04926*.
- ZADROZNY, B. (2003): “Policy mining: Learning decision policies from fixed sets of data,” Ph.D. thesis, University of California, San Diego.
- ZHANG, B., A. A. TSIATIS, M. DAVIDIAN, M. ZHANG, AND E. LABER (2012): “Estimating optimal treatment regimes from a classification perspective,” *Stat*, 1, 103–114.
- ZHAO, Y., D. ZENG, A. J. RUSH, AND M. R. KOSOROK (2012): “Estimating individualized treatment rules using outcome weighted learning,” *Journal of the American Statistical Association*, 107, 1106–1118.

Figure 1: Variation in treatment propensity across individuals in the JTPA Study sample



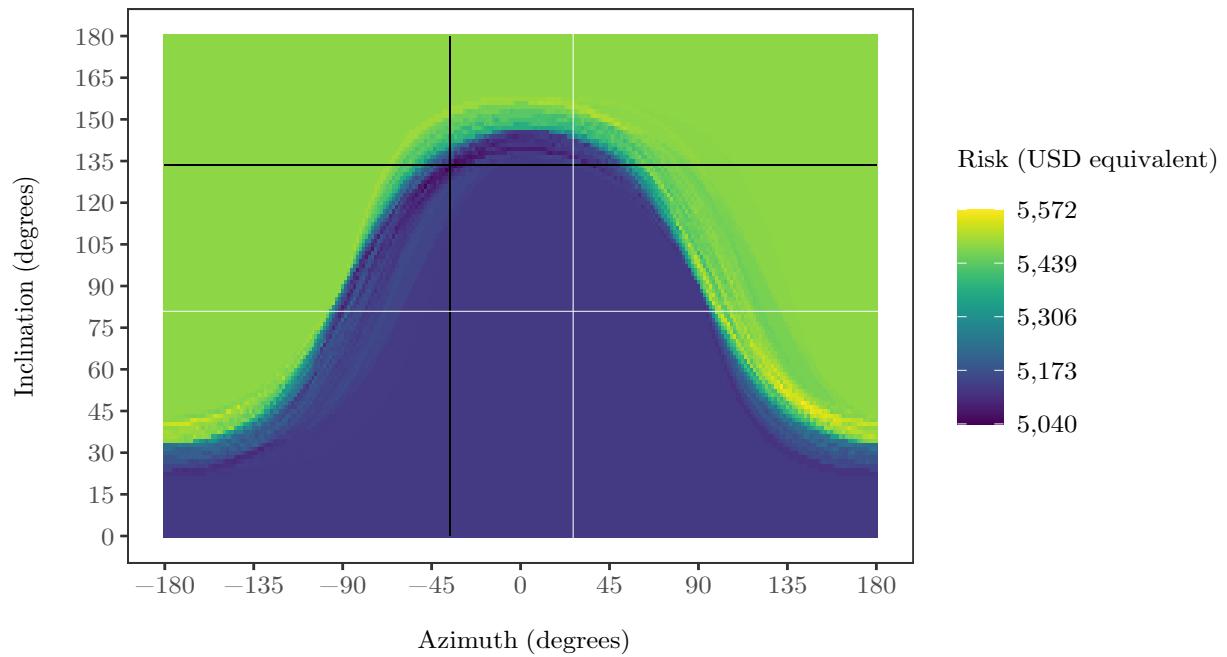
Notes: JTPA Study sample. This figure illustrates the treatment propensity of individuals under the posterior assignment rule that is induced by (κ^*, μ^*) . Each point represents the individual characteristics of an individual or several individuals in the (crosses denote individuals with zero in-sample weight). For comparison, individuals to the left of the solid diagonal line are assigned treatment under the optimal deterministic assignment rule of Kitagawa and Tetenov (2018).

Figure 2: Behaviour of the objective function at μ^* given variation in κ



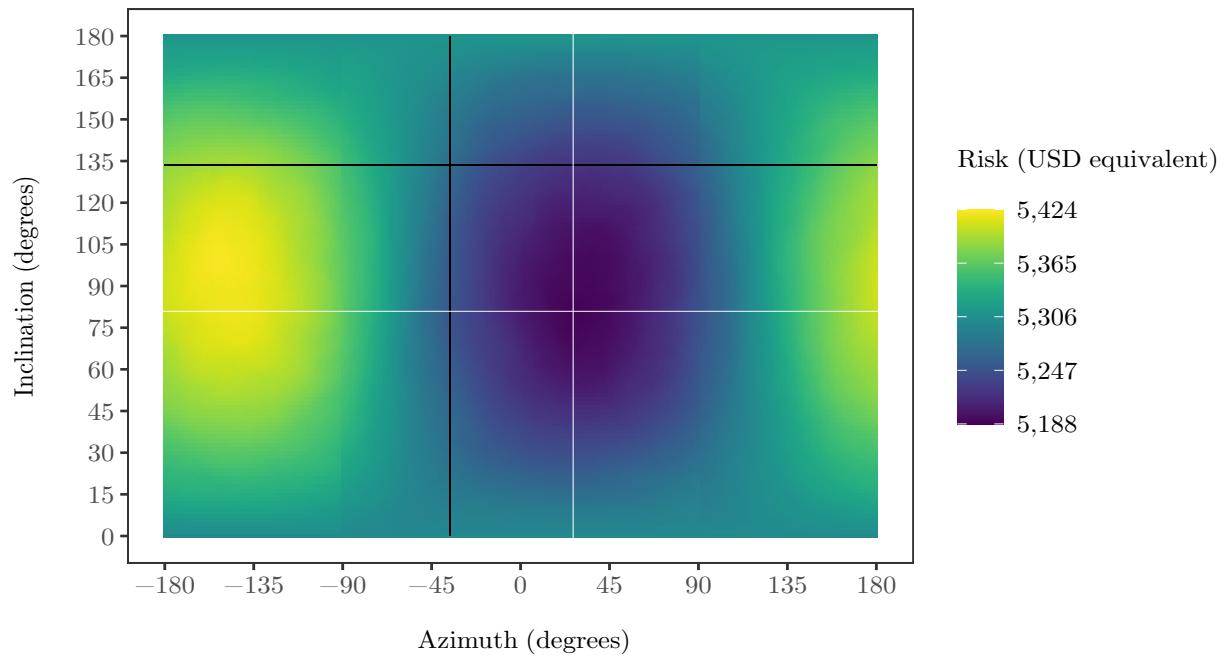
Notes: JTPA Study sample. This figure illustrates the shape of the objective function and its risk component at μ^* as κ is varied; high concentration is associated with low risk but incurs a large penalty for divergence from the uniform prior.

Figure 3: Deterministic assignment rules and empirical welfare risk



Notes: JTPA Study sample. This figure illustrates the risk that is associated with (deterministic) assignment rules in \mathcal{F} . A spherical coordinate mapping is implemented. The intersection of the two white lines is located at μ^* . The intersection of the two black lines is located at the optimal deterministic assignment rule of Kitagawa and Tetenov (2018), which attains the minimal regret amongst all deterministic linear rules.

Figure 4: Behaviour of the objective function at κ^* given variation in the mean direction μ



Notes: JTPA Study sample. This figure illustrates the risk that is associated with (stochastic) assignment rules in \mathcal{V} ; the concentration parameter is fixed at κ^* whilst μ is varied. The intersection of the two white lines is located at μ^* . The intersection of the two black lines is located at the optimal deterministic assignment rule of Kitagawa and Tetenov (2018), which attains the minimal regret amongst all deterministic linear rules.

Supplement to *Stochastic treatment choice with empirical welfare updating*

Toru Kitagawa

Hugo Lopez

Jeff Rowley

February 22, 2023

Abstract

We apply the methods that are outlined in *Stochastic treatment choice with empirical welfare updating* to experimental data from the Job Training Partnership Act Study and to several numerical simulations.

B Accounting for the cost of treatment in the JTPA Study sample

In what follows, we distinguish between the JTPA Study sample, which we sometimes refer to as the raw data, and the cost-adjusted JTPA Study sample, which we sometimes refer to as the costed data. The costed data subtracts the cost of treatment from the outcome of interest (post-programme earnings) of all treated individuals.¹ Following Kitagawa & Tetenov (2018), we assume that the cost of treatment is \$774. We then proceed to search for the optimal stochastic assignment rule, applying the same grid search approach as we outlined in the main text for the raw data, using the costed data.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

We find that the objective function is minimised by the stochastic assignment rule with $\kappa = 0.560$ and $\mu = (0.872, 0.490, 0.018)$, which we label κ^a and μ^a , respectively.² The value of the objective function and the empirical welfare risk induced by this assignment rule are equivalent to \$9,302 and \$5,155, respectively.

*The authors thank participants at the Bristol Econometrics Study Group and Midwest Econometrics Group conferences for beneficial comments. The authors also gratefully acknowledge financial support from the ERC (grant number 715940) and the ESRC Centre for Microdata Methods and Practice (CeMMAP; grant number RES-589-28-0001).

[†]Department of Economics, Brown University. Email: toru_kitagawa@brown.edu

[‡]Department of Economics, University of Chicago.

[§]Department of Economics, University College London.

¹We observe that this leads individuals with zero post-programme earnings to have a negative outcome. This violates Assumption 1, which requires that all outcomes be bounded and non-negative. We, nonetheless, proceed with this adjustment of post-programme earnings as is. An alternative would be to add \$774 to all post-programme earnings before subtracting the assumed cost of treatment.

²This directional vector can be represented by an azimuth of 29° and an inclination of 89° using spherical coordinates.

In comparison, Kitagawa & Tetanov (2018) estimates that the deterministic assignment rule defined by $\beta = (0.117, -0.990, -0.086)$ minimises empirical welfare risk. Relative to the raw analysis, this assignment rule is much flatter, and essentially determines that all individuals with low pre-programme earnings (earnings of around \$5,000 or less) should be treated, irrespective of education.

Figures 5 to 8 are intended to be comparable to the figures in the main text. There are two differences of note between the results of the raw and costed analyses. First, in Figure 6, the relationship between κ and empirical welfare risk is much more muted, with risk only decreasing by around \$50 as κ increases (over the interval that we study). In contrast, empirical welfare risk decreases by around \$200 when the cost of treatment is not accounted for. This reflects the second difference, which is apparent in Figure 7. Specifically, that the difference in empirical welfare risk between those assignment rules that assign everyone to treatment versus those that assign no-one to treatment is much smaller. The *benefit* to an increase in κ , which is a reduction in the probability mass allocated to those assignment rules that assign no-one to treatment is, accordingly, smaller. The implication is that the penalty term starts to dominate the objective function for smaller values of κ .

To supplement our analysis, we include two additional figures in Figures 9 and 10. These figures are intended to give some idea about how the concentration of the von Mises-Fisher distribution influences the propensity with which individuals are assigned to the experimental group. For example, given a particular instance of the von Mises-Fisher distribution, one individual might have a high propensity of assignment whereas another might have a low propensity. In other words, the number of directional vectors that are drawn from the von Mises-Fisher distribution for which the first individual is assigned to the experimental group is greater than for the second individual. As κ increases, so these directional vectors concentrate around the mean direction, and the propensity of assignment to the experimental group approaches zero or one for each individual. We see from Figure 9 that, when μ is fixed at μ^* , the propensity of assignment in the costed data tends towards one for all individuals. In contrast, and to illustrate the possibility of complete dichotomy, for the raw data, we see from Figure 10 that the propensity of assignment diverges as κ increases. We observe in both datasets that much of the variation in propensity is driven by a few individuals in the tail, who are likely those individuals whose individual characteristics make them outliers. In both cases, the median propensity of treatment (i.e., the probability with which the average individual in the sample is assigned to the experimental group) tends towards one.

C Numerical simulations

We propose several experiments that investigate how various aspects of the sample data that is available to the social planner affect the posterior distribution and its shape when the specified prior distribution is uniform over the sphere. We conduct these experiments to better understand some of the empirical results that we obtain for the main paper.

Aspects of the sample data that we vary include the number of observations, the outcome of interest, and individual characteristics. We conduct these experiments with simulated data and through manipulation of the existing empirical application. Each experiment is comparable to the existing empirical application in any case, in that observed individual characteristics are taken to be education and pre-programme earnings alongside an intercept term, and the outcome of interest is taken to be post-programme earnings. We investigate how these aspects and our variation of them affect the objective function.

The specific questions that we ask, and that inform the design of our experiments, are as follows.

- How does the number of observations influence the shape of the objective function?
- How does the relative influence of education and pre-programme earnings on post-programme earnings influence empirical welfare risk?
- How does the distribution of individual characteristics influence the shape of the objective function?

To address these questions we propose a series of linear specifications that satisfy the bounded outcomes assumption (i.e., the outcome of interest can be restricted to some subset $[0, M]$, where $M < \infty$) that we require. We concede that these specifications are somewhat contrived, reflecting our need to balance

tractability with the requirements of the bounded outcomes assumption. Denoting individual characteristics by X , we suppose that

$$X = (1, X_{\text{earn}}, X_{\text{educ}}) \in \{1\} \times [0, 1]^2, \quad (\text{C.1})$$

which can always be maintained via an appropriate affine map of the individual characteristics (a step that we undertake in any case), and propose that

$$\begin{aligned} Y_1 &= X^\top \eta + \epsilon_1, & \epsilon_1 &\sim N(0, \sigma_1^2) \text{ truncated to } (0, m), \\ Y_0 &= X^\top \alpha + \epsilon_0, & \epsilon_0 &\sim N(0, \sigma_0^2) \text{ truncated to } (0, m), \end{aligned} \quad (\text{C.2})$$

with

$$\begin{aligned} \eta &\in \{v \in \mathbb{R}^3 : v_1 + v_2 + v_3 \geq 0\}, \\ \alpha &\in \{v \in \mathbb{R}^3 : v_1 + v_2 + v_3 \geq 0\}, \end{aligned} \quad (\text{C.3})$$

where Y_1 and Y_0 denote the potential outcomes associated with treatment and non-treatment, respectively. The advantage of this specification is that it provides a clear interpretation of the influence of education and pre-programme earnings on post-programme earnings, and facilitates addressing all of the questions that we pose. Moreover, the potential outcomes are restricted to the lie in the union of the sets $[0, \eta + m]$ and $[0, \alpha + m]$. We note that the conditional expectations of the potential outcomes are

$$\begin{aligned} E(Y_1|x) &= x^\top \eta + \frac{\phi(0) - \phi(m/\sigma_1)}{\Phi(m/\sigma_1) - \Phi(0)} \cdot \sigma_1 = x^\top \tilde{\eta}, \\ E(Y_0|x) &= x^\top \alpha + \frac{\phi(0) - \phi(m/\sigma_0)}{\Phi(m/\sigma_0) - \Phi(0)} \cdot \sigma_0 = x^\top \tilde{\alpha}, \end{aligned} \quad (\text{C.4})$$

which are both affine functions of individual characteristics.³ This is an attractive property that we exploit.

Knowledge of the data generating process (Equation (C.2) in our framework) is sufficient to determine the optimal assignment rule.⁴ Specifically, the optimal assignment rule delivers the partition

$$G = \{x : E(Y_1|x) \geq E(Y_0|x)\}. \quad (\text{C.5})$$

When the conditional expectation of the potential outcomes are affine functions of individual characteristics (and assignment is at random, which is an assumption that we implicitly maintain throughout) then the optimal assignment rule belongs to the LES class. Hence, Equation (C.4) guarantees that the optimal policy has the specific form

$$G = \{x : x^\top \beta \geq 0, \beta = (\tilde{\eta} - \tilde{\alpha})/\|\tilde{\eta} - \tilde{\alpha}\|_2\}, \quad (\text{C.6})$$

which is not only a member of the LES class but emphasises that the optimal assignment rule can be summarised by a vector on the sphere (provided that the potential outcomes have distinct process). This result provides motivation for our reliance on spherical distributions and, in particular, the von Mises-Fisher distribution. We observe that, abstracting from the issue of sampling variation, empirical welfare risk is minimised when the mean direction of the von Mises-Fisher distribution coincides with the policy defined in Equation (C.6). This insight allows us to focus exclusively on the influence of the concentration parameter on empirical welfare risk and the objective function.

Throughout, we are careful to sample using inversion-based pseudo-random sampling methods where possible. Compared to rejection sampling-based pseudo-random sampling methods, inversion-based methods are able to guarantee comparability across experiments despite differences in parameter values. We also note that where we use data from the JTPA Study sample, this data is not adjusted for the cost of treatment (i.e., we use the raw data). We continue to label those values of the parameters that minimise the objective function by κ^* and μ^* , which we emphasise can vary across the various experiments that we conduct.

³Other commonly invoked models fail to meet our test of tractability or the requirements of the bounded outcomes assumption: the standard censored outcome model generates a rectified normal distribution that is non-linear in individual characteristics and is unbounded from above; logarithmic transformation of the outcome implies a non-linear transformation of individual characteristics.

⁴We do not make any claims about uniqueness in what follows and, indeed, presented with a finite sample of individual characteristics, it is likely that several assignment rules can attain the same partition of individuals as what we refer to as the optimal assignment rule.

C.1 Varying the number of observations

[Figure 11 about here.]

To investigate the influence that the number of observations has on the objective function and on empirical welfare risk, we copy each observation in the JTPA Study sample either once or three times, thereby doubling and quadrupling, the number of individuals in the sample. The mechanical effect of this change is to mute the influence of the penalty term in the objective function, without affecting empirical welfare risk. The immediate implication is that larger values of the concentration parameter can be sustained, since the penalty due to deviating from the uniform distribution is relatively smaller for every such deviation.

We plot the objective function and empirical welfare risk for the doubled and quadrupled JTPA Study samples in Figure 11. We observe that increasing the number of observations has the direct effect of decreasing the magnitude of the objective function. Moreover, increasing the number of observations also leads to an increase in κ^* , which increases from 1.550 to 1.890 and then to 2.490. We emphasise that μ^* is also not the same across the two cases. The intuition here is that an increase in κ leads to the concentric contour map of the density function becoming more tightly arranged around μ (whatever that may be). Locating μ closer to the boundary between high and moderate regret regions say, such as where the deterministic assignment rule of Kitagawa & Tetenov (2018) is located, incurs less of a penalty in this instance since the density function assigns less probability mass to the high regret region than it would for a smaller value of the κ . For the doubled sample we find that $\mu^* = (0.812, 0.577, 0.088)$, whilst for the quadrupled sample we find that $\mu^* = (0.917, 0.394, 0.053)$.⁵

C.2 The variance of post-programme earnings

In the following three experiments, we investigate the influence of the variance of post-programme earnings on the results of our method. Specifically, we investigate how altering σ_0 and σ_1 in the specification outlined in Equation (C.2) affects our results. In particular, we are interested in whether there is a fundamental change in how empirical welfare risk varies with the parameters of the von Mises-Fisher distribution.

[Figure 12 about here.]

The JTPA Study sample consists of 9,223 observations, and we extract treatment status and individual characteristics directly from this dataset. We then generate potential outcomes according to Equation (C.2) (Page 34). We implement the affine map of Equation (C.1) via the transformations

$$\begin{aligned} X_{\text{earn}} &\mapsto X_{\text{earn}} / \max\{X_{\text{earn}}\}, \\ X_{\text{educ}} &\mapsto X_{\text{educ}} / \max\{X_{\text{educ}}\}. \end{aligned} \quad (\text{C.7})$$

We then regress post-programme earnings in the experimental group and the control group on these characteristics, separately and together, so that the baseline experiment somewhat mimics the JTPA Study sample. Using our simple regressions as a rough guide, we let

$$\begin{aligned} \eta &= (+3, 040, +86, 446, +14, 008), & \sigma_1 &= 15, 914, & m &= 5\sigma_1 = 5\sigma_0, \\ \alpha &= (-1, 086, +82, 458, +18, 804), & \sigma_0 &= 15, 914, \end{aligned} \quad (\text{C.8})$$

so that the optimal policy is

$$G = \{x : x^\top \beta \geq 0, \beta = (+0.552, +0.533, -0.641)\}, \quad (\text{C.9})$$

which we plot in Figure 12. We note that the coefficients of the linear specification are so large because our simple regressions scale individual characteristics but do not scale post-programme earnings, which inflates the effect of pre-programme earnings and years of education. This constitutes our first experiment.

We then propose two further experiments. Our second experiment increases the variance of post-programme earnings by inflating σ_1 , σ_0 and m in Equation (C.8) by a factor of five. Our third experiment

⁵These directions translate (azimuth:inclination) to $35^\circ : 85^\circ$ and $23^\circ : 87^\circ$, respectively, as compared to $\mu^* = (0.883, 0.442, 0.158)$ or $27^\circ : 81^\circ$ in the original sample.

similarly reduces the variance of post-programme earnings by deflating σ_1 , σ_0 and m in Equation (C.8) by a factor of five. We note that both experiments leave the oracle policy unchanged from Equation (C.9). We present the corresponding estimates of the parameters of the posterior distribution immediately below.

Experiment	μ^*	κ^*
1. Baseline	(+0.822,+0.565,+0.078)	1.850
2. High variance	(+0.641,+0.703,-0.309)	0.340
3. Low variance	(+0.831,+0.547,+0.105)	1.970

An immediate conclusion that we can draw from these results is that μ^* does not align with the oracle assignment rule. That being said, it is apparent from Figure 12 that individuals who are assigned treatment under the oracle assignment rule are more likely to be assigned treatment under the posterior distribution that we obtain.

[Figure 13 about here.]

We plot the behaviour of the objective function and of empirical welfare risk in Figure 13. We emphasise the non-smoothness of empirical welfare risk for the second experiment, and we suggest that increasing the variance of post-programme earnings makes the problem of finding the optimal assignment rule more difficult.

C.3 The location of the oracle assignment rule

A common feature of the data generating process outlined in Equation (C.8) for Experiments 1 through 3 is the oracle assignment rule, which is located in the south-east corner of the covariate space (Figure 12). Two things can be inferred from this. First, that only a small minority of individuals that are educated to post-graduate level are not assigned treatment under the oracle assignment rule. In other words, a small subset of the data with extreme characteristics. Second, that, all else being equal, average outcomes in the experimental group are higher than average outcomes in the control group, and that this baseline difference is relatively important as compared to pre-programme earnings and years of education. These features, we suggest, are indicative of a relatively easy problem (of determining the optimal assignment rule).

[Figure 14 about here.]

The following three experiments largely replicate Experiments 1 through 3 but shift the oracle assignment rule to the left, narrowing the difference between the baseline average outcomes in the experimental and control groups. In Experiment 4, we assume that

$$\begin{aligned} \eta &= (+2, 442, +86, 446, +14, 008), & \sigma_1 &= 15,914, & m &= 5\sigma_1 = 5\sigma_0, \\ \alpha &= (-489, +82, 458, +18, 804), & \sigma_0 &= 15,914, \end{aligned} \quad (\text{C.10})$$

so that the optimal policy is

$$G = \{x : x \cdot \beta \geq 0, \beta = (+0.425, +0.579, -0.696)\}, \quad (\text{C.11})$$

which we plot in Figure 14. Experiments 5 and 6 then mirror Experiments 2 and 3 in that they inflate and deflate, respectively, the variance of post-programme earnings. We suggest that Equation (C.10) is a more difficult problem than Equation (C.8).

We specifically design these experiments so that the oracle assignment rule is such that as close to 50% of the sample is assigned to treatment as is possible without altering the importance of pre-programme earnings relative to years of education in the earnings process. We present the corresponding estimates of the parameters of the posterior distribution immediately below.

Experiment	μ^*	κ^*
4. Baseline	(+0.932,+0.362,-0.035)	0.00
5. High variance	(-0.810,-0.526,-0.259)	0.390
6. Low variance	(+0.899,+0.425,+0.105)	0.00

[Figure 15 about here.]

A curious feature of these results is that, for Experiments 4 and 6, the posterior distribution is uniform. We note that the mean direction is irrelevant in this case, but we include it anyway because knowledge of it is necessary to interpret Figure 15. This uniformity of the posterior distribution is apparently driven by how empirical welfare risk increases alongside κ in Figure 15. A possible explanation for this feature is that this data generating process is indeed hard. We suggest that the narrow gap between the baseline average outcomes of the experimental and control groups makes treating everyone versus treating no-one equally appealing (or unappealing). For instance, contrasting Figure 7 with its analogue in the main text (for the raw data), we see that the difference in empirical welfare risk between the *treat everyone rules* and *treat no-one rules* that occupy the southern and northern regions of the heatmaps, respectively, narrows. A similar effect happens here. Given the fact that the density contours of the von Mises-Fisher distribution are concentric, there is then no natural region of the sphere to locate in order to minimise empirical welfare risk unless κ is particularly large and the posterior distribution allocates substantial probability mass to an extremely localised area on the frontier between the collections of *treat everyone rules* and *treat no-one rules*. High concentration is heavily penalised by the Kullback-Leibler divergence though, and so uniformity cannot be improved upon. The fact that a different pattern is observed for Experiment 5 is compatible with this argument, and could be achieved if the high variance of post-programme earnings makes a few observations in the sample pivotal (recall that the weight attached to each observation is proportional to post-programme earnings; inflating the variance of post-programme earnings can lead to greater concentration of weight on a few observations, since extreme outliers are more likely).

C.4 Individual characteristics and their distribution

[Figure 16 about here.]

The final experiments that we undertake vary the distribution of individual characteristics. Our intention is to understand whether the JTPA Study sample is, in some sense, special and whether altering individual characteristics substantially alters our results.

Experiments 7 and 8 mirror Experiment 1 in how the outcome is generated: both experiments rely on Equation (C.2) and the parameter values that we outline in Equation (C.8) for Experiment 1. The distinction between Experiments 7 and 8 and Experiment 1 is how individual characteristics are generated. Whereas Experiment 1 uses data taken from the JTPA Study sample, Experiments 7 and 8 generate individual characteristics according to a bivariate normal distribution with the mean vector equal to half of the maximum of pre-programme earnings and the average number of years of education, respectively. We estimate the covariance matrix of pre-programme earnings and years of education in the JTPA Study sample, and set the covariance of the bivariate normal distribution equal to this (Pearson correlation coefficient of 0.126). In the case of Experiment 7, we discretise years of education by assigning each observation to one of 12 equal-sized bins. We then map each characteristic to the unit interval by means of the aforementioned linear transformation.

We plot the individual characteristics that we use in Experiments 7 and 8 in Figure 16. Absent from either plot is the oracle assignment rule. This is not an oversight. Rather, the oracle assignment rule is such that it recommends that all individuals be assigned treatment (i.e., it to the south-east of the plotting area). We plot the criterion (left-hand panel) and empirical welfare risk (right-hand panel) for Experiments 7 and 8 in Figure 17 (Page 51). Due to the similarity of Figure 17 to the other figures that we have presented, we do not provide any further discussion of this figure. What is somewhat interesting though is how the propensity of treatment is highest for those individuals with individual characteristics located around their mean values.

We present the corresponding estimates of the parameters of the posterior distribution immediately below.

Experiment	μ^*	κ^*
7. Discrete	(+0.794,+0.405,+0.454)	3.090
8. Continuous	(+0.801,+0.408,+0.438)	3.120

We find that μ^* does not align with the oracle assignment rule, but does dictate that all individuals are assigned treatment (as they are under the oracle assignment rule); κ^* is also relatively large, as compared to its value in the previous experiments.

So as to make the assignment problem harder – at least, what we understand to be harder – we propose Experiments 9 and 10. These experiments follow Experiments 7 and 8 in how individual characteristics are generated, but differ slightly in how they generate the outcome. Whilst Experiments 9 and 10 broadly follow Experiments 7 and 8 with respect to how the outcome is generated, they shift the process for the potential outcomes and, thereby, the oracle assignment rule. Specifically, in Experiments 9 and 10, we assume that

$$\begin{aligned}\eta &= (+1, 286, +86, 446, +14, 008), & \sigma_1 &= 15,914, & m &= 5\sigma_1 = 5\sigma_0, \\ \alpha &= (-668, +82, 458, +18, 804), & \sigma_0 &= 15,914,\end{aligned}\tag{C.12}$$

such that the oracle policy is

$$G = \{x : x \cdot \beta \geq 0, \beta = (+0.098, +0.636, -0.765)\}.\tag{C.13}$$

The effect of this change is to maintain the slope of the oracle assignment rule (i.e., the contribution of pre-programme earnings relative to years of education) but to alter the intercept. In other words, to narrow the difference in baseline average outcomes between the experimental and control groups. In this regard, Experiments 9 and 10 are similar in intent to Experiment 4.

[Figure 17 about here.]

[Figure 18 about here.]

We plot the individual characteristics that we use in Experiments 9 and 10 in Figure 18 (Page 52). The oracle assignment rule is designed to partition the sample into approximately two groups of equal size. We plot the criterion (left-hand panel) and empirical welfare risk (right-hand panel) for Experiments 9 and 10 in Figure 19 (Page 53). We draw attention to several differences between the results that we obtain for Experiments 9 and 10 versus Experiments 7 and 8. First, we highlight the fall in the average propensity of treatment, as is apparent in Figure 18. Whereas in Experiments 7 and 8 (and the preceding experiments too), the propensity of treatment is close to one (i.e., everyone is assigned to the experimental group), here the propensity is closer to one third. There is more randomisation. Second, individuals that have a relatively high propensity of treatment as compared to their peers in Experiments 7 and 8 have a relatively low propensity of treatment in Experiments 9 and 10 (visually, the colours are inverted), as is apparent in Figure 18. That is, individuals whose characteristics are close to the mean values have a low propensity of treatment. Third, the value of the concentration parameter decreases from above three to below one (i.e., the posterior distribution is less concentrated and more uniform), as is apparent in Figure 19.

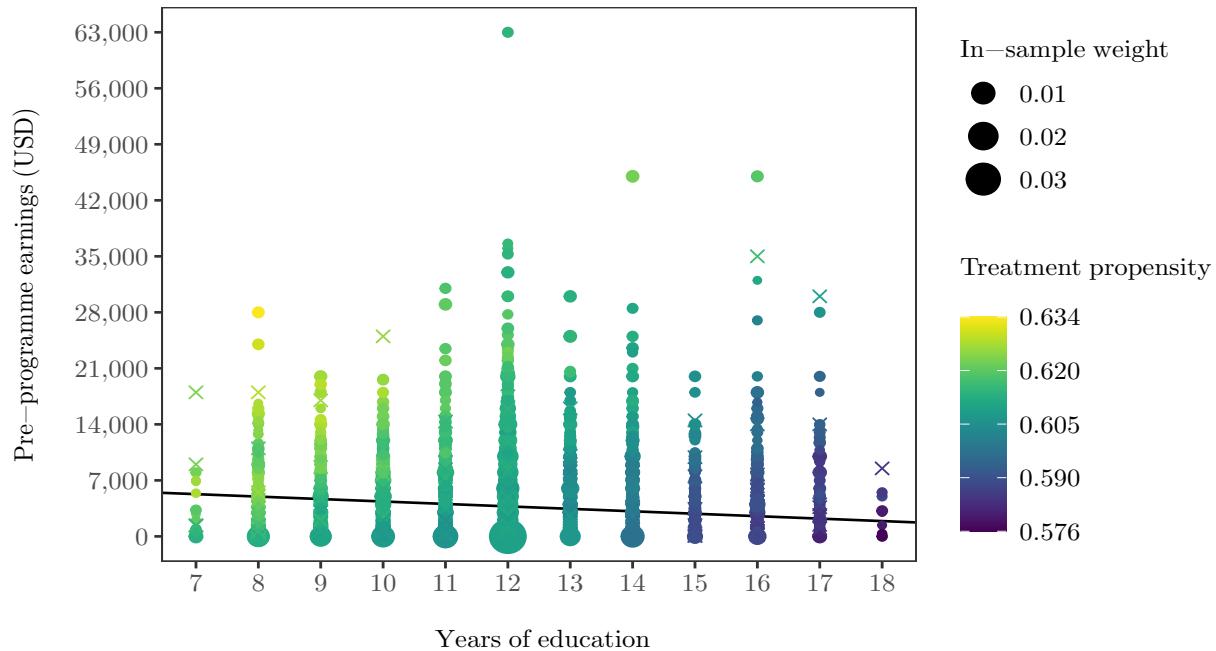
We present the corresponding estimates of the parameters of the posterior distribution immediately below.

Experiment	μ	κ
9. Discrete	(-0.766,-0.470,-0.438)	0.770
10. Continuous	(-0.766,-0.470,-0.438)	0.770

Interestingly, the mean direction is, in both cases, almost the exact opposite of the corresponding mean direction in Experiments 7 and 8 (i.e., the mean direction is approximately the negative of the mean direction in Experiments 9 and 10).

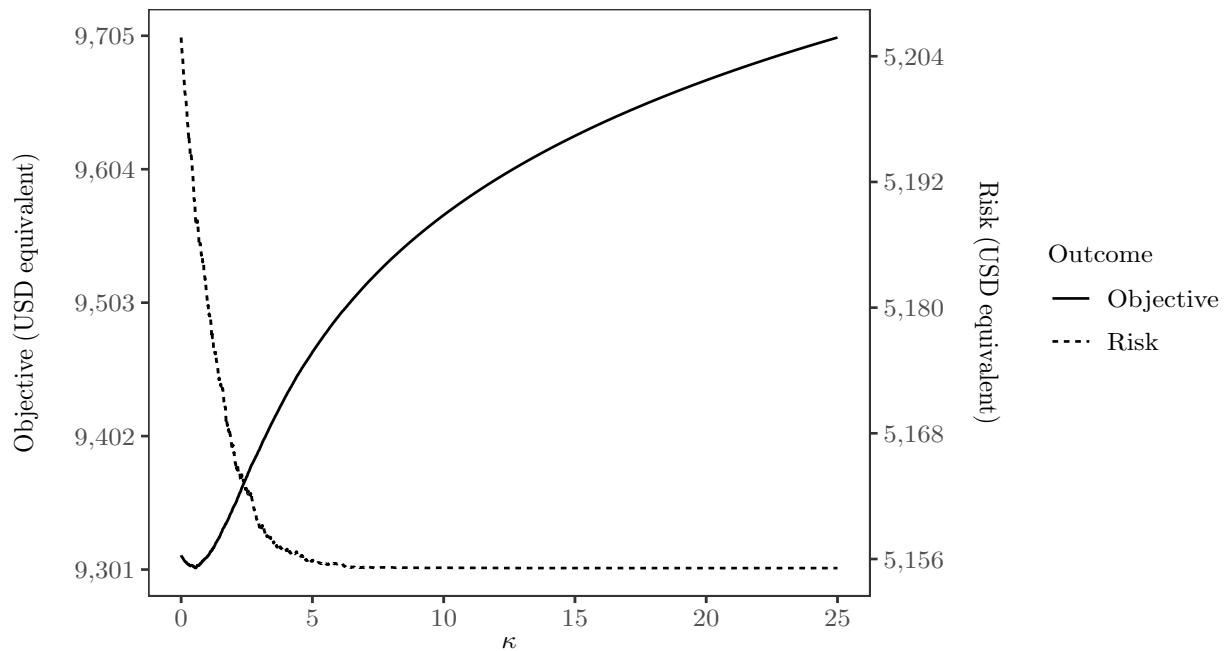
[Figure 19 about here.]

Figure 5: Variation in treatment propensity across individuals in the cost-adjusted JTPA Study sample



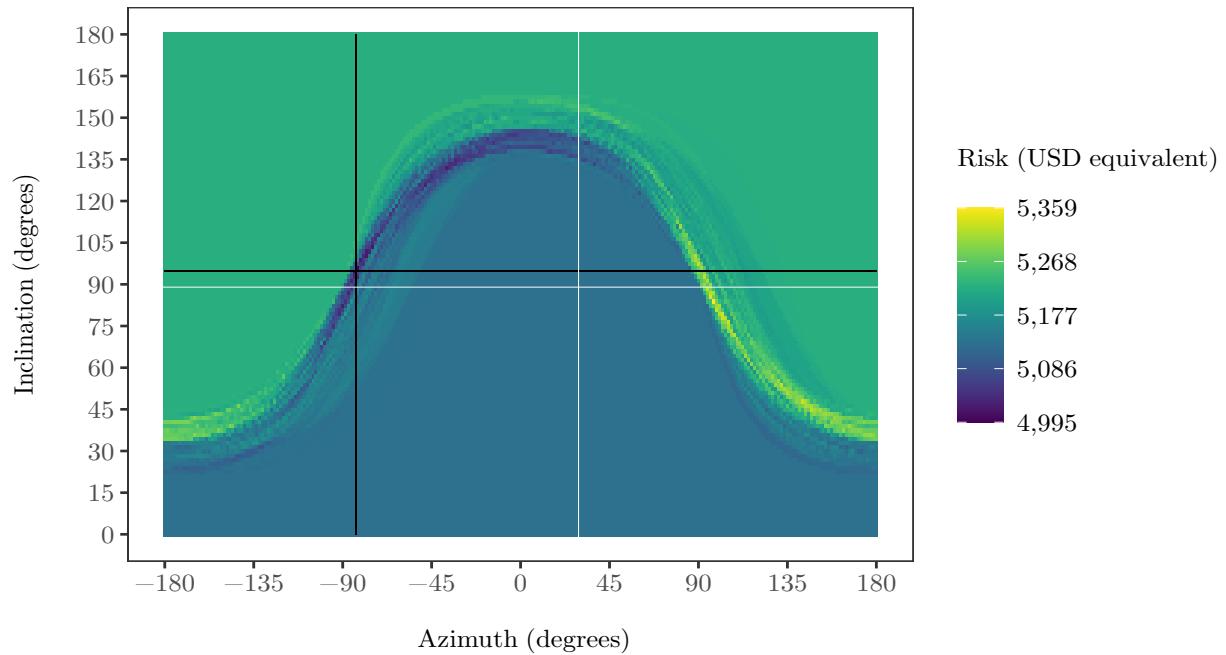
Notes: Cost-adjusted JTPA Study sample. This figure illustrates the treatment propensity of individuals under the posterior assignment rule that is induced by (κ^a, μ^a) . Each point represents the individual characteristics of an individual or several individuals (crosses denote individuals with negative in-sample weight). For comparison, individuals to the left of the solid diagonal line are assigned treatment under the optimal deterministic assignment rule of Kitagawa and Tetenov (2018).

Figure 6: Behaviour of the objective function at μ^* given variation in κ



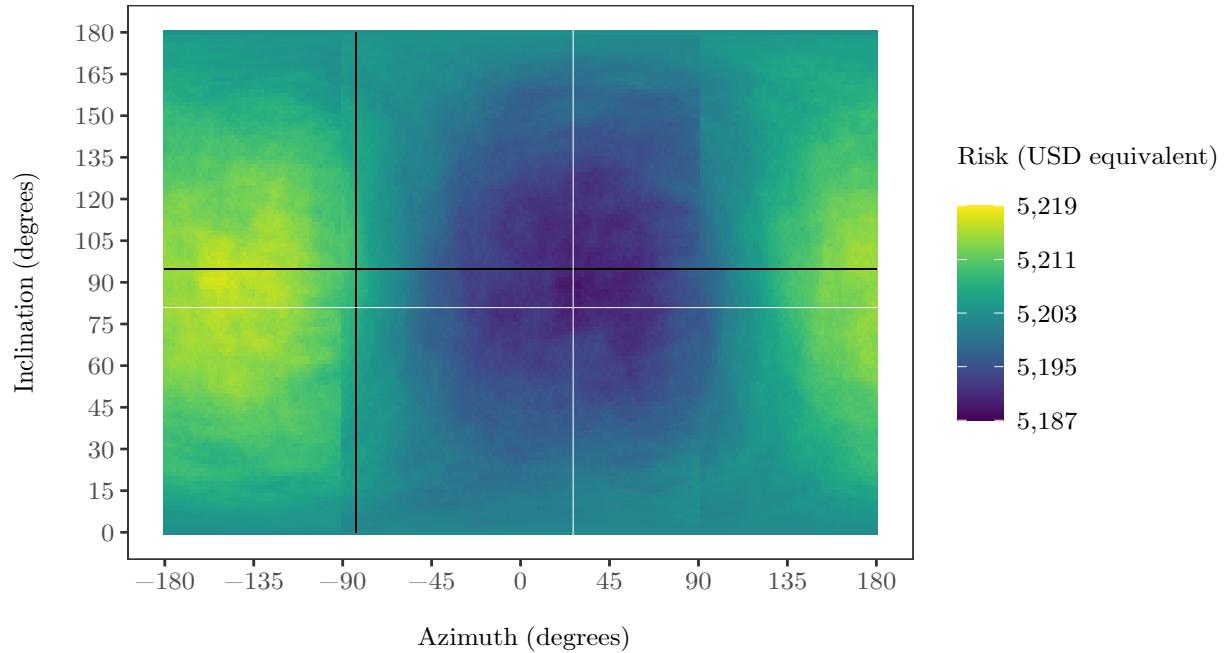
Notes: Cost-adjusted JTPA Study sample. This figure illustrates the shape of the objective function and its risk component at μ^a as κ is varied.

Figure 7: Deterministic assignment rules and empirical welfare risk



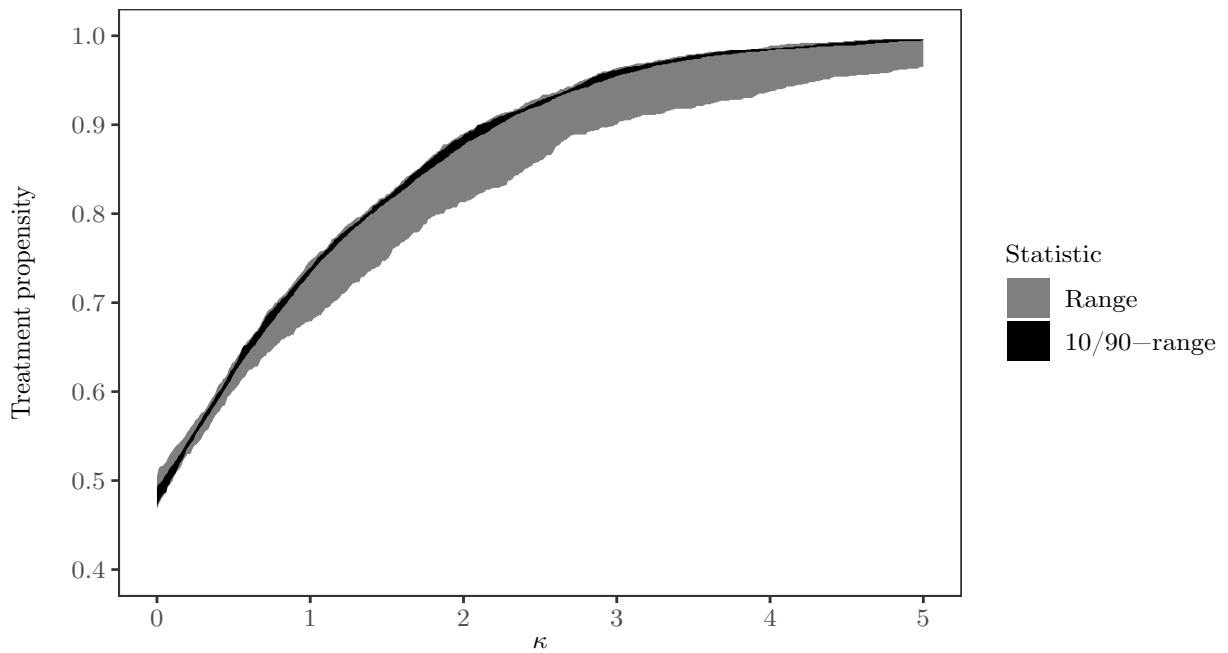
Notes: Cost-adjusted JTPA Study sample. This figure illustrates the risk that is associated with (deterministic) assignment rules in \mathcal{F} . A spherical coordinate mapping is implemented. The intersection of the two white lines is located at μ^a . The intersection of the two black lines is located at the optimal deterministic assignment rule of Kitagawa and Tetenov (2018), which attains the minimal regret amongst all deterministic linear rules.

Figure 8: Behaviour of the objective function at κ^a given variation in μ



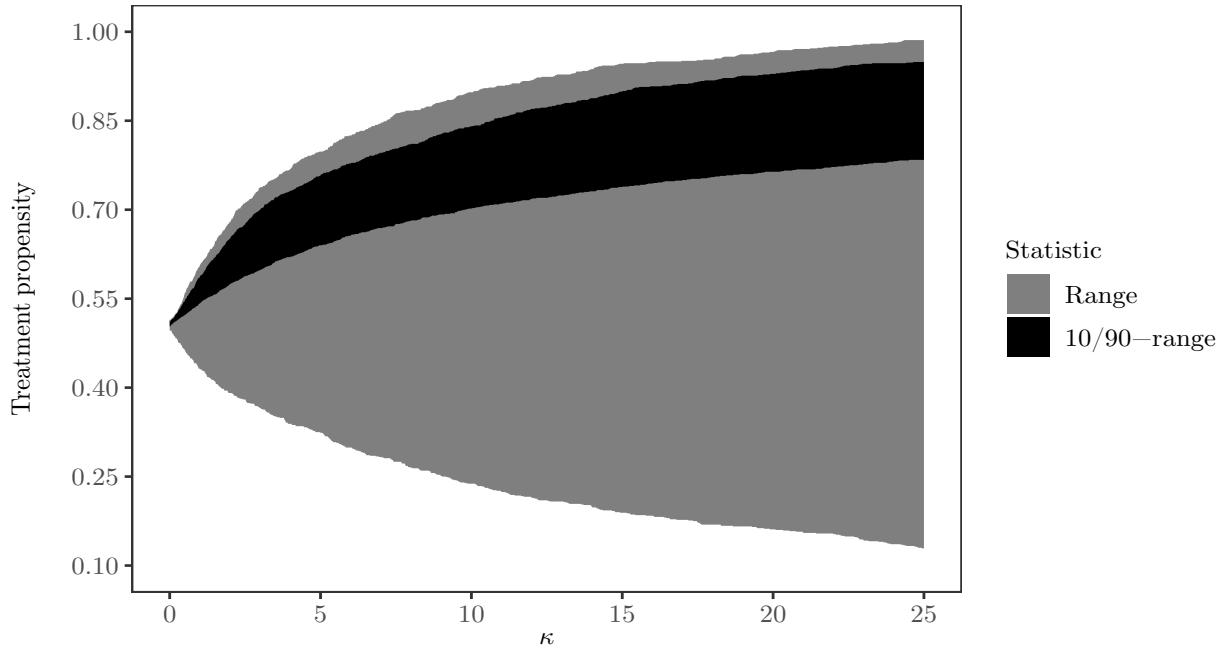
Notes: Cost-adjusted JTPA Study sample. This figure illustrates the risk that is associated with (stochastic) assignment rules in \mathcal{V} ; the concentration parameter is fixed at κ^a whilst μ is varied. The intersection of the two white lines is located at μ^a . The intersection of the two black lines is located at the optimal deterministic assignment rule of Kitagawa and Tetenov (2018), which attains the minimal regret amongst all deterministic linear rules.

Figure 9: Distribution of propensity of treatment across individuals (adjusted) at (κ^a, μ^a)



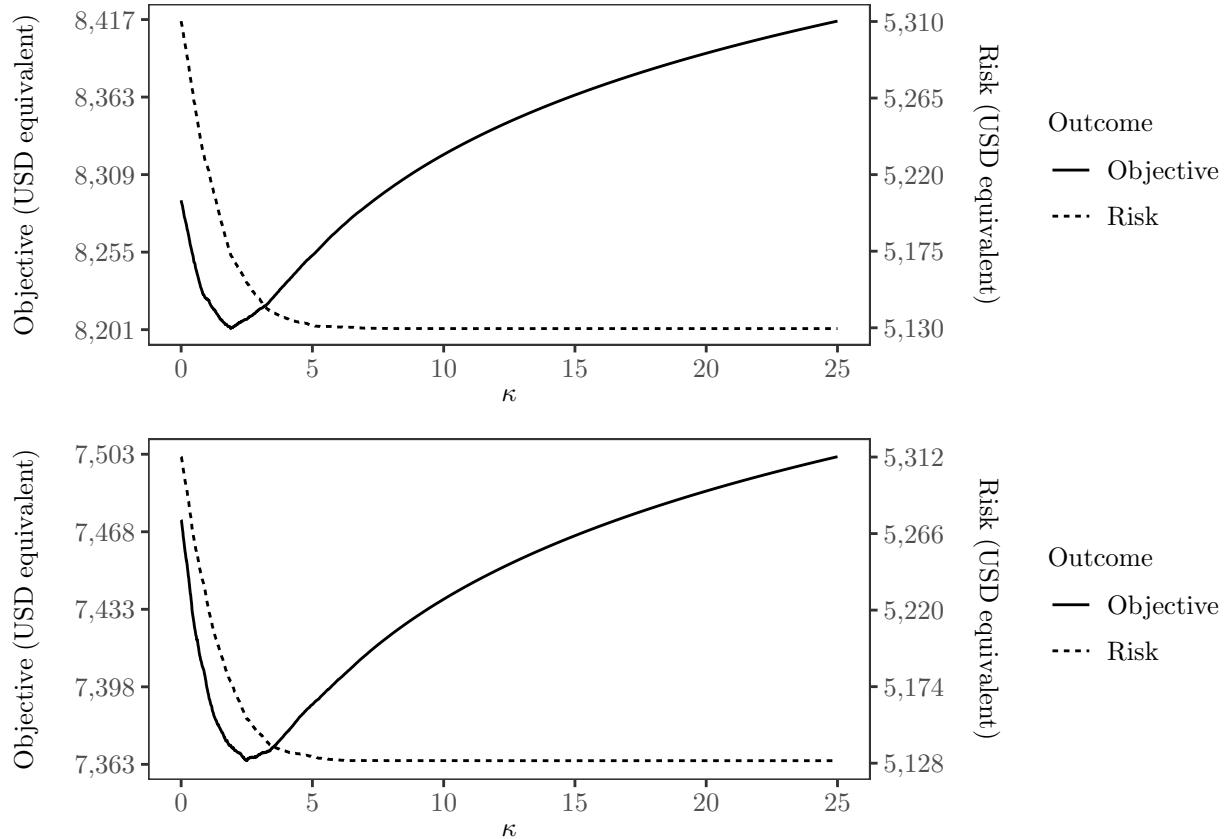
Notes: Cost-adjusted JTPA Study sample. This figure illustrates the variation in treatment propensity across individuals. We draw 1,000 directional vectors and count how many of these vectors, implemented as deterministic assignment rules, assign each individual to treatment. We sort individuals by how often they are assigned to treatment and summarise the implied distribution by the range and the 10th and 90th quantiles.

Figure 10: Distribution of propensity of treatment across individuals (raw) at (κ^*, μ^*)



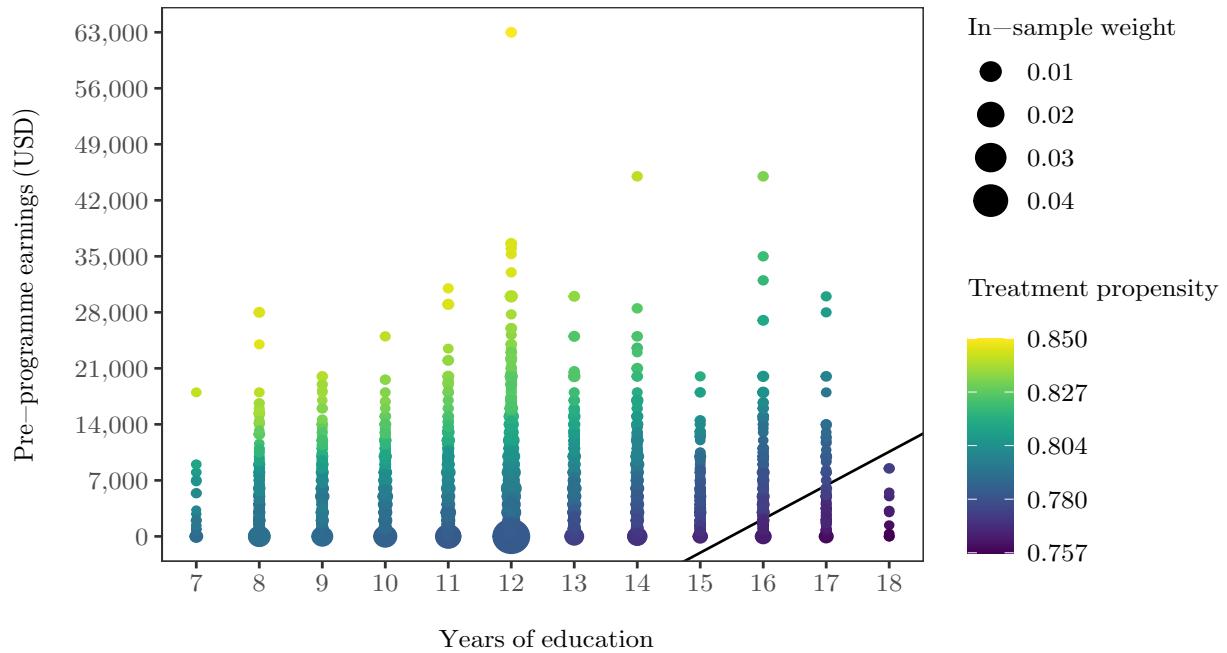
Notes: JTPA Study sample. This figure illustrates the variation in treatment propensity across individuals. We draw 1,000 directional vectors and count how many of these vectors, implemented as deterministic assignment rules, assign each individual to treatment. We sort individuals by how often they are assigned to treatment and summarise the implied distribution by the range and the 10th and 90th quantiles.

Figure 11: Behaviour of the objective function at μ^* given variation in κ



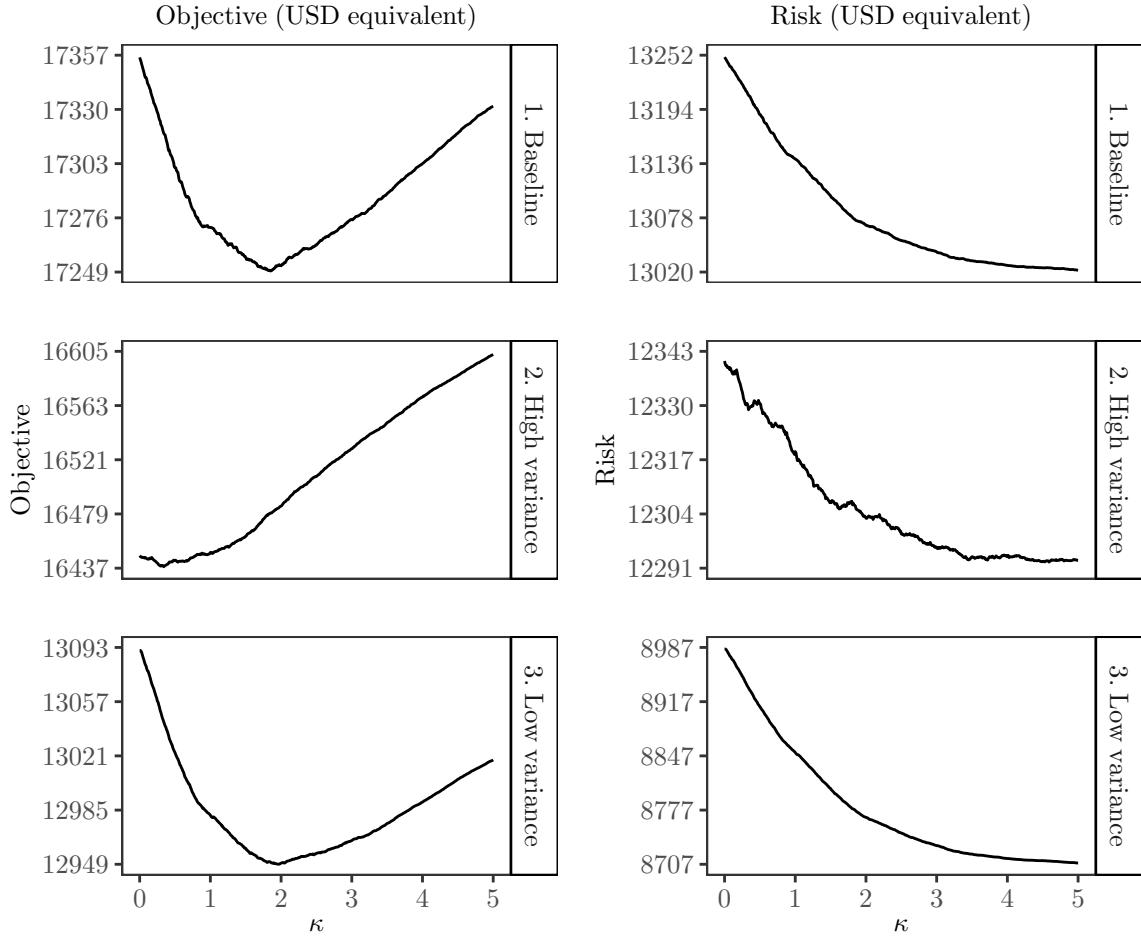
Notes: JTPA Study sample (artificially inflated by copying observations). This figure illustrates how the objective function and its risk component change as the number of observations is doubled (upper panel) and quadrupled (lower panel). We hold μ fixed at μ^* in each case, which differs according to the sample size, and vary κ .

Figure 12: (Experiment 1) Variation in treatment propensity across individuals



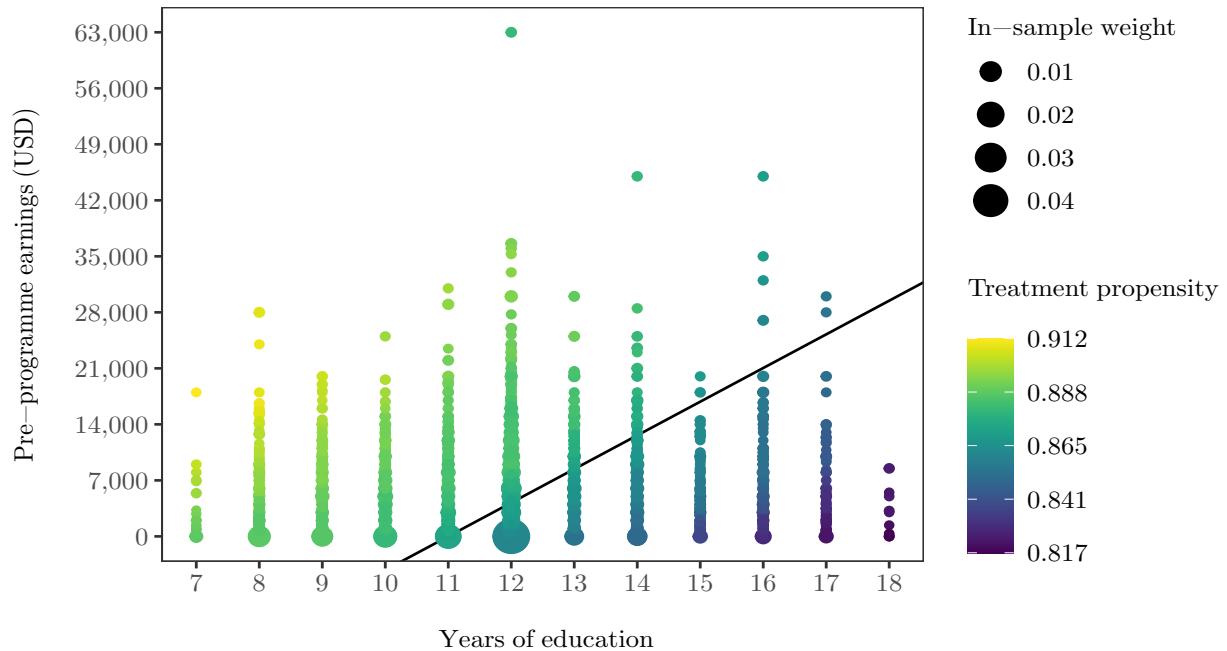
Notes: Simulated data generated using Equation (C.8) in conjunction with data from the JTPA Study sample. This figure illustrates the treatment propensity of individuals in a simulated sample under the posterior assignment rule that is induced by (κ^*, μ^*) . Each point represents the individual characteristics of an individual or several individuals. For comparison, individuals to the left of the solid diagonal line are assigned treatment under the oracle assignment rule of Equation (C.9).

Figure 13: (Experiments 1–3) Behaviour of the objective function at μ^* given variation in κ



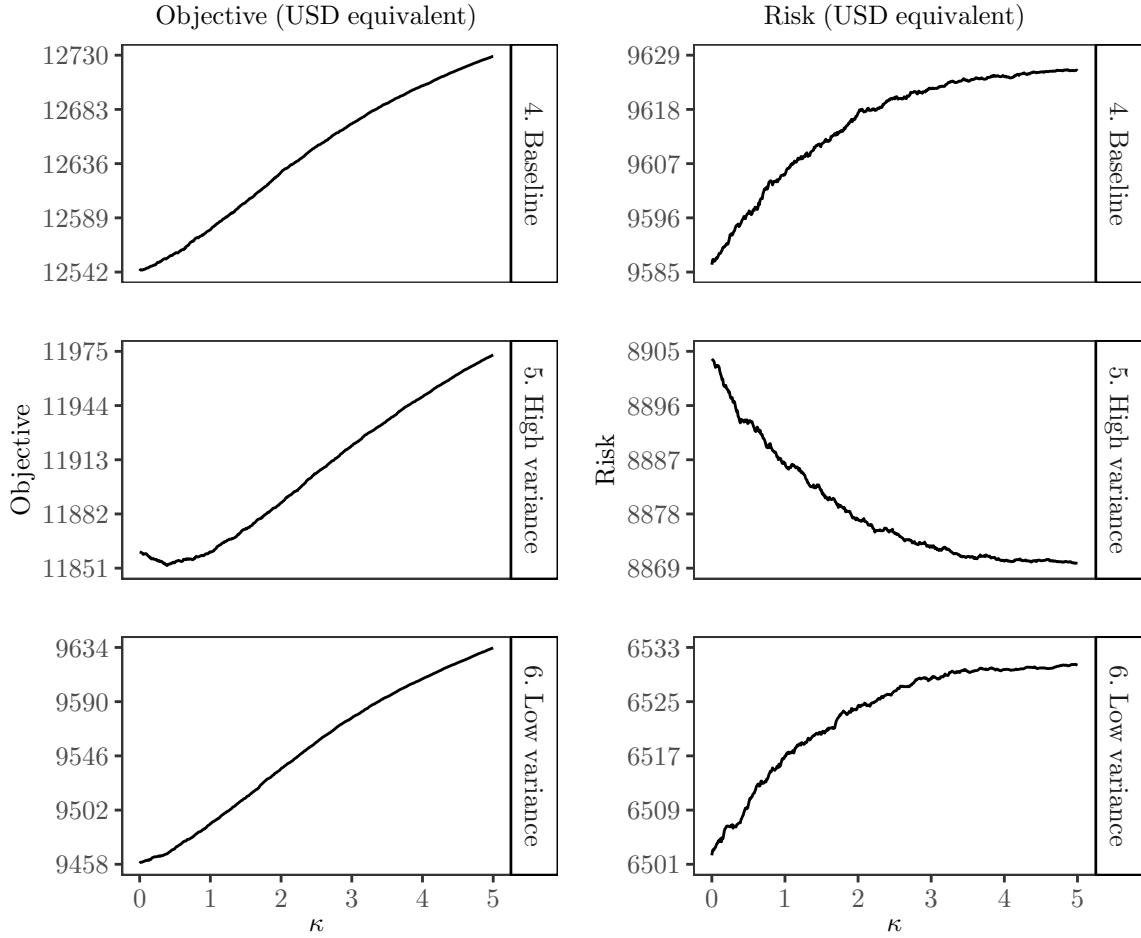
Notes: Simulated data generated using Equation (C.8) in conjunction with data from the JTPA Study sample. This figure illustrates how the shape of the objective function and its risk component changes as the variance of post-programme earnings is increased (middle panel) and decreased (lower panel). We hold μ fixed at μ^* in each case, which varies according to the sample size, and vary κ .

Figure 14: (Experiment 4) Variation in treatment propensity across individuals



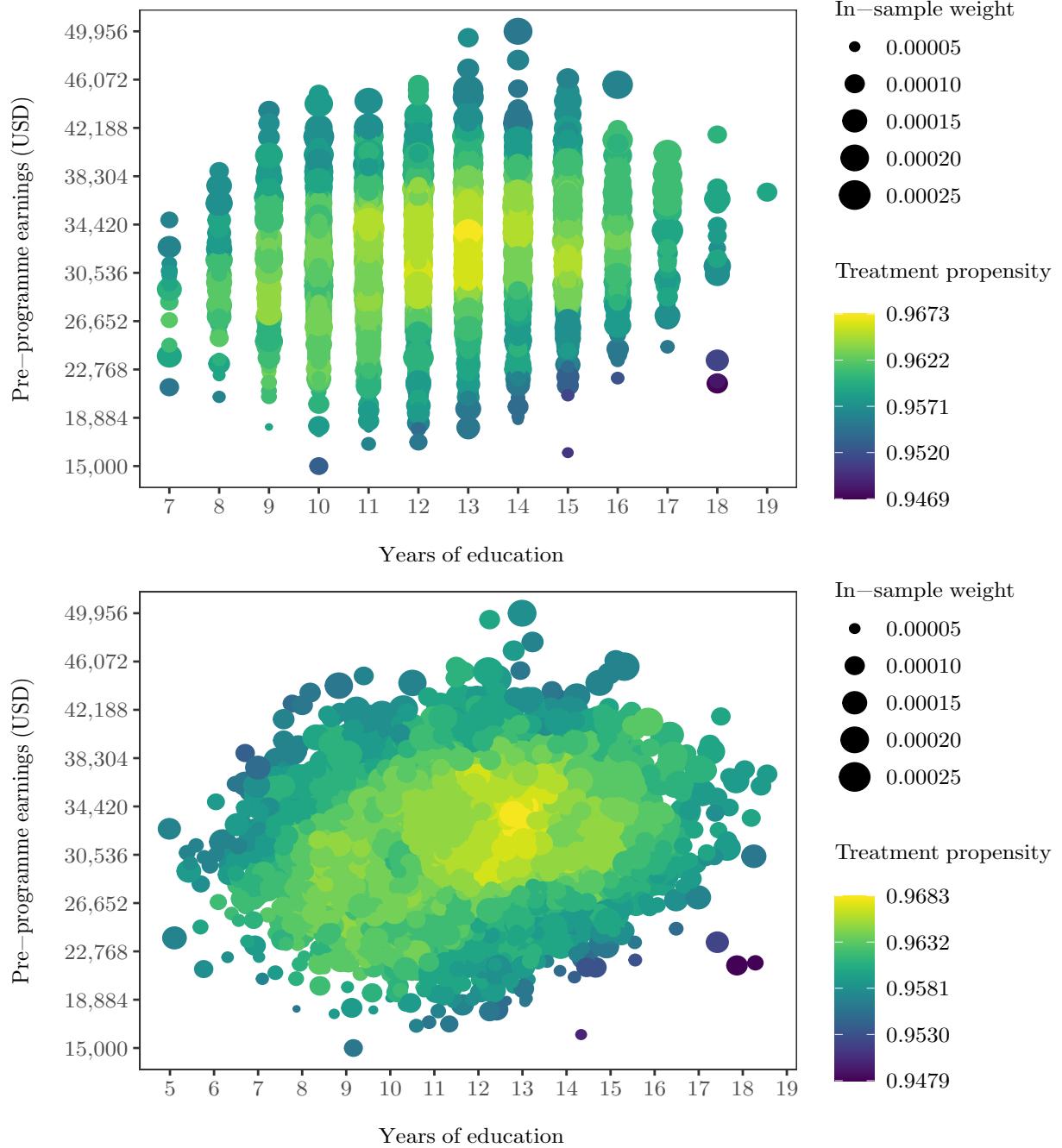
Notes: Simulated data generated using Equation (C.10) in conjunction with data from the JTPA Study sample. This figure illustrates the treatment propensity of individuals in a simulated sample under the posterior assignment rule that is induced by (κ^*, μ^*) . Each point represents the individual characteristics of an individual or several individuals. For comparison, individuals to the left of the solid diagonal line are assigned treatment under the oracle assignment rule of Equation (C.11).

Figure 15: (Experiments 4–6) Behaviour of the objective function at μ^* given variation in κ



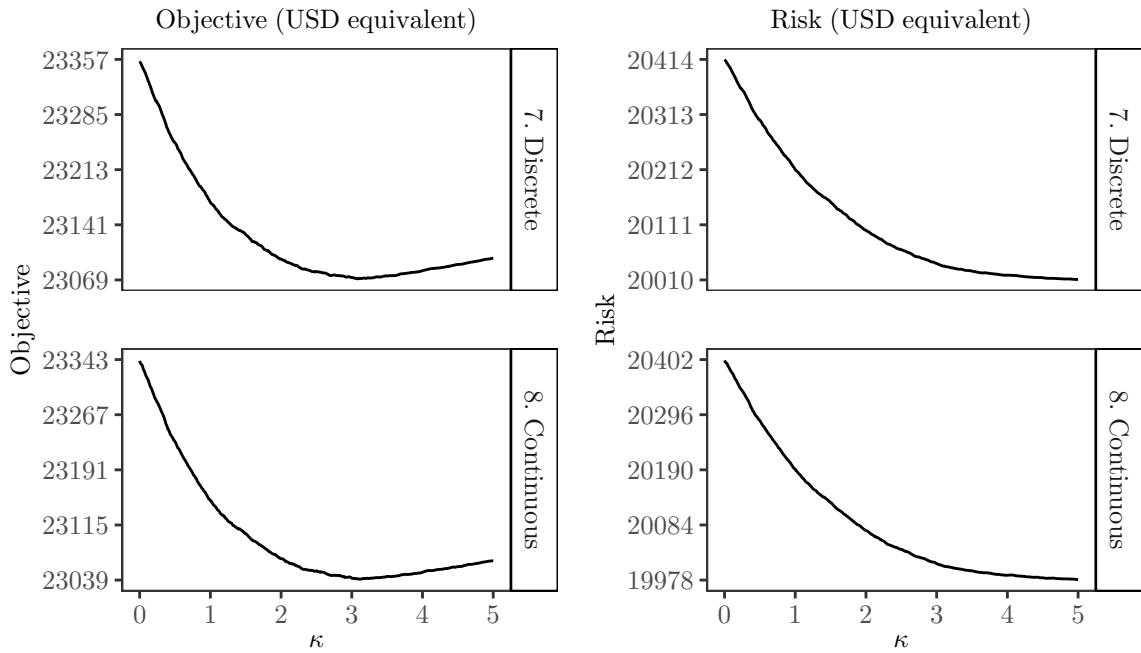
Notes: Simulated data generated using Equation (C.10) in conjunction with data from the JTPA Study sample. This figure illustrates how the shape of the objective function and its risk component changes as the variance of post-programme earnings is increased (middle panel) and decreased (lower panel). We hold μ fixed at μ^* in each case, which differs according to the sample size, and vary κ .

Figure 16: (Experiments 7–8) Variation in treatment propensity across individuals



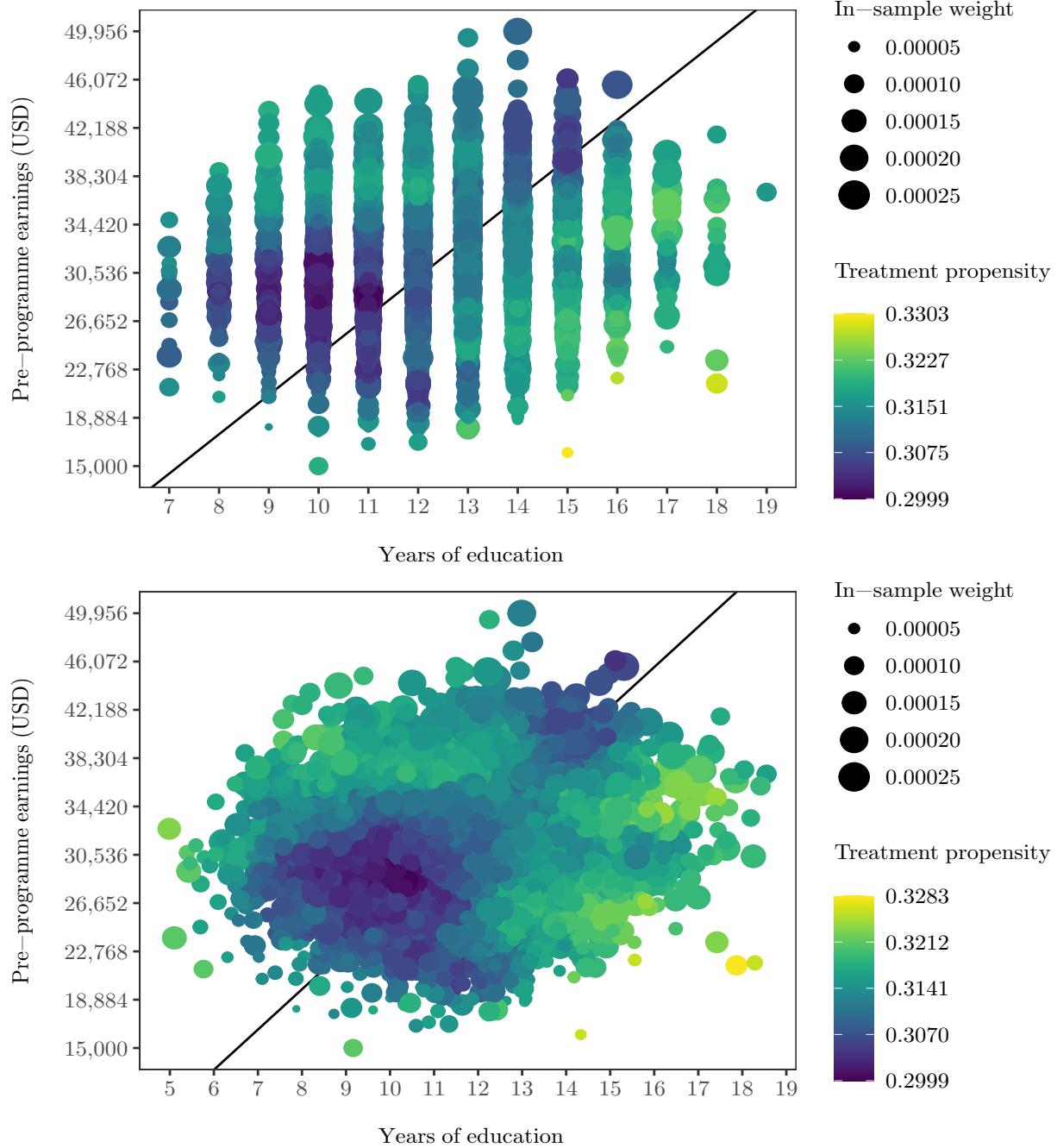
Notes: Simulated data generated using Equation (C.8) and a bivariate normal distribution. This figure illustrates the treatment propensity of individuals in simulated samples under the posterior assignment rule that is induced by (κ^*, μ^*) . Each point represents the individual characteristics of an individual or several individuals. Every individual is assigned to treatment under the oracle assignment rule.

Figure 17: (Experiments 7–8) Behaviour of the objective function at μ^* given variation in κ



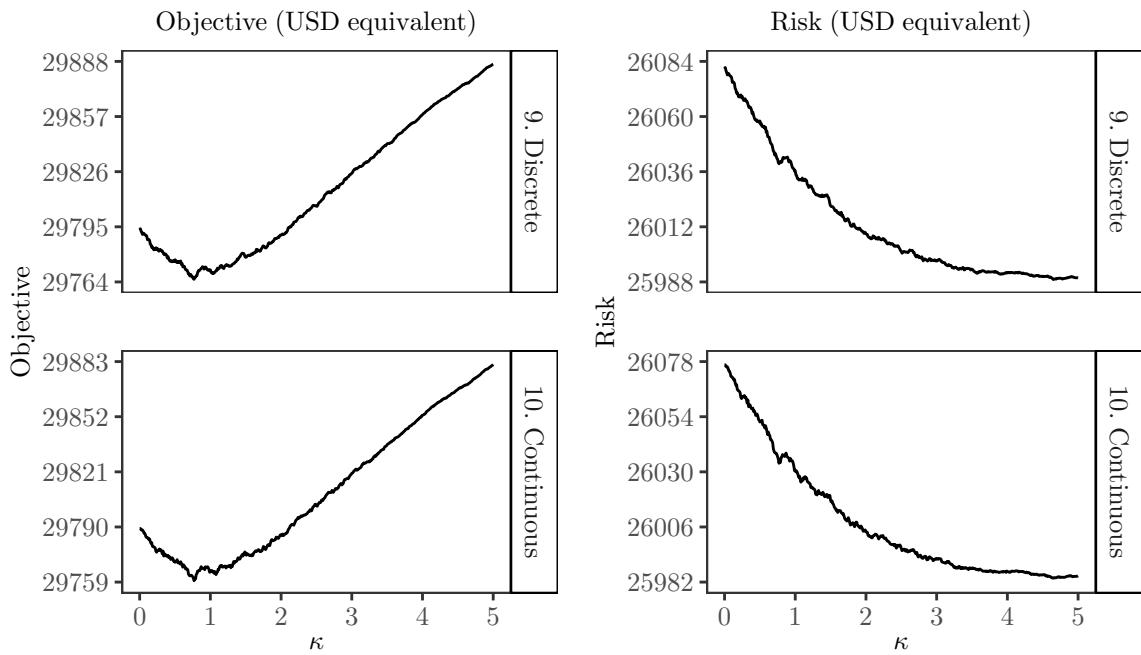
Notes: Simulated data generated using Equation (C.8) and a bivariate normal distribution. This figure illustrates how the shape of the objective function and its risk component at μ^* changes as κ increases.

Figure 18: (Experiments 9–10) Variation in treatment propensity across individuals



Notes: Simulated data generated using Equation (C.12) and a bivariate normal distribution. This figure illustrates the treatment propensity of individuals in simulated samples under the posterior assignment rule that is induced by (κ^*, μ^*) . Each point represents the individual characteristics of an individual or several individuals. For comparison, individuals to the left of the solid diagonal line are assigned treatment under the oracle assignment rule, which partitions the sample into two approximately equal-sized groups.

Figure 19: (Experiments 9–10) Behaviour of the objective function at μ^* given variation in κ



Notes: Simulated data generated using Equation (C.12) and a bivariate normal distribution. This figure illustrates how the shape of the objective function and its risk component at μ^* changes as κ increases.