cemmap
centre for microdata methods and practice

Policy learning with confidence

Victor Chernozhukov
Sokbae Lee
Adam M. Rosen
Liyang Sun

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP15/25

Economic
and Social
Research Council

POLICY LEARNING WITH CONFIDENCE*

VICTOR CHERNOZHUKOV†, SOKBAE LEE‡, ADAM M. ROSEN§, AND LIYANG SUN¶

Abstract. This paper introduces a framework for selecting policies that maximize expected welfare under estimation uncertainty. The proposed method explicitly balances the size of the estimated welfare against the uncertainty inherent in its estimation, ensuring that chosen policies meet a reporting guarantee, namely, that actual welfare is guaranteed not to fall below the reported estimate with a pre-specified confidence level. We produce the efficient decision frontier, describing policies that offer maximum estimated welfare for a given acceptable level of estimation risk. We apply this approach to a variety of settings, including the selection of policy rules that allocate individuals to treatments and the allocation of limited budgets among competing social programs.

Keywords: budget allocation, risk-aware policy learning, statistical decision theory

JEL classification codes: C14, C44, C52.

Date: July 2025.

*We thank Isaiah Andrews, Nathan Hendren, Matt Masten, Whitney Newey, Vira Semenova, Jesse Shapiro, Aleksey Tetenov and conference and seminar participants at the Munich Econometrics Workshop, Duke University, CUNY, University of Geneva, Tilburg University, the Encounters in Econometric Theory workshop, CUHK Shenzhen, and IAAE for helpful comments and discussion. Tian Xie provided excellent research assistance. Liyang Sun acknowledges generous support from the Economic and Social Research Council (new investigator grant UKRI607).

†Department of Economics, MIT.

‡Department of Economics, Columbia University.

§Department of Economics, Duke University.

¶Department of Economics, UCL and CEMFI. Email: liyang.sun@ucl.ac.uk.

1. Introduction

Consider a decision-maker (DM) faced with choosing from a menu of policies to maximize expected welfare. The DM could for example be a planner deciding how to allocate treatments to heterogeneous individuals, a firm choosing different potential innovations in which to invest resources, an auctioneer choosing an optimal auction design or reserve price, or a policy-maker choosing expenditure shares to allocate to government programs to maximize a measure of welfare to the public. If the DM knew the welfare that would be obtained from implementing each policy, the choice would be easy; she would then simply choose the policy that would yield the highest welfare. However, the DM lacks such precise knowledge, but instead has data that can be used to consistently estimate the welfare that would be achieved by each policy. The estimates are measured with varying degrees of precision, as reflected by their standard errors. How should the DM use the available information for policy selection?

An intuitively appealing choice is to select the policy with the highest estimated welfare, the so-called plug-in rule, which simply replaces the population objective with its sample analog in determining policy choice. In the literature on statistical treatment rules, this is referred to as empirical welfare maximization (EWM).[1] The same logical task applies more broadly to a variety of economic contexts, for example when researchers use structural models, they obtain estimates of the performance of different policies that can then be used to inform policy choice. Manski (2021) calls this "as-if" optimization, which he describes as, "specification of a model, point estimation of its parameters, and use of the point estimate to make a decision that would be optimal if the estimate were accurate." Here we refer to such rules interchangeably as plug-in rules or empirical welfare maximization (EWM) rules.

Unfortunately, the available estimates for the welfare of different choices are generally not identical to their population values. Estimation error could result in the EWM rule delivering suboptimal policy choice. It may be observed, for example, that among a menu of options policy A has the highest estimated welfare with policy B coming in a close second, but that the standard error of policy B is much smaller than that of policy A, reflecting that the welfare of policy B is estimated much more precisely. Could it be that policy B is actually better than policy A, and that the higher estimated welfare of policy A is simply down to a lack of precision in the estimates? Should the DM consider choosing policy B since it is estimated more precisely to hedge against estimation error?

[1]See for example Kitagawa and Tetenov (2018) and Athey and Wager (2021). Andrews, Kitagawa and McCloskey (2024) alternatively refers to this as the "natural rule" or "picking the winners".

This paper proposes a framework for selecting policies with the goal of maximizing expected welfare in the presence of estimation uncertainty, explicitly accounting for estimation risk in the decision rule. As inputs to her decision the DM has consistent estimates for the welfare of each available policy, and their corresponding sample variance matrix, with sample estimates approximately normally distributed around their population means.[2] We define and analyze a class of risk-aware policy rules that provide a principled manner for explicitly balancing the estimated welfare of each policy against its estimation risk. Policies that perform well in sample but are estimated imprecisely may be passed over in favor of less risky policies that are not optimal according to the EWM rule but that can be more confidently assured to achieve high performance. Risk-aware rules select a policy from the (Pareto) efficient decision frontier that balances performance and precision as measured by trading off higher sample estimates against smaller sample variance, where the exact tradeoff is governed by the DM's choice of policy rule from the class of risk-aware rules. Consequently, such rules select the best performing policy among those within the DM's choice of acceptable estimation error. They can alternatively be interpreted as selection rules obtained by a risk-averse Bayesian DM who minimizes an expected loss function that takes a weighted average of regret and estimation error.[3] We establish favorable regret properties for this class of rules, as well as conditions under which they deliver a reporting guarantee, ensuring with high confidence that the actual welfare exceeds a lower threshold.

To illustrate, consider a setting in which a budget-conscious DM needs to allocate funds across several social programs that affect the same group of individuals, either in response to a slight budget increase or for incremental cost-cutting. Suppose the DM wishes to use the Marginal Value of Public Funds (MVPF), spear-headed by Hendren and Sprung-Keyser (2020), for this purpose. The MVPF of each social program is the ratio of its marginal benefit to its net marginal cost to the government, inclusive of the impact of any behavioral responses on the government budget. For instance, an MVPF of 1.5 indicates that each \$1 of net government spending generates \$1.50 in benefits for beneficiaries. For a utilitarian DM, the MVPF provides a metric to compare the "bang for the buck" of different policies that target the same beneficiaries. A DM interested in allocating additional marginal expenditure will prioritize programs with the highest MVPF to maximize welfare.[4]

[2]If the estimates are from independent samples the sample variance matrix will simply be a diagonal matrix with each policy's sample variance along the diagonal.

[3]This is shown in Section 2.5. Such dual-objective loss functions have been explored in the statistics literature in e.g. Gupta and Miescke (1986), Gupta and Miescke (1990), and see also Liese and Miescke (2008), as a way to unify the problems of selection and estimation when sampling from independent normal populations.

[4]Here we abstract from distributional incidence and assume the DM values dollars equally across the beneficiaries of each policy. If welfare weights vary across beneficiaries, as in the original framework of Hendren and Sprung-Keyser (2020), our analysis can accommodate this by adjusting each MVPF in the value function

In practice the DM must make decisions based on estimated MVPFs. The Policy Impacts Library, Hendren and Sprung-Keyser (2021), provides easily accessible MVPF estimates online while also noting significant estimation uncertainty in some cases. Using the EWM rule to allocate funds would direct the entire budget to the program with the highest MVPF estimate. However, the DM may be reluctant to direct the entire budget to a single social program if this MVPF estimate has a large standard error relative to other programs whose MVPF estimates are nearly as high, and which have smaller standard errors. How should the DM balance the tradeoff between the size of the estimated MVPF and the precision with which it is estimated?

(A) Programs targeting age > 25 (B) Programs targeting age ≤ 25

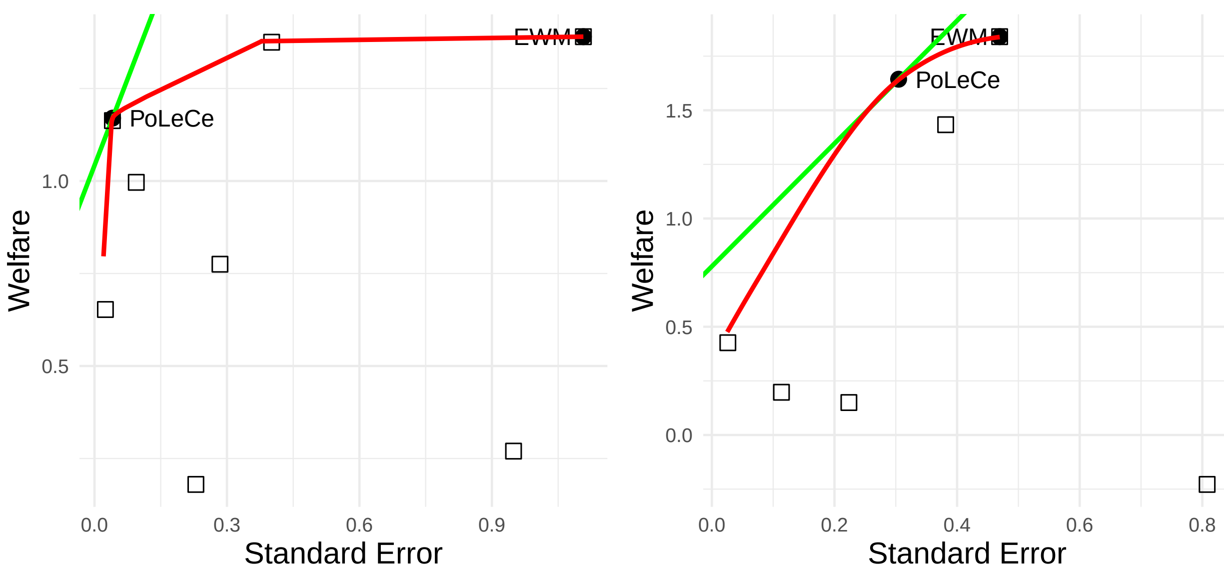


FIGURE 1. Efficient Decision Frontier.

Notes: Hollow squares represent MVPF estimates and their standard errors for 14 programs collected by the Policy Impact Library. The red curve shows the decision frontier for budget allocation across these programs. The solid square and circle mark allocations selected by EWM and PoLeCe, respectively. For programs targeting age > 25, the programs included in the PoLeCe allocation with large shares, listed in order of their shares, are: Housing Vouchers in Chicago (0.65 ± 0.03), OHIE, Single Adults (1.16 ± 0.04), where each value represents the point estimate ($\pm$ standard error). The program selected by EWM is Holistic Wrap-around Services Can Improve Employment Rates (1.39 ± 1.11). For programs targeting age ≤ 25, the programs included in the PoLeCe allocation are: Head Start (1.84 ± 0.47) and Wisconsin Scholars Grant (1.43 ± 0.38). The program selected by EWM is Head Start (1.84 ± 0.47).

using the DM's designated weights. Example welfare weights have been derived in Hendren (2020). Other metrics for measuring program's marginal benefit and cost have also been discussed in García and Heckman (2022).

This tradeoff is illustrated in Figure 1 using the MVPF estimates from the Policy Impacts Library, which comprises 14 US programs after we focus on estimates based on randomized control trials (RCTs) with finite standard errors. To mimic a DM who applies constant welfare weights within certain age groups but not necessarily across them due to distributional concerns, we use the Policy Impacts Library classification of programs based on the intended beneficiaries' age. We then consider budget allocation separately for two distinct age groups. Six of these programs target beneficiaries below age 25, covering early childhood education programs, college financial aid, and job training.[5] Eight of these programs target beneficiaries above age 25, covering federal social assistance programs, health insurance, and housing vouchers. In both cases, the MVPF of the programs selected by the EWM rule are not estimated as precisely as the welfare delivered by diversifying the budget across several programs whose MVPFs are estimated with significantly higher precision. If a DM values both welfare and precision, how should these goals be balanced?

We propose *Policy Learning with Confidence*, or PoLeCe, among the class of risk-aware policy rules that lie on the efficient decision frontier. Like all risk-aware rules, PoLeCe uses observable measures of statistical variation to adjust policy choice relative to a simple plug-in rule by penalizing the observed performance of each policy by an amount proportional to its standard error. Specifically PoLeCe is a risk-aware decision rule that uses a data-dependent penalization factor so that the resulting objective value automatically provides the lower bound of a one-sided confidence interval for the welfare achieved by the selected rule. Thus, the rule delivers a reporting guarantee, ensuring with high confidence that the actual welfare exceeds a lower threshold; no adjustment for post-selection inference is required. Importantly, policies that allocate fractional shares to different treatments or programs can be allowed, producing a richer frontier than that available from singleton allocations such as the individual programs indicated by hollow squares in Figure 1.

In the above example, all that was required to map out the efficient decision frontier, and indeed all that is needed to determine the optimal allocation following our proposal, are point estimates for the welfare of each available policy and their joint variance-covariance.[6] This can be obtained in applications that feature a variety of different models and sampling

[5]As explained in the Policy Impacts Library, alternative specifications have led to larger MVPF estimates for early childhood education programs depending on how lifetime earnings are forecasted and whether one accounts for the transfer value of preschool subsidies to parents.

[6]Our proposed PoLeCe rule uses the sample correlation of the estimates to calibrate the precision penalty to achieve the aforementioned reporting guarantee. The general class of risk-aware decision rules described in Section 2.1 allows for different choices for the penalization factor, which corresponds to the green slope highlighted in Figure 1. Other risk-aware rules may thus only require standard errors rather than estimates of the entire correlation structure, depending on their choice of penalization factor.

processes. An important special case to which our framework applies is the analysis of optimal treatment assignment using data on individual-specific treatments and allocations, such as from an RCT. For example, in a real-world setting of treatment assignment, while GiveWell uses EWM based on RCT results, they have recently highlighted the importance of accounting for estimation uncertainty in decision-making for the sake of "transparency" (Salisbury, 2024). Our proposal offers a practical solution to this concern.

The analysis of statistical treatment rules in such settings has received renewed attention in the recent literature, starting with the pioneering work of Manski (2004), who considers treatment choice on the basis of minimax (expected) regret. Follow-up works include a decision theoretic framework introduced by Dehejia (2005), finite-sample bounds considered by Stoye (2009), and an asymptotic framework introduced by Hirano and Porter (2009, 2020). Hypothesis testing is a conventional approach to decision-making under uncertainty. As argued by Tetenov (2012), it can be interpreted as a DM having asymmetric aversion to Type I and Type II regret.

Kitagawa and Tetenov (2018) proposes empirical welfare maximization (EWM) rules for experimental data, which is an "as-if" optimization rule that selects a policy with the highest estimated welfare (empirical welfare), thus constituting a plug-in rule for treatment effect models with experimental data. Kitagawa and Tetenov (2018) proves the optimality of EWM rules in the sense that as the sample size increases expected regret converges to zero at the minimax rate. Athey and Wager (2021) and Demirer, Syrgkanis, Lewis and Chernozhukov (2019) propose doubly-robust estimation of average welfare, in which case this result obtains even with quasi-experimental data. In the context of this literature, PoLeCe offers a modification of EWM that incorporates estimation risk into the decision rule. The EWM rule corresponds to the PoLeCe rule with penalization factor of zero on the standard error, putting no weight on estimation risk. Inference approaches for optimal treatment assignment and/or welfare complementary to EWM, but which do not propose alternative assignment rules to account for uncertainty, include those of Armstrong and Shen (2023), Ponomarev and Semenova (2024), and Rai (2019), as well as the online supplement of the original paper by Kitagawa and Tetenov (2018). As an extension of EWM, Mbakop and Tabord-Meehan (2021) proposes a penalized welfare maximization rule that penalizes for the complexity of the policy, and establishes an oracle property for model selection.

Andrews and Chen (2025) proposes the use of certified decisions, which comprise the pairing of a decision rule with an accompanying inferential guarantee. Their results provide a decision theoretic justification for the PoLeCe rule, showing that it belongs to an essentially complete class of decision rules for a DM who requires a coverage guarantee, and therefore dominates the pairing of EWM with the studentized projection approach in Andrews et al. (2024).

PoLeCe also contributes to a recent literature featuring some alternative proposals to directly account for estimation uncertainty in the decision rule. In settings with a budget constraint, Sun (2021) proposes a rule that penalizes policies with large uncertainty in meeting the constraint, and Moon (2025) proposes using empirical Bayes estimates, leveraging shared information across related policies. Kitagawa, Lee and Qiu (2022) focuses on the variability of regret and shows that the EWM rule is dominated by fractional rules when the objective function is the sum of squared expected regret and the variance of regret. If the estimated welfare is independent across policies, then one may appeal to the literature on optimal selection such as Gupta and Miescke (1986), referenced above. However, in the context of policy learning, since the groups assigned by different policies might not be mutually exclusive, the estimated welfare associated with different policies may not be independent. For instance, when treatment allocation is determined by an income threshold, the welfare estimates for groups falling below different thresholds will be correlated by construction.

The paper proceeds as follows. Section 2 sets out our approach to risk-aware policy choice that selects policies from the estimated efficient decision frontier. Section 2.1 motivates the approach using selection of policies for treatment choice and investment allocation in public programs as key illustrative examples. Section 2.2 develops favorable regret and lower confidence guarantees for the general class of risk-aware policy rules, and Section 2.3 formally sets out the proposed PoLeCe rule as the risk-aware rule that maximizes a lower confidence bound on maximal welfare. Section 2.4 provides evidence from Monte Carlo experiments calibrated to empirical applications. Section 2.5 provides alternative motivations for risk-aware decisions. Section 2.6 discusses refinements and further theoretical points. Section 3 provides four empirical applications that demonstrate the application of PoLeCe, and Section 4 concludes.

2. Risk-Aware Policy Choices

2.1. **Motivation and Risk-Aware Policy Choice.** We begin with the following key problem as motivation.

Problem of Choosing Better Treatment Policies. Given a finite treatment set $\mathcal{T}$ and potential outcomes $Y(t)$ for each treatment t, a policy $\pi \in \Pi$ specifies probabilities $\pi(t \mid X)$ of assigning treatment t given covariates X, where Π denotes the set of allowable policies. For example, if any fractional allocation is allowed Π is the unit simplex, while if treatment assignment is required to be a deterministic function of covariates then Π is the set of vertices of the unit simplex. Other constraints can also be accommodated by appropriate specification of Π.

The welfare function of a policy is then defined as

$$V(\pi) \;=\; \mathrm{E}\Big[\sum_{t\in\mathcal{T}} Y(t)\,\pi\big(t \mid X\big)\Big],$$

which captures the expected outcome under policy π.

We assume we have data $(W_i)_{i=1}^n$ in the form of i.i.d. copies of $W = (Y, T, X)$, where T is the treatment, X are covariates, and Y is the outcome. We also assume the data were collected under the unconfoundedness assumption, namely that the assigned treatment T is independent of the potential outcome $Y(t)$ given X. This implies that $\mathrm{E}[Y(t) \mid X]$ is identified from the regression

$$g(t, X) = \mathrm{E}[Y \mid T = t, X],$$

provided that $p(t \mid X) = \Pr[T = t \mid X] > 0$.

Let $p(t \mid X) = \Pr[T = t \mid X]$ be the propensity score, and define the Riesz representer

$$H(\pi) \;=\; \sum_{t\in\mathcal{T}} \mathbf{1}\{T = t\}\,\frac{\pi(t \mid X)}{p(t \mid X)}.$$

Under standard conditions, including $\mathrm{E}[H^2(\pi)] < \infty$ – a relevant overlap condition[7] – we obtain the following regression and representer formulas for identifying the policy's welfare:

$$V(\pi) \;=\; \mathrm{E}\left[\sum_{t\in\mathcal{T}} g(t, X)\,\pi(t \mid X)\right] \;=\; \mathrm{E}\big[Y\,H(\pi)\big], \tag{1}$$

which can be viewed either as aggregating up regressions or as reweighting observed outcomes by how the policy π differs from the propensity score p. Combining these two approaches leads to efficient representation and estimation of the welfare function via the influence function (e.g., Newey, 1994):

$$V(\pi) \;=\; \mathrm{E}[\psi_\pi(W)], \quad \text{where} \quad \psi_\pi(W) \;:=\; \sum_{t\in\mathcal{T}} g(t, X)\,\pi(t \mid X) \;+\; H(\pi)\,\big[Y - g(T, X)\big],$$

also known as the doubly-robust score in a parametric estimation context.

Added Value of Policies vs. Control Policy. It is often the case that we want to make decisions depending upon whether policies improve upon a control treatment. Let us denote

[7]This requires that policies do not place weight on treatments for which there is insufficient randomization at particular values of X. Note that the standard overlap condition can fail, but the welfare of a particular policy is identified, for example, if $H(\pi)$ and Y have a finite second moment.

by $t = 0$ the control state. In this case, we can re-define:

$$V(\pi) \; = \; \mathrm{E}\Big[\sum_{t \in \mathcal{T}} \{Y(t) - Y(0)\}\, \pi\big(t \mid X\big)\Big]. \tag{2}$$

The efficient score for estimation and inference for $V(\pi)$ is now

$$\psi_\pi(W) := \sum_{t \in \mathcal{T}} (g(t, X) - g(0, X))\pi(t \mid X) + H(\pi)\, \big[Y - g(T, X)\big] \tag{3}$$

where we re-define $H(\pi) = \sum_{t \in \mathcal{T}} \left[\frac{\mathbf{1}\{T=t\}}{p(t|X)} - \frac{\mathbf{1}\{T=0\}}{p(0|X)}\right] \pi(t \mid X)$.

We follow the convention of encapsulating in X both covariates that are confounders as well as those for which the planner may assign different treatments.

In randomized experiments, the representers are known, and we only need consistent cross-fitted learners for the regression functions to obtain consistent, asymptotically normal, efficient estimators $\widehat{V}(\pi)$.[8] In observational studies, one can employ cross-fitted, modern statistical or machine learning methods to estimate the Riesz representers and the regressions. The resulting estimators satisfy

$$\widehat{V}(\pi) = V(\pi) + \mathbb{E}_n\big(\psi_\pi(W_i) - V(\pi)\big) + o_p\big(1/\sqrt{n}\big),$$

uniformly over $\pi \in \Pi$, where Π is either a finite set or a set whose complexity does not grow too fast. By further applying high-dimensional central limit theorems, we obtain

$$\{\widehat{V}(\pi)\} \;\overset{a}{\sim}\; \{V(\pi) + s(\pi)\, Z_\pi\}, \quad Z_\pi \sim N(0, 1), \quad s^2(\pi) = \mathrm{Var}(\psi_\pi)/n,$$

where $\{Z_\pi\}_{\pi \in \Pi}$ is a centered Gaussian process, with each component having standard normal distribution; and $\overset{a}{\sim}$ means "approximately distributed as" in the sense stated more formally below.

Risk-Aware Policy Choices. Different treatment policies π can have markedly different welfare and estimation risk profiles, and some may be much riskier than others. For example, policies that "turn on" treatments with near-zero propensities may be especially risky, since

$$s^2(\pi) \propto \mathrm{E}[H^2(\pi)]/n$$

can be large in such cases. What should a decision maker do?

The decision maker has to choose an optimal policy in Π based on data $\widehat{V}(\pi)$ that is (approximately) Gaussian. We argue that an optimal risk-aware decision can be achieved

[8]With randomized experimental data, the pseudo-consistency of a regression estimator for some function (though not necessarily the true function) is sufficient to satisfy all desirable properties except for asymptotic efficiency.

by maximizing the empirical welfare *offset* by the estimation risk times a critical value:

$$\widehat{\pi}_{\mathrm{RW}} \in \arg\max\Big\{\widehat{V}(\pi) - k\,\widehat{s}(\pi) : \pi \in \Pi\Big\}, \tag{4}$$

where k is a critical value and $\widehat{s}(\pi)$ is a consistent estimator of the estimation risk $s(\pi)$. By varying k, we trace out the *efficient decision frontier* of treatment policies, as illustrated by Figure 1 in the introduction. Each point on the frontier corresponds to a particular $k > 0$. We provide fast algorithms in Appendix B to efficiently compute the frontier for any application.

Our leading proposal to choose k is to meet certain reporting guarantees with high confidence, thereby conducting "policy learning with confidence" (PoLeCe).[9] The resulting choice generally differs from the empirical welfare maximizer (EWM):

$$\widehat{\pi}_{\mathrm{EWM}} \in \arg\max\{\widehat{V}(\pi) : \pi \in \Pi\}. \tag{5}$$

The EWM rule arises when $k \searrow 0$. We provide a comparative analysis of these two rules in what follows. Before we present details of our PoLeCe proposal, we discuss another class of applications.

The Problem of Choosing Investment Allocations in Public Programs. A public or private agency is considering programs $\{1, ..., J\}$ and must decide on investment shares

$$\pi = (\pi_1, ..., \pi_J),$$

subject to

$$\Pi = \Big\{ (\pi_1, \ldots, \pi_J) : \sum_{j=1}^{J} \pi_j = 1,\ 0 \le a_j \le \pi_j \le b_j \le 1 \Big\},$$

which is the unit simplex intersected with a rectangular set of constraints. These additional constraints may reflect diversity or other requirements that allocations must satisfy. The welfare of the allocation π is

$$V(\pi) = \pi' R,$$

where R is a vector of the rate of return measures, for instance the ratios of marginal benefit to net cost of funds used in public finance. The agency is given estimates $\widehat{R}$ that are approximately Gaussian[10]

$$\widehat{R} \overset{a}{\sim} N(R, \Omega/n),$$

[9]Other choices of k are of course possible and can be rationalized as optimal under risk aversion. We discuss those later in Section 2.5.

[10]In the empirical application we consider, Ω is block-diagonal, and n is the notional sample size used to study the behavior of rules as more information is acquired.

so that the estimated welfare of the allocations are also approximately Gaussian:

$$\widehat{V}(\pi) = \pi'\widehat{R} \overset{a}{\sim} \pi' R + s(\pi) Z_\pi, \quad Z_\pi \sim N(0,1), \quad s^2(\pi) = \pi'\Omega\,\pi/n,$$

where $\{Z_\pi\}_{\pi\in\Pi}$ is a Gaussian vector with standard normal marginal distributions.

Hence, for each fund allocation π, the agency is given $\widehat{V}(\pi)$ along with its associated estimation risk $s(\pi)$. What should the agency do? Our proposal is again to use the risk-aware rule (4). If estimation risk is ignored, one might rely on the empirical welfare maximizer, but in the absence of additional constraints, that rule would simply allocate all funds to the program with the highest estimated return. This is generally unappealing on intuitive grounds. In contrast, the risk-aware approach (4) would spread out funds in an "efficient" portfolio of programs, similar to Markowitz portfolio allocation in finance. However, the motivation and formulation of our approach are distinct from the Markowitz model.

2.2. **Regret Properties and Guarantees of Any Risk-Aware Policy.** We now provide regret bounds for any risk-aware decision rule $\widehat{\pi}_{\mathrm{RW}}(\widehat{k})$ that solves the risk-adjusted empirical welfare problem:

$$\widehat{\pi}_{\mathrm{RW}}(\widehat{k}) \;\in\; \arg\max_{\pi\in\Pi}\big\{\widehat{V}(\pi) \;-\; \widehat{k}\,\widehat{s}(\pi)\big\},$$

where $\widehat{k} \geq 0$ may depend on both the decision maker's preferences and the data. Note that $\widehat{k} = 0$ corresponds to the EWM rule.

We focus on the case where Π has finite cardinality $p = p_n$, which can be very large, potentially much larger than n. (A theoretical treatment for infinite Π follows by suitable discretization, as we comment on later.)

Scaled Estimation Error and Gaussian Approximation. Define

$$\widehat{Z}_\pi \;:=\; \frac{\widehat{V}(\pi) - V(\pi)}{\widehat{s}(\pi)}$$

to be the scaled estimation error process. We use the following Gaussian approximation condition, denoted by (G). Let $\mathcal{A}$ represent the collection of rectangular sets in $\mathbb{R}^p$.

(G) There exists a sequence of non-negative constants r_n with $r_n \searrow 0$ such that,

$$\Big|\Pr\big((\widehat{Z}_\pi)_{\pi\in\Pi} \in A\big) \;-\; \Pr\big((Z_\pi)_{\pi\in\Pi} \in A\big)\Big| \;\leq\; r_n, \quad \text{for all } A \subset \mathcal{A}. \tag{G}$$

These conditions are known to hold under mild assumptions.[11] Furthermore, (G) implicitly requires consistency of risk estimates: $\max_{\pi\in\Pi}|\widehat{s}(\pi)/s(\pi) - 1| \to_P 0$.

Quantiles of the Maximal Estimation Error. For a subset $K \subseteq \Pi$, let

$$q_{1-\beta,K} \;:=\; (1-\beta)\text{-Quantile}\Big(\sup_{\pi\in K} Z_\pi\Big)$$

denote the $(1-\beta)$-quantile of the estimation error over K under the Gaussian approximation. Under (G), with probability at least $1-\beta-r_n$, $\max_{\pi\in K}\widehat{Z}_\pi \le q_{1-\beta,K}$. Let $V_{\max} = \sup_{\pi\in\Pi} V(\pi)$ be the maximal true welfare and $\Pi_0 = \{\pi \in \Pi : V(\pi) = V_{\max}\}$ be the set of policies with maximal true welfare, and denote its cardinality by $p_0 = |\Pi_0|$. Define

$$\bar{\sigma}_\Pi \;:=\; \sqrt{n}\max_{\pi\in\Pi}\widehat{s}(\pi), \quad \underline{\sigma}_{\Pi_0} \;:=\; \sqrt{n}\min_{\pi\in\Pi_0}\widehat{s}(\pi).$$

These represent, respectively, an upper bound on the estimation risk of all policies in Π and a lower bound on the estimation risk of the best policies Π_0.

Proposition 1 (Regret Bounds for Risk-Aware Decisions)**.** *Under condition* (G)*, the regret of any RW policy selection rule* $\widehat{\pi}_{RW}(\widehat{k})$ *is bounded with probability at least* $1-2\beta-2r_n$ *as:*

$$V_{\max} - V(\widehat{\pi}_{RW}(\widehat{k})) \;\le\; \frac{\underline{\sigma}_{\Pi_0}}{\sqrt{n}}\Big\{q_{1-\beta,\Pi_0} + \widehat{k}\Big\} \;+\; \frac{\bar{\sigma}_\Pi}{\sqrt{n}}\Big\{\big(q_{1-\beta,\Pi} - \widehat{k}\big)_+\Big\}.$$

In particular, if $0 \le \widehat{k} \lesssim q_{1-\beta,\Pi}$*, then with the same probability,*

$$V_{\max} - V(\widehat{\pi}_{RW}(\widehat{k})) \;\lesssim\; \big(\bar{\sigma}_\Pi/\sqrt{n}\big)\, q_{1-\beta,\Pi}.$$

Moreover, if $q_{1-\beta,\Pi} \le \widehat{k} \lesssim q_{1-\beta,\Pi}$*, then with the same probability,*

$$V_{\max} - V(\widehat{\pi}_{RW}(\widehat{k})) \;\lesssim\; \big(\underline{\sigma}_{\Pi_0}/\sqrt{n}\big)\, q_{1-\beta,\Pi}.$$

As β decreases, the probability that the regret bound holds increases, since the failure probability is at most $2\beta + 2r_n$. However, this comes at the cost of a looser bound because the quantile $q_{1-\beta,\Pi}$ increases as β gets smaller. This reflects a basic trade-off: a smaller β provides a more conservative bound that holds with higher probability, while a larger β results in a tighter bound that is less certain to be valid. This behavior is analogous to confidence intervals, where greater confidence requires a wider interval.

[11] See Chernozhukov, Chetverikov, Kato and Koike (2022); Chernozhukov, Chetverikov and Koike (2023a) for sharp forms of Gaussian approximations, and Chernozhukov, Chetverikov, Kato and Koike (2023b) for a review. Additional references, e.g. Belloni, Chernozhukov, Chetverikov and Wei (2018); Quintas-Martinez (2022) discuss Gaussian approximations for many or a continuum of target parameters, including those learned via debiased machine learning.

Proposition 1 provides dimension-free bounds. To obtain simpler dimension-based bounds, observe that by standard concentration inequalities,[12]

$$q_{1-\beta,K} \;\leq\; \mathrm{E}\Big(\sup_{\pi \in K} Z_\pi\Big) \;+\; \sqrt{2\log(1/\beta)} \;\leq\; \sqrt{2\log|K|} \;+\; \sqrt{2\log(1/\beta)} \;=\; u_{1-\beta,|K|}.$$

We now see that the broad class of RW-rules with limited risk aversion has regret bounded by

$$\bar{\sigma}_\Pi \sqrt{\frac{\log p}{n}},$$

as $p \to \infty$ (and $n \to \infty$). This class includes the empirical welfare maximization rule ($k = 0$) and matches known optimal minimax rates when available (see Athey and Wager (2021)). Additionally, choosing $\widehat{k} > 0$ sufficiently large can yield a regret bound of

$$\underline{\sigma}_{\Pi_0} \sqrt{\frac{\log p}{n}},$$

which has the same rate but may have a *much smaller multiplicative constant* –

$$\underline{\sigma}_{\Pi_0} < \bar{\sigma}_\Pi$$

if there is significant heterogeneity in estimation risks across policies. The empirical risk minimization literature has also used penalties on estimation risk to improve error bounds, albeit in different contexts; see, for example, Maurer and Pontil (2009); Swaminathan and Joachims (2015); Foster and Syrgkanis (2023).

Moving Toward a Data-Driven $\widehat{k}$. These ideas naturally lead to the next section, where we propose a main policy rule that carefully constructs a data-driven $\widehat{k} \approx q_{1-\alpha,\Pi}$ to achieve small regret and provide additional reporting guarantees.

Reporting Guarantees of Risk-Aware Decisions. In practice, to obtain reporting guarantees for RW decisions, one needs to approximate the quantiles $q_{1-\alpha,\Pi}$. We do so via the bootstrap, where we set:

$$\widehat{q}_{1-\alpha,\Pi} \;:=\; (1-\alpha)\text{-Quantile}\Big(\max_{\pi \in \Pi} \widehat{Z}^*_\pi\Big),$$

where $\widehat{Z}^*_\pi := N\big(0, \widehat{C}\big)$, and $\widehat{C}$ is a consistent estimator of the covariance matrix $C = \mathrm{Cov}((Z_\pi)_{\pi\in\Pi})$. Denote by $\Pr^*$ the bootstrap-induced probability measure computed conditional on $\widehat{C}$. This construction relies on the following condition:

[12]We can also employ Talagrand's style upper bounds, but to make our points it suffices to use these simple concentration bounds.

(B) With probability at least $1-\delta_n$ (where $\delta_n \searrow 0$),

$$\left|\Pr^*\big((\widehat{Z}^*_\pi)_{\pi\in\Pi} \in A\big) \;-\; \Pr\big((Z_\pi)_{\pi\in\Pi} \in A\big)\right| \;\le\; r_n, \quad \text{for all } A \subset \mathcal{A}. \tag{B}$$

Like (G), condition (B) is satisfied under mild assumptions even when p is much larger than n, and has been verified for a variety of estimation methods, including debiased machine learning.[13]

A key consequence is that, letting $r'_n := 2r_n + \delta_n$, we have

$$\sup_{\pi\in\Pi} \widehat{Z}_\pi \;\le\; \widehat{q}_{1-\alpha,\Pi} \quad \text{with probability at least } 1-\alpha-r'_n, \tag{6}$$

(see Lemma 1 for a proof). Inequality (6) then directly implies a uniform lower confidence bound (LCB) on the performance of all policies:

Proposition 2 (LCB Guarantees for RW Decisions)**.** *With probability at least* $1-\alpha-r'_n$*, the true value of any policy* $\pi \in \Pi$ *is bounded below as follows:*

$$V(\pi) \;\ge\; LV_{1-\alpha}(\pi) \;:=\; \widehat{V}(\pi) \;-\; \widehat{q}_{1-\alpha,\Pi}\,\widehat{s}(\pi) \;\; \textit{for all } \pi \in \Pi.$$

2.3. **The PoLeCe Rule and Its Properties.** We now introduce *policy learning with confidence* (PoLeCe), a risk-aware rule that provides a key reporting guarantee by maximizing the LCB on policy welfare:

$$\widehat{\pi}_{\text{PoLeCe}} \;\in\; \arg\max_{\pi\in\Pi} LV_{1-\alpha}(\pi). \tag{7}$$

Notably, PoLeCe is an RW decision that uses

$$\widehat{k} = \widehat{q}_{1-\alpha,\Pi}$$

(the bootstrap quantile) as the penalty for estimation risk. Furthermore, the maximized LCB

$$LV_{1-\alpha,\Pi} \;:=\; \max_{\pi\in\Pi} LV_{1-\alpha}(\pi) \;=\; \max_{\pi\in\Pi}\Big\{\widehat{V}(\pi) \;-\; \widehat{q}_{1-\alpha,\Pi}\,\widehat{s}(\pi)\Big\}$$

serves as a natural performance guarantee.

Remark 1 (PoLeCe as Minimizer of the UCB on Regret)*.* We can also view $\widehat{\pi}_{\text{PoLeCe}}$ as minimizing an upper confidence bound on regret. Indeed, for any $\pi \in \Pi$,

$$\text{Regret}(\pi) \;=\; V_{\max} - V(\pi) \;\le\; V_{\max} \;-\; \big(\widehat{V}(\pi) - \widehat{k}\,\widehat{s}(\pi)\big)$$

[13]See Chernozhukov et al. (2022, 2023a) and Chernozhukov et al. (2023b) for more details, along with other references on "approximate means" settings such as debiased ML (Belloni et al. (2018); Quintas-Martinez (2022)).

with probability at least $1-\alpha-r_n'$. Hence, by its construction, $\widehat{\pi}_{\text{PoLeCe}}$ minimizes this UCB on regret:

$$\min_{\pi\in\Pi}\Big\{V_{\max} \;-\; \big(\widehat{V}(\pi)-\widehat{k}\,\widehat{s}(\pi)\big)\Big\}.$$

The following result establishes formal properties of the PoLeCe rule.

Proposition 3 (Reporting Guarantees and Regret Bounds for PoLeCe). *With probability at least* $1-\alpha-r_n'$,

$$V_{\max} \;\geq\; V(\widehat{\pi}_{\text{PoLeCe}}) \;\geq\; LV_{1-\alpha,\Pi}.$$

Moreover, with probability at least $1-\alpha-\beta-r_n-r_n'-\delta_n$, *the regret of* $\widehat{\pi}_{\text{PoLeCe}}$ *satisfies*

$$V_{\max}-V(\widehat{\pi}_{\text{PoLeCe}}) \;\leq\; V_{\max}-LV_{1-\alpha,\Pi} \;\leq\; \big(\underline{\sigma}_{\Pi_0}/\sqrt{n}\big)\Big\{q_{1-\beta,\Pi_0} \;+\; q_{1-\alpha+r_n,\Pi}\Big\},$$

where $\underline{\sigma}_{\Pi_0} = \sqrt{n}\,\min_{\pi\in\Pi_0}\widehat{s}(\pi)$ *is determined by the* least risky *policy among those with maximal true policy welfare.*

The first result in Proposition 3 shows that $LV_{1-\alpha,\Pi}$ provides a valid high-confidence lower bound on the true policy welfare of the PoLeCe choice. The second result sharpens the general regret bounds presented in Proposition 1. As noted following Proposition 1, the parameter β governs the trade-off between failure probability and bound tightness, while the constant α, chosen by the researcher, determines the confidence level of the lower bound $LV_{1-\alpha,\Pi}$. The refinement in Proposition 3 depends on the relationship between α and β. When $\alpha=\beta$, Propositions 1 and 3 yield qualitatively identical guarantees. When $\alpha<\beta$, the bounds in both propositions remain qualitatively identical, but the bound in Proposition 3 holds with higher probability. When $\alpha>\beta$, the bound in Proposition 3 is tighter, but it may hold with lower probability. For a large α, the bound may fail to hold; in such cases, the general bound from Proposition 1 remains valid and informative, as it applies to all risk-aware decision rules.

Another key observation is that the leading constant in the regret bound is governed by the least risky among the best policies, and can therefore be significantly smaller than that for the EWM rule in settings with substantial risk heterogeneity. This theoretical advantage is corroborated by the computational experiments we present below.

2.4. **Computational Experiments.** Here we provide results of computational experiments calibrated to three empirical applications in Section 3.2. Specifically, we take the welfare estimates and their standard errors $(\widehat{V}(\pi),\widehat{s}(\pi))$ as shown in Figure 2 (DGP 1 and 2), Figure 3a (DGP 3) and Figure 3b (DGP 4) as true parameter values $(V(\pi),s(\pi))$. For each DGP, we take independent draws from $N(V(\pi),s(\pi))$ for $\pi\in\Pi$ and then assess the performance of EWM and PoLeCe based on $(\widehat{V}^*(\pi), s(\pi))$ where $\widehat{V}^*(\pi)$ are the simulation draws. We focus on $\alpha = 0.05$ and since we take $\widehat{V}^*(\pi)$ to be independent across $\pi \in \Pi$, we obtain the critical value $\widehat{q}_{1-\alpha,\Pi}$ via the bootstrap as discussed in Condition (B) by setting the correlation matrix to be the identity matrix. As a benchmark, we also report the performance of choosing the control policy, whose identity is known in each experiment.

Table 1 summarizes the simulation results. In line with our theoretical results, when there is heterogeneity in the estimation risk for policies and $\underline{\sigma}_{\Pi_0}$ (the lower bound on the risk of the best policies Π_0) is small, as in DGP 1, 2 and 4, PoLeCe outperforms EWM in terms of tail regret. Since PoLeCe is data-dependent, there is still variability in regret, but it only loses to the control policy in rare cases, as shown in DGP 4. Finally, by construction, the lower confidence band for PoLeCe is the highest, and is much higher than EWM in all four DGPs considered in this simulation.

Table 1. Calibrated Simulation Results

		Average regret	Median regret	95%-percentile regret	Average Welfare LCB
Measles shots (DGP 1)	EWM	**7.88%**	**2.12%**	32.10%	-36.20%
	PoLeCe	8.83%	10.02%	**12.84%**	**-21.53%**
	Control	52.64%	52.64%	52.64%	-79.92%
Shots per dollar (DGP 2)	EWM	1.74%	1.77%	4.89%	-6.64%
	PoLeCe	**1.00%**	**0.00%**	**3.83%**	**-5.98%**
	Control	4.89%	4.89%	4.89%	-16.31%
Informal Saving Technology (DGP 3)	EWM	**15.91%**	**17.21%**	**33.85%**	-54.17%
	PoLeCe	28.14%	30.77%	41.45%	**-45.02%**
	Control	66.39%	66.39%	66.39%	-87.82%
Alcohol and Self-Control (DGP 4)	EWM	1.38%	1.08%	4.46%	-3.66%
	PoLeCe	**0.84%**	**0.47%**	2.21%	**-3.46%**
	Control	1.43%	1.43%	**1.43%**	-7.86%

Note. Based on 10,000 simulation draws with nominal level $\alpha = 0.05$. Regret is measured in percentage $(V_{\max} - V(\widehat{\pi}))/V_{\max}$. The Welfare LCB is $\widehat{V}(\widehat{\pi}) - \widehat{q}_{1-\alpha,\Pi}\widehat{s}(\widehat{\pi})$, reported as percentage below $V_{\max}$, that is, $(\widehat{V}(\widehat{\pi}) - \widehat{q}_{1-\alpha,\Pi}\widehat{s}(\widehat{\pi}) - V_{\max})/V_{\max}$.

2.5. **Alternative Motivations for Risk-Aware Decisions.** We consider a decision-maker (DM) who holds posterior beliefs B about policy welfare $\{V(\pi)\}$ given by

$$V(\pi) \;\sim\; \widehat{V}(\pi) \;+\; \widehat{s}(\pi)\, Z_\pi, \tag{8}$$

where $\{Z_\pi\}$ is a centered Gaussian vector with unit variance $\big(\mathrm{E}_B[Z_\pi^2]=1\big)$.[14] Note that $\widehat{V}(\pi)$ and $\widehat{s}(\pi)$ are fixed parameters under the DM's posterior B. They may arise from empirical estimates, incorporate prior information, or reflect other model uncertainties.

The following result shows that this framework encompasses all risk-aware, rational policy choices of a DM whose risk preferences follow the von Neumann-Morgenstern expected utility model. Specifically, suppose the DM has a strictly increasing and concave utility function U and aims to maximize

$$\max_{\pi\in\Pi} \mathrm{E}_B\big[U\big(V(\pi)\big)\big].$$

Proposition 4 (Posterior Expected Utility Maximization)**.** *Consider the DM described above. Assume $\widehat{V}(\pi)$ and $\widehat{s}(\pi)$ are continuous in π, and Π is compact.*

(1) *If U is strictly concave, then the DM's choice can be implemented by solving*

$$\max_{\pi\in\Pi}\left\{\widehat{V}(\pi)\;-\;k_{UB}\,\widehat{s}(\pi)\right\} \tag{9}$$

for some constant k_{UB} that depends on U and B.

(2) *If U is linear (the DM is risk-neutral), then the DM solves the above program with $k_{UB}=0$. Furthermore, any limit point of the solutions to (9) as $k_{UB}\searrow 0$ lies within the solution set for $k_{UB}=0$.*

Thus, a risk-averse DM must solve a variant of (9). As an example, for exponential utility $U(v)=1-\exp(-\lambda v)$, one obtains

$$\mathrm{E}[U(V(\pi))]\;=\;1\;-\;\exp\Big(-\lambda\,\widehat{V}(\pi)\;+\;\tfrac{1}{2}\,\lambda^2\,\widehat{s}^2(\pi)\Big).$$

Maximizing this over $\pi\in\Pi$ is equivalent to (9) for some $k_{UB}>0$. In contrast, setting $k_{UB}=0$ recovers the risk-neutral (EWM) decision. Notably, as k_{UB} approaches zero, solutions to the risk-weighted program (9) converge to the lowest-risk points among the EWM solutions (if the latter is set-valued).

We can gain another perspective by considering a DM with Gaussian beliefs as described above in (8), but whose risk preferences are more directly shaped by a combination of regret and estimation risk, as suggested by Liese and Miescke (2008).[15]

Consider the loss function

$$L(V(\pi))=V_{\max}-V(\pi)+k_R\big|\widehat{V}(\pi)-V(\pi)\big|\sqrt{\boldsymbol{\pi}/2},$$

[14]This is formally justified by the approximate Bayesian framework of Doksum and Lo (1990), in which the DM treats empirical estimates as the input data and uses approximate likelihoods of these estimators to form beliefs about parameters of interest.

[15]The suggestion was made to unify the problems of selection and estimation.

where boldface $\boldsymbol{\pi}$ denotes the area of the unit circle. This loss function has regret component

$$V_{\max} - V(\pi) = \max_{\bar{\pi} \in \Pi} V(\bar{\pi}) - V(\pi)$$

stemming from using the policy rule π rather than the unknown welfare-maximizing choice, and the estimation risk component $\left|\widehat{V}(\pi) - V(\pi)\right|$. Consider a DM who minimizes the expected regret risk-aware loss function:

$$\min_{\pi \in \Pi} \mathrm{E}_B\big[L(V(\pi))\big].$$

Proposition 5 (Best Posterior Regret-Risk Aware Decision)**.** *Assume $\widehat{V}(\pi)$ and $\widehat{s}(\pi)$ are continuous in π, and that Π is compact. The DM described above finds the optimal policy by solving:*

$$\max_{\pi \in \Pi} \Big\{ \widehat{V}(\pi) \ - \ k_R\, \widehat{s}(\pi) \Big\}.$$

Propositions 4 and 5 explain why DMs may prefer a strictly positive k, but the exact value of k is application-specific. One way to set it is by conducting experiments where DMs respond to hypothetical or incentivized lotteries/choices, then calibrating k based on their observed choices. Another approach is to present a range of efficient decision frontiers, have DMs select their preferred allocations, and reverse-engineer k that aligns with those selections. Generally, there is no single universal way to determine the "right" k; its choice depends on context and stakeholder preferences. Using a critical value for k as in PoLeCe provides both interpretable regret and reporting guarantees.

2.6. **Refinements and Theoretical Points.**

2.6.1. *Preliminary Reductions of* Π.

Proofs and theoretical arguments suggest that regret bounds and confidence guarantees can be improved by a preliminary reduction of the consideration set of policies Π. Applicable methods for reducing Π include those based on Chernozhukov, Lee and Rosen (2013) or Romano, Shaikh and Wolf (2014). This reduction can theoretically yield less risk-averse choices of $\hat{k}$, while improving regret performance. Our experiments so far have not shown substantial gains to such reductions in calibrated settings. Hence, we do not elaborate here and leave this as a potential topic for further research.

2.6.2. *Discretization for Theoretical Analysis.*

If the original set of policies, denoted by Π^o, is infinite, we replace it with a finite Π of cardinality p. In this case we need a discretization that is both fine enough and not overly complex, so that with probability at least $1 - \delta_n$,

$$\sqrt{\log p}\ \sup_{\pi \in \Pi^o} \Big(\min_{\nu \in \Pi} |\widehat{Z}_\pi - \widehat{Z}_\nu| \Big) \ \leq \ c r_n,$$

where $0 < c < 1$ is a small enough constant. We then work with Π for theoretical purposes. This condition is often satisfied when Π^o is not too large. For example, in investment allocation,

$$\log p \ \propto \ J \log\big(\tfrac{1}{r_n}\big),$$

where J is the number of programs to invest in. In policy learning,

$$\log p \ \propto \ V \log\big(\tfrac{1}{r_n}\big),$$

where V is the effective dimension (e.g., the VC dimension) of Π^o. Consequently, regret bounds often scale as

$$\sqrt{(V/n)\log n},$$

requiring V to be small relative to n. Note that p can be much larger than the sample size.

3. Empirical Examples

Here we provide two sets of applications. In the first, we investigate the allocation of government funds across public programs, continuing the discussion started in the introduction. In the second, we estimate the best treatment policies using data from three prominent RCTs. Overall, these empirical examples show that it is important to incorporate sampling uncertainty into policy learning.

3.1. **Investment Allocations for Public Programs.** This section continues the discussion from the introduction, and illustrates how to solve the problem of investment allocations for public programs described in Section 2.1. We obtain MVPF estimates $\widehat{R}$ from the Policy Impacts Library, Hendren and Sprung-Keyser (2021). There are 172 programs in total, and we focus on those in the United States whose MVPF is estimated by an RCT. We assume the reported upper and lower bounds correspond to 95% confidence intervals and infer the standard errors for each $\widehat{R}$ from these bounds.[16] This yields 14 programs and we provide the Policy Impacts Library's description of each corresponding social program in Appendix D. We also assume the estimation errors across programs are independent, so their covariance matrix is diagonal, with the squared standard errors on the diagonal. Note that here we only account for estimation error, holding the specification in the original papers that produced these MVPF estimates fixed. Alternative specifications can lead to different MVPF estimates for the same program. For example, accounting for transfers to parents could lead to larger MVPF estimates for early childhood programs.

[16]We therefore exclude programs for which the upper and lower bounds are infinite.

In the following two illustrations, we implement PoLeCe by reformulating the problem as a root-finding task that involves solving concave second-order conic programs, which has superior computational efficiency relative to grid search as detailed in Appendix C. We set $\alpha = 0.05$ throughout.

Table 2 reports the allocation selected by PoLeCe and EWM if we restrict the set of programs to those targeting individuals older than 25. As expected, the EWM rule is a corner solution that places all investment in the best-performing program according to the empirical estimates, Holistic Wrap-around Services Can Improve Employment Rates, as its MVPF estimate is the highest. The PoLeCe method, in sharp contrast to EWM, places a large weight on housing vouchers in Chicago and Medicaid for single adults as their MVPF estimates are among the highest, and are highly precise.

TABLE 2. Results for Investment Allocations to Adult Programs (Age > 25)

	EWM	PoLeCe
Panel A. Outcome: MVPF		
Estimated Value: $\widehat{V}(\widehat{\pi})$	1.39	1.17
LCB: $\widehat{V}(\widehat{\pi}) - \widehat{q}_{0.95,\Pi}\,\widehat{s}(\widehat{\pi})$	-2.06	1.04
Panel B. Selected Policy		
Holistic Wrap-around Services Can Improve Employment Rates: 1.39 (1.11)	1	0.00
Padua Pilot: 0.18 (0.23)	0	0.01
Housing Vouchers in Chicago: 0.65 (0.02)	0	0.68
Oregon Health Insurance Experiment (Provided to Single Adults): 1.16 (0.04)	0	0.25
Income Maintenance Experiment in Seattle and Denver: 0.27 (0.95)	0	0.00
Paycheck Plus: EITC to Adults without Dependents: 1.00 (0.09)	0	0.05
Work Advance: 0.78 (0.28)	0	0.01
Job Training Partnership Act, Adults: 1.38 (0.40)	0	0.00

Note: EWM denotes empirical welfare maximization, and PoLeCe denotes policy learning with confidence. In Panel A, we report the EWM and the PoLeCe lower confidence bound with $\alpha = 0.05$. In Panel B, we report the selected allocation for each program. MVPF estimates and associated standard errors (in parentheses) are included alongside each program name.

Table 3 reports the allocation selected by PoLeCe and EWM if we restrict the set of programs to those targeting individuals aged 25 or under. Compared to allocations for programs for those over age 25 in Table 2, the PoLeCe solutions here allocate a large share towards programs whose MVPF estimates are less precise. This is because $\widehat{q}_{0.95,\Pi}$ is smaller, resulting in less aversion to estimation uncertainty.

3.2. **Choosing Better Treatment Policies.** This section illustrates how to solve the problem of choosing better treatment policies described in Section 2.1. We use RCT data from Banerjee, Chandrasekhar, Dalpath, Duflo, Floretta, Jackson, Kannan, Loza, Sankar,

Table 3. Results for Investment Allocations to Youth Programs (Age $\leq$ 25)

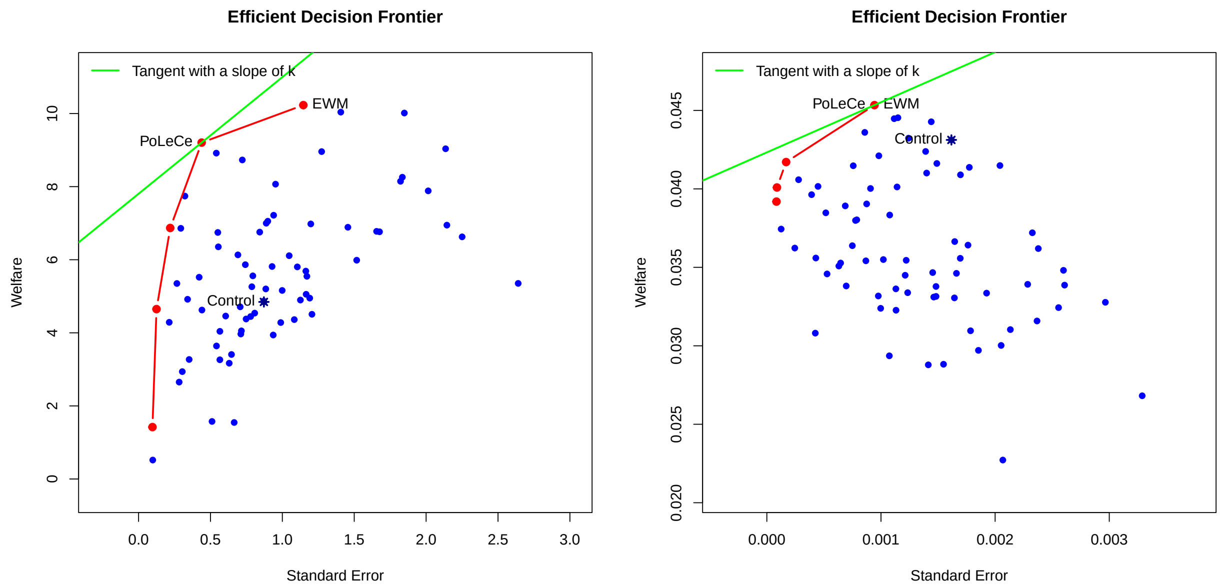
	EWM	PoLeCe
Panel A. Outcome: MVPF		
Estimated Value: $\widehat{V}(\widehat{\pi})$	1.84	1.64
LCB: $\widehat{V}(\widehat{\pi}) - \widehat{q}_{0.95,\Pi}\widehat{s}(\widehat{\pi})$	0.51	0.78
Panel B. Selected Policy		
Head Start Impact Study: 1.84 (0.47)	1	0.52
Wisconsin Scholar Grant to Low-Income College Students: 1.43 (0.38)	0	0.48
JobStart: 0.20 (0.11)	0	0
Year Up: 0.43 (0.03)	0	0
Job Corps: 0.15 (0.22)	0	0
Job Training Partnership Act, Youth: -0.23 (0.81)	0	0

Note: EWM denotes empirical welfare maximization, and PoLeCe denotes policy learning with confidence. In Panel A, we report the EWM and the PoLeCe lower confidence bound with $\alpha = 0.05$. In Panel B, we report the selected allocation for each program. MVPF estimates and associated standard errors (in parentheses) are included alongside each program name.

Schrimpf and Shrestha (2024), Dupas and Robinson (2013), and Schilbach (2019) to illustrate the usefulness of our method and demonstrate how the results can be incorporated into research papers that analyze RCTs. These papers focus on estimating the added value relative to control (as defined in (2)). In contrast, we focus on each policy's value to fully account for the estimation error for estimating the average outcome under control. While these papers originally used regression adjustment, we rely on efficient influence functions (doubly robust scores via (3)) to obtain welfare estimates for potential precision gains.

3.2.1. *Immunization Nudges.* The first example illustrates the trade-off between welfare and precision in selecting the best arm. Banerjee et al. (2024) conducted a large-scale RCT of nudges to encourage immunization in the state of Haryana, Northern India. The study employed a cross-randomized design involving three main types of nudges: providing incentives, sending short messaging service (SMS) reminders, and seeding community ambassadors. Each type included multiple variations, resulting in 74 distinct treatments and a control. While Banerjee et al. (2024) proposed a method to group similar treatments for analysis, we retain the original set of treatments for illustrative purposes.

The outcomes include the number of measles shots administered, and number of shots per dollar spent. We observe 7,370 households and use doubly robust scores to obtain the welfare estimates $\widehat{V}(\pi)$ and their standard errors. Following Banerjee et al. (2024), we weight these village-level regressions by village population, and standard errors are clustered at the SC level. We calculate $\widehat{k} = \widehat{q}_{1-\alpha,\Pi}$ by the bootstrap procedure in Condition (B).



(A) Measles shots outcome (B) Shots per dollar outcome

FIGURE 2. Welfare estimates against precision

Notes: Each blue circle represents a pair of welfare estimate and its standard error $(\widehat{V}(\pi), \widehat{s}(\pi))$ for a given treatment π. The connected red circles show the decision frontier.

As shown in Figure 2 , there is substantial variation in the precision of welfare estimates across treatments. While EWM selects the treatment with the highest welfare estimate $\widehat{V}(\pi)$, PoLeCe adjusts for estimation uncertainty and may choose differently. Here we set $\alpha = 0.05$ which places large emphasis on precision. For example, for shots per dollar, EWM and PoLeCe both select `trusted info hub, no incentive, no reminder`, as there is minimal trade-off between welfare and precision. However, for number of shots, EWM selects `trusted info hub, high slope, high SMS`, while PoLeCe selects `info hub, low slope, low SMS` because the latter has comparable welfare but much higher precision.

3.2.2. *Informal Savings Technologies.* Dupas and Robinson (2013) conducted a field experiment in Kenya to study how simple informal savings technologies can increase investment in preventative health and reduce vulnerability to health shocks. The experiment was run through Rotating Savings and Credit Associations (ROSCAs), and all participants were required to be enrolled in a ROSCA at the start.[17]

Specifically, 771 ROSCA participants were randomized to five treatment groups: (i) *Control*; (ii) *Safe Box*: participants were given a simple locked box made out of metal, while

[17]ROSCAs are informal savings groups. Members come together on a regular basis and contribute to a common pot of money which is taken home by one member on a rotating basis.

Table 4. Results for Immunization Nudges

	EWM	PoLeCe	Control
Panel A. Outcome: Measles shots			
Estimated Value: $\widehat{V}(\widehat{\pi})$	10.23	9.21	4.85
LCB: $\widehat{V}(\widehat{\pi}) - \widehat{q}_{0.95,\Pi}\widehat{s}(\widehat{\pi})$	6.57	7.81	2.06
Panel B. Outcome: Shots per dollar			
Estimated Value: $\widehat{V}(\widehat{\pi})$	0.045	0.045	0.043
LCB: $\widehat{V}(\widehat{\pi}) - \widehat{q}_{0.95,\Pi}\widehat{s}(\widehat{\pi})$	0.042	0.042	0.038
Panel C. Selected Policy:			
`Measles shots`	*trusted info hub, high slope high SMS*	*info hub, low slope low SMS*	*no info, no incentive no reminder*
`Shots per dollar`	*trusted info hub, no incentive no reminder*	*trusted info hub, no incentive no reminder*	*no info, no incentive no reminder*

Note. EWM refers to empirical welfare maximization, and PoLeCe corresponds to policy learning with confidence. In Panels A and B, the EWM and PoLeCe lower confidence bounds with $\alpha = 0.05$ are provided. In Panel C, selected treatment combinations are given.

they were asked to record what health product they were saving for, and its cost, on a passbook; (iii) *Lockbox*: participants were given a passbook and a locked box identical to those in the *Safe Box* treatment, except that the program officer kept the key; (iv) *Health Pot*: participants were given a side pot that the members could contribute to in addition to the regular ROSCA pot;[18] (v) *Health Savings Account, or HSA*: participants were encouraged to make regular deposits into an individual HSA managed by the ROSCA treasurer.

The outcomes include investments in health and measures of whether people have trouble affording medical treatments. As an illustration, we focus on health investments in terms of amount (in Ksh) spent on preventative health products. We choose two baseline covariates: `Female` and `Married` purely for simplicity of illustration, without taking a stand on whether they *should* be used in this setting.[19] We obtain 691 individuals after removing 80 participants who received multiple treatments. The randomization was done after stratifying on some ROSCA characteristics. We incorporate strata indicators as well as `Female` and `Married` for calculating the doubly robust scores.

There are $|\mathcal{T}| = 625 = 5^4$ possible treatment policies all together because there are five treatment values and four possible values in the support of the two covariates. Figure 3a plots all welfare estimates relative to their precision. The highest welfare is observed below

[18]Unlike the regular pot, this pot would be earmarked for a specific health product.

[19]In practice, equity, fairness or other concerns could make the use of these covariates for heterogeneous policy assignment unwarranted, and the use of other covariates, either in addition or separately, may be preferred. One could then compute the PoLeCe rule that incorporates the desired covariates.

Table 5. Results for Informal Savings Technologies

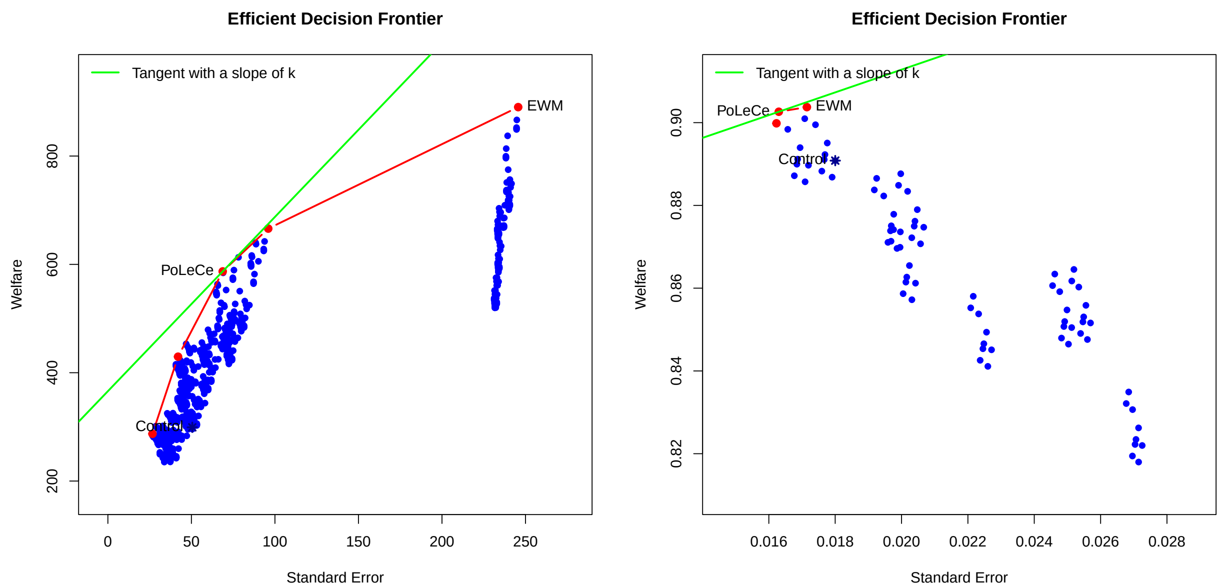
	EWM	PoLeCe	Control
Panel A. Outcome: Health Investment (in Ksh)			
Estimated Value: $\widehat{V}(\widehat{\pi})$	890	587	299
LCB: $\widehat{V}(\widehat{\pi}) - \widehat{q}_{0.95,\Pi}\widehat{s}(\widehat{\pi})$	100	366	137
Panel B. Selected Policy:			
$(\texttt{Female}, \texttt{Married}) = (0, 0)$	*Health Pot*	*Health Pot*	*Control*
$(\texttt{Female}, \texttt{Married}) = (0, 1)$	*Health Pot*	*Health Pot*	*Control*
$(\texttt{Female}, \texttt{Married}) = (1, 0)$	*Health Pot*	*Safe Box*	*Control*
$(\texttt{Female}, \texttt{Married}) = (1, 1)$	*Safe Box*	*Health Pot*	*Control*

Note. EWM refers to empirical welfare maximization, and PoLeCe corresponds to policy learning with confidence. In Panel A, EWM and PoLeCe lower confidence bounds with $\alpha = 0.05$ are provided along with that for the control group policy. In Panel B, selected treatment for each vector of (`Female`, `Married`) is given.

900 Ksh, but it is accompanied by the largest standard error. Therefore, it is advisable to account for the precision of these estimates when selecting an optimal policy. To implement PoLeCe at the level $\alpha = 0.05$ and calculate $\widehat{q}_{0.95,\Pi}$, we apply the bootstrap procedure from Condition (B). Table 5 summarizes the empirical results. In Panel A, the maximized welfare by EWM is 890 Ksh, which corresponds to the highest welfare in Figure 3a. The adjusted estimate of maximum welfare $\widehat{V}(\widehat{\pi}_{\text{PoLeCe}})$ at $\alpha = 0.05$ is 587, which is much less than no-precision-corrected $\widehat{V}(\widehat{\pi}_{\text{EWM}})$ but substantially greater than $\widehat{V}(\widehat{\pi}_{\text{EWM}}) - \widehat{q}_{0.95,\Pi}\widehat{s}(\widehat{\pi}_{\text{EWM}})$. Furthermore, the chosen optimal policies differ for females. EWM selects *Health Pot* for unmarried women and *Safe Box* for married women. In contrast, after accounting for standard errors, PoLeCe recommends *Safe Box* for unmarried women and *Health Pot* for married women.

3.2.3. *Alcohol and Self-Control.* Schilbach (2019) studied alcohol consumption among low-income workers in India. Specifically, 229 cycle-rickshaw drivers in India were randomized to three treatment groups: (i) *control group*: participants were paid Rs 90 ($1.50) each day for visiting the study office; (ii) *incentive group*: drivers were given incentives to remain sober—the payment was Rs 60 ($1.00) if they arrived at the office with a positive blood alcohol content (BAC) and Rs 120 ($2.00) if they arrived sober; (iii) *choice group*: individuals were given the same incentives earlier in the study and then offered to choose between incentives and unconditional payments in the later phase of the study. See Schilbach (2019) for details on the experiment and background.

As an illustration, we focus on labor supply measured by the fraction of days individuals worked from day 5 (the first day of sobriety incentives) through day 19 (the last day of



(A) Health investment (in Ksh) (B) Labor supply

FIGURE 3. Welfare estimates against precision

Notes: Each blue circle represents a pair of welfare estimate and its standard error $(\widehat{V}(\pi), \widehat{s}(\pi))$ for a given treatment π. The connected red circles show the decision frontier.

sobriety incentives). For simplicity, we choose two baseline covariates: `Baseline sober` $= 1$ if the baseline fraction sober is 1 and `Baseline sober` $= 0$ otherwise; `Owns rickshaw` $= 1$ if a driver own a rickshaw and `Owns rickshaw` $= 0$ otherwise. After removing observations with missing values, there are 222 individuals. Given complete randomization, we calculate doubly robust scores based on the two baseline covariates.

TABLE 6. Results for Self-control

	EWM	PoLeCe	Control
Panel A. Outcome: Labor Supply			
Estimated Value: $\widehat{V}(\widehat{\pi})$	0.904	0.903	0.891
LCB: $\widehat{V}(\widehat{\pi}) - \widehat{q}_{0.95,\Pi}\widehat{s}(\widehat{\pi})$	0.857	0.858	0.841
Panel B. Selected Policy:			
(`Baseline sober`, `Owns rickshaw`) $= (0, 0)$	Control	Control	Control
(`Baseline sober`, `Owns rickshaw`) $= (0, 1)$	Control	Incentive	Control
(`Baseline sober`, `Owns rickshaw`) $= (1, 0)$	Choice	Choice	Control
(`Baseline sober`, `Owns rickshaw`) $= (1, 1)$	Choice	Choice	Control

Note. EWM refers to empirical welfare maximization, and PoLeCe corresponds to policy learning with confidence. In Panel A, the EWM point estimate and the PoLeCe with $\alpha = 0.05$ are provided along the control group policy. In Panel B, selected treatment for each vector of (`Baseline sober`, `Owns rickshaw`) is given.

There are $|\mathcal{T}| = 81 = 3^4$ possible treatment policies in total since there are three treatment levels and four possible values of the two covariates. Figure 3b plots the welfare estimates against their precision. Although the highest welfare appears with a relatively small standard error, this does not necessarily imply statistical significance. To implement PoLeCe at the level $\alpha = 0.05$ and calculate $\widehat{q}_{0.95,\Pi}$, we apply the bootstrap procedure from Condition (B). Table 6 summarizes the empirical results. The EWM and PoLeCe estimates, along with their lower confidence bounds, are very similar and only slightly exceed those of the control group.

4. Conclusion

In this paper we have focused on the problem faced by a DM who wishes to choose from a menu of policies that which will deliver the highest welfare using imperfect sample estimates. In order to balance the estimated performance of each policy with its associated statistical uncertainty, we proposed and analyzed the properties of a class of risk-aware policies that make the precision/performance tradeoff explicit. Such policies solve the expected loss function of a risk-averse planner and achieve favorable regret properties. We proposed a specific rule from the class of risk-aware policies, namely the PoLeCe rule, which uses a data-dependent construction for balancing the inherent tradeoff between estimated performance and sample uncertainty, such that a lower confidence bound on the welfare obtained by the chosen policy is automatically provided. Application of the rule was demonstrated across four settings; one featuring the allocation of government funds across social programs using the MVPF, and three others featuring optimal treatment assignment. Intuitively, the approach offers a principled and robust way for a decision maker to hedge against selecting the best-performing policy in sample, when that policy exhibits large sample variation relative to other well-performing policies. Importantly, risk-aware policy rules allow policies with fractional allocations.

A large body of work synthesized in Manski (2013, 2019, 2024) has advocated for greater acknowledgement and incorporation of uncertainty in planning problems. This paper contributes by proposing a principled way to acknowledge and incorporate statistical uncertainty into decision making. Statistical uncertainty is however only one of the many different types of uncertainty that may be present.[20] We have focused on settings in which the welfare of each policy is point identified, such that consistent estimates of the welfare and sample variance of the policies are available. Application of the concepts developed here to settings

[20]For example, Manski (2019) discusses transitory uncertainty, permanent uncertainty, and conceptual uncertainty. The statistical uncertainty considered here is one type of permanent uncertainty.

that feature so-called "deep uncertainty", or ambiguity, would seem an important direction for further development given the large number of settings in which the mean performance of various policies is only credibly partially identified.[21] This would require balancing statistical uncertainty with a collection of interval estimates for each policy's performance.

While the nuance required to formally develop such an approach is beyond the scope of the present paper, a rough prescription could be made based on an extension of the reporting guarantee established here. To see how, consider the definition of the PoLeCe rule in (7). The rule selects the policy that maximizes a $1-\alpha$ lower confidence band for the maximal achievable welfare. Construction of such a confidence band does not require that the optimal level of welfare or the optimal rule be point identified, and could be implemented using techniques developed in e.g. Chernozhukov et al. (2013) under partial identification. This is of course not the only way to balance statistical uncertainty and ambiguity.[22] A more thorough study of the performance of such an approach, and the implicit tradeoff between statistical uncertainty and ambiguity due to partial identification would seem a useful direction for future research.

Appendix A. Deferred Proofs

In what follows let

$$\kappa_{1-\alpha,K} := (1-\alpha)\text{-Quantile}\left(\sup_{\pi\in K} \widehat{Z}_\pi\right)$$
$$q_{1-\alpha,K} := (1-\alpha)\text{-Quantile}\left(\sup_{\pi\in K} Z_\pi\right)$$
$$\widehat{q}_{1-\alpha,K} := (1-\alpha)\text{-Quantile}^*\left(\sup_{\pi\in K} Z^*_\pi\right)$$

Lemma 1 (Quantile Comparison)**.** *With probability at least $1-\delta_n$, critical values obey the following inequalities:*

$$\kappa_{1-\alpha-2r_n,K} \leq q_{1-\alpha-r_n,K} \leq \widehat{q}_{1-\alpha,K} \leq q_{1-\alpha+r_n,K}$$

Furthermore,

$$q_{1-\alpha,K} \leq \mathrm{E}\sup_{\pi\in K} Z_\pi + \sqrt{2\log 1/\alpha} \leq \sqrt{2\log|K|} + \sqrt{2\log 1/\alpha}.$$

[21]There is now a long line of research on treatment choice when mean treatment performance is only partially identified, going back to Manski (2000). Recent research with references to the broader literature includes Russell (2020), Ishihara and Kitagawa (2021), Yata (2021), Christensen, Moon and Schorfheide (2023), Kido (2023), and Montiel Olea, Qiu and Stoye (2023).

[22]For instance, Ben-Michael, Greiner, Imai and Jiang (2025) develops an alternative approach employing a minimax rule over a uniform two-sided confidence band for the conditional mean of potential outcomes in a setting with partial identification due to lack of overlap.

We also note that a lower bound of the form $\sqrt{\log |K|}$ holds, provided that $Z'_\pi s$ are not too correlated, using Sudakov's minoration. The upper bound here follows from the standard Gaussian concentration bounds.

Proof of Lemma 1. By Condition (G) we have

$$\sup_{K \subset \Pi} \sup_{x \in \mathbb{R}} \left| \Pr\left(\sup_{\pi \in K} \widehat{Z}_\pi \leq x \right) - \Pr\left(\sup_{\pi \in K} Z_\pi \leq x \right) \right| \leq r_n,$$

By Condition (B) we have with probability at least $1 - \delta_n$:

$$\sup_{K \subset \Pi} \sup_{x \in \mathbb{R}} \left| \Pr^*\left(\sup_{\pi \in K} \widehat{Z}^*_\pi \leq x \right) - \Pr\left(\sup_{\pi \in K} Z_\pi \leq x \right) \right| \leq r_n,$$

These relations imply that w.p. $\geq 1 - \delta_n$:

$$(1 - \alpha - 2r_n) - \text{Quantile}\left(\sup_{\pi \in K} \widehat{Z}_\pi \right) \leq (1 - \alpha - r_n) - \text{Quantile}\left(\sup_{\pi \in K} Z_\pi \right)$$
$$\leq (1 - \alpha)\text{-Quantile}^*\left(\sup_{\pi \in K} Z^*_\pi \right) \leq (1 - \alpha + r_n)\text{-Quantile}\left(\sup_{\pi \in K} Z_\pi \right).$$

Thus, the first claim follows.

The second claim follows by the Gaussian concentration inequality of Borell-Sudakov-Tsirelson:

$$q_{1-\alpha,K} \leq \mathrm{E} \sup_{\pi \in K} Z_\pi + \sqrt{2 \log 1/\alpha},$$

and by the standard calculation we have $\mathrm{E} \sup_{\pi \in K} Z_\pi \leq \sqrt{2 \log |K|}$, where $|K| \leq p$ is the cardinality of K. $\square$

Proof of Proposition 1. Let

$$\Delta^+_\Pi := \max_{\pi \in \Pi} (\widehat{Z}_\pi); \quad \Delta^-_\Pi := \min_{\pi \in \Pi} (\widehat{Z}_\pi);$$
$$\bar{\sigma}_\Pi := \sqrt{n} \max_{\pi \in \Pi} \widehat{s}(\pi); \quad \widehat{\pi} := \widehat{\pi}_{\mathrm{RW}}(\widehat{k});$$
$$\underline{\sigma}_{\Pi_0} := \min_{\pi \in \Pi_0} \sqrt{n} \widehat{s}(\pi); \quad \widehat{\pi}_0 = \arg\min_{\pi \in \Pi_0} \sqrt{n} \widehat{s}(\pi);$$
$$L(\widehat{k}) := \max_{\pi \in \Pi} \{ \widehat{V}(\pi) - \widehat{k} \widehat{s}(\pi) \} = \widehat{V}(\widehat{\pi}) - \widehat{k} \widehat{s}(\widehat{\pi}).$$

We then have

$$\begin{aligned} L(\widehat{k}) - V_{\max} &= \widehat{V}(\widehat{\pi}) - \widehat{k}\widehat{s}(\widehat{\pi}) - V_{\max} \\ &\geq \widehat{V}(\widehat{\pi}_0) - \widehat{k}\widehat{s}(\widehat{\pi}_0) - V_{\max} \\ &= V(\widehat{\pi}_0) + (\widehat{Z}_{\widehat{\pi}_0} - \widehat{k})\widehat{s}(\widehat{\pi}_0) - V_{\max} \\ &\geq (\widehat{Z}_{\widehat{\pi}_0} - \widehat{k})\widehat{s}(\widehat{\pi}_0) \\ &= (\widehat{Z}_{\widehat{\pi}_0} - \widehat{k})\underline{\sigma}_{\Pi_0}/\sqrt{n} \\ &\geq (\Delta^-_{\Pi_0} - \widehat{k})\underline{\sigma}_{\Pi_0}/\sqrt{n}. \end{aligned}$$

Also, we have

$$\begin{aligned} L(\widehat{k}) - V(\widehat{\pi}) &= \widehat{V}(\widehat{\pi}) - \widehat{k}\widehat{s}(\widehat{\pi}) - V(\widehat{\pi}) \\ &= \widehat{Z}_{\widehat{\pi}}\widehat{s}(\widehat{\pi}) - \widehat{k}\widehat{s}(\widehat{\pi}) \\ &\leq (\widehat{Z}_{\widehat{\pi}} - \widehat{k})_+\widehat{s}(\widehat{\pi}) \\ &\leq (\widehat{Z}_{\widehat{\pi}} - \widehat{k})_+\bar{\sigma}_{\Pi}/\sqrt{n} \\ &\leq (\Delta^+_{\Pi} - \widehat{k})_+\bar{\sigma}_{\Pi}/\sqrt{n} \end{aligned}$$

Combining the two bounds, we have

$$\begin{aligned} V_{\max} - V(\widehat{\pi}) &= V_{\max} - L(\widehat{k}) + L(\widehat{k}) - V(\widehat{\pi}) \\ &\leq (-\Delta^-_{\Pi_0} + \widehat{k})\underline{\sigma}_{\Pi_0}/\sqrt{n} + (\Delta^+_{\Pi} - \widehat{k})_+\bar{\sigma}_{\Pi}/\sqrt{n} \end{aligned}$$

By the proof of Lemma 1, we have that

$$\Pr(\Delta^+_{\Pi} \leq q_{1-\beta,\Pi}) \geq 1 - \beta - r_n, \quad \Pr(-\Delta^-_{\Pi_0} \leq q_{1-\beta,\Pi_0}) \geq 1 - \beta - r_n.$$

It then follows that with probability at least $1 - 2\beta - 2r_n$, we have

$$V_{\max} - V(\widehat{\pi}) \leq (q_{1-\beta,\Pi_0} + \widehat{k})\underline{\sigma}_{\Pi_0}/\sqrt{n} + (q_{1-\beta,\Pi} - \widehat{k})_+\bar{\sigma}_{\Pi}/\sqrt{n}.$$

□

Proof of Proposition 2. Consider the event

$$\begin{aligned} \mathcal{E}_{\Pi} &:= \{V(\pi) \geq \widehat{V}(\pi) - \widehat{k}\widehat{s}(\pi),\ \forall \pi \in \Pi\} \\ &= \{-\widehat{Z}_{\pi} \geq -\widehat{k},\ \forall \pi \in \Pi\} \\ &= \{\max_{\pi \in \Pi} \widehat{Z}_{\pi} \leq \widehat{k}\}. \end{aligned}$$

By Lemma 1, our bootstrap choice $\widehat{k} = \widehat{q}_{1-\alpha,\Pi}$ leads to

$$\kappa_{1-\alpha-2r_n,\Pi} \leq \widehat{k} \leq q_{1-\alpha+r_n,\Pi}$$

with probability at least $1-\delta_n$, and so it follows that the event $\mathcal{E}_\Pi$ occurs with probability at least $1-\alpha-r'_n$ for

$$r'_n = 2r_n + \delta_n,$$

so the claim follows. □

Proof of Proposition 3. The first claim was shown in the proof of the previous proposition.

We next show the second claim about the regret bound. Since $V_{\max} = V(\widehat{\pi}_0)$, on the event $\mathcal{E}_\Pi$ we have

$$\begin{aligned} V_{\max} - V(\widehat{\pi}) &\leq V_{\max} - \max_{\pi\in\Pi}\{\widehat{V}(\pi) - \widehat{k}\widehat{s}(\pi)\} \\ &\leq V(\widehat{\pi}_0) - (\widehat{V}(\widehat{\pi}_0) - \widehat{k}\widehat{s}(\widehat{\pi}_0)) \\ &= -\widehat{Z}_{\widehat{\pi}_0}\widehat{s}(\widehat{\pi}_0) + \widehat{k}\widehat{s}(\widehat{\pi}_0) \\ &= (\underline{\sigma}_{\Pi_0}/\sqrt{n})(\widehat{k} - \widehat{Z}_{\widehat{\pi}_0}) \\ &\leq (\underline{\sigma}_{\Pi_0}/\sqrt{n})(\widehat{k} - \Delta^-_{\Pi_0}). \end{aligned}$$

By the proof of Lemma 1, we have that $\Pr(-\Delta^-_\Pi \leq q_{1-\beta,\Pi_0}) \geq 1-\beta-r_n$.

It then follows with probability $1-\alpha-\beta-r_n-r'_n-\delta_n$:

$$V_{\max} - V(\widehat{\pi}) \leq [q_{1-\alpha+r_n,\Pi} + q_{1-\beta,\Pi_0}](\underline{\sigma}_{\Pi_0}/\sqrt{n})$$

□

Proof of Proposition 4. Claim (1). Let $v = \widehat{V}(\pi)$ and $s = \widehat{s}(\pi)$. Then

$$\mathrm{E}_B\big[U\big(V(\pi)\big)\big] \;=\; \mathrm{E}\big[U(v+sZ)\big], \quad Z \sim N(0,1).$$

Define $f(v,s) := \mathrm{E}\big[U(v+sZ)\big]$. The function $(v,s) \mapsto f(v,s)$ is *increasing* in v and *decreasing* in s. Indeed,

$$\frac{\partial}{\partial v} f(v,s) \;=\; \mathrm{E}[U'(v+sZ)] \;>\; 0, \quad \frac{\partial}{\partial s} f(v,s) \;=\; \mathrm{E}\big[U'(v+sZ)\,Z\big] \;<\; 0,$$

because $U'(\cdot) > 0$ and $-U'(v+sZ)$ is comonotonic with Z (given that U'' exists by Alexandrov's theorem almost everywhere and $U'' < 0$ by strict concavity). The second inequality follows from Chebyshev's association inequality, and the first holds because the expectation of a strictly positive random variable is positive.

Let $\big(\widehat{V}(\widehat{\pi}), \widehat{s}(\widehat{\pi})\big)$ correspond to a choice $\widehat{\pi}$ that maximizes $\mathrm{E}_B[U(V(\pi))]$ over $\pi \in \Pi$. By the monotonicity properties, we have

$$\widehat{V}(\widehat{\pi}) \;=\; \max_{\pi\in\Pi}\Big\{\widehat{V}(\pi) : \widehat{s}(\pi) \;\le\; \widehat{s}(\widehat{\pi})\Big\}.$$

This maximization can be written in a Lagrangian form:

$$\max_{\pi\in\Pi}\Big\{\widehat{V}(\pi) - k\,(\widehat{s}(\pi) - \widehat{s}(\widehat{\pi}))\Big\},$$

where $k > 0$ is the Lagrange multiplier, which depends on U and the data $\{\widehat{V}(\pi), \widehat{s}(\pi)\}$ defining the beliefs B. We can then remove $k\widehat{s}(\widehat{\pi})$ from the program without affecting its solution.

Claim (2). The first assertion of the claim for $k_{UB} = 0$ follows from $\mathrm{E}_B[V(\pi)] = \hat{V}(\pi)$. The second assertion of the claim follows from Berge's maximum theorem. □

Proof of Proposition 5. Observing that

$$\mathrm{E}_B\big[V_{\max} - V(\pi)\big] = \mathrm{E}_B\big[V_{\max}\big] - \widehat{V}(\pi),$$

and

$$\mathrm{E}_B\big|\widehat{V}(\pi) - V(\pi)\big| = \widehat{s}(\pi)\sqrt{\frac{2}{\boldsymbol{\pi}}},$$

the program becomes

$$\min_{\pi\in\Pi}\Big\{\mathrm{E}_B\big[V_{\max}\big] - \widehat{V}(\pi) + k_R\,\widehat{s}(\pi)\Big\}.$$

We can now omit the term $\mathrm{E}_B\big[V_{\max}\big]$, resulting in the expression stated in the claim, without affecting the solution of the program. □

Appendix B. Efficient Decision Frontier Computation

In this part of the appendix, we provide pseudocode of the algorithm to compute the efficient decision frontier. The code below builds a monotonically nondecreasing, concave frontier from the input of points.

Note that the algorithm loops over sorted points, skipping any new point with a lower Value than the frontier's last point, thereby enforcing a monotonically nondecreasing frontier. It also ensures the concavity of the frontier by removing a "bulge," thus guaranteeing that the slopes decrease as we move right. Here, the bulge is checked by whether the following holds:

$$\frac{p.Value - p1.Value}{p.Risk - p1.Risk} > \frac{p2.Value - p1.Value}{p2.Risk - p1.Risk}.$$

Algorithm 1 Build a Monotonically Nondecreasing, Concave Frontier

Require: $points = (Risk, Value)$ such that $Risk$ and $Value$ have the same length
Ensure: The set of points that form the efficient frontier

```
procedure FrontierEstimation(points)
    Create data structure from input
    Sort by ascending Risk, tie-break by descending Value
    frontier ← empty list
    for i ← 1 to the number of rows of points do
        p ← row i of points                                  ▷ extract row i from points
        if frontier is not empty and p.Value < frontier[last].Value then
            continue                    ▷ skip to next iteration; ensure monotonicity
        end if
        while there are at least two elements in frontier do
            p2 ← last frontier point
            p1 ← second to last frontier point
            if (p.Value − p1.Value)/(p.Risk − p1.Risk) > (p2.Value − p1.Value)/(p2.Risk − p1.Risk) then
                remove p2 from frontier                      ▷ ensure concavity
            else
                break
            end if
        end while
        append p to frontier
    end for
    return frontier                                          ▷ the set of frontier points
end procedure
```

That is, each new point to the frontier must produce a slope that is less than or equal to the previous segment's slope.

For the investment allocation problem, it may be infeasible to discretize Π and apply the above algorithm to plot the efficiency frontier. Instead, we start with a fine grid for k centered at $\hat{q}_{0.95,\Pi}$, and directly trace out the efficiency decision frontier by solving the corresponding optimization problem as in (4). Appendix C provides computationally efficient way to solve such optimization and therefore to plot the efficiency frontier.

Appendix C. Gaussian Bootstrap for Investment Allocation Problem

The Gaussian bootstrap is a resampling technique widely used in statistical inference to approximate the distribution of a statistic. A critical step in this procedure involves computing the statistic

$$\bar{Z}^* = \max_{\pi \in \Pi} Z^*_\pi,$$

where

$$Z_\pi^* = \frac{\pi^\top Z^*}{\sqrt{\pi^\top \widehat{\Omega}\, \pi}},$$

and

$$Z^* = \widehat{\Omega}^{1/2}\, N^*(0, I).$$

Here, $N^*(0, I)$ denotes a standard Gaussian vector in $\mathbb{R}^d$, $\widehat{\Omega}$ is a positive-definite covariance estimator, and $\Pi \subset \mathbb{R}^d$ is a convex, bounded subset of the probability simplex.

At first glance, the computation of $\bar{Z}^*$ appears to require solving a nonconcave fractional optimization problem:

$$\max_{\pi \in \Pi} \frac{\pi^\top Z^*}{\sqrt{\pi^\top \widehat{\Omega}\, \pi}},$$

which can be computationally prohibitive. However, by reformulating the problem as a root-finding task involving concave second-order conic programs, we can achieve computational efficiency.

Reformulation as a Root-Finding Problem. Instead of directly tackling the fractional optimization, we consider the following equivalence:

$$\bar{Z}^* = \mathrm{root}_{t \geq 0} \left\{ \max_{\pi \in \Pi} \left(\pi^\top Z^* - t\sqrt{\pi^\top \widehat{\Omega}\, \pi} \right) = 0 \right\}.$$

This reformulation allows us to solve for $\bar{Z}^*$ by identifying the unique value of t that zeroes the function

$$f(t) = \max_{\pi \in \Pi} \left(\pi^\top z - t\sqrt{\pi^\top \widehat{\Omega}\, \pi} \right)$$

when $Z^* = z$. The following lemma formalizes the properties of $f(t)$ necessary for this approach.

Lemma 2 (Existence, Uniqueness, and Differentiability of the Root)**.** *Let $\Pi \subset \mathbb{R}^d$ be a convex, closed subset of the probability simplex, and let $\widehat{\Omega}$ be a positive-definite matrix. Define*

$$f(t) = \max_{\pi \in \Pi} \left(\pi^\top z - t\sqrt{\pi^\top \widehat{\Omega}\, \pi} \right) \quad \text{for } t \geq 0.$$

Then:

(1) *For each fixed $t \geq 0$, the optimization problem*

$$\max_{\pi \in \Pi} \left(\pi^\top z - t\sqrt{\pi^\top \widehat{\Omega}\, \pi} \right)$$

can be reformulated as a concave second-order conic optimization problem.

(2) *There exists a unique $t^* \geq 0$ such that $f(t^*) = 0$.*

(3) *The derivative of f at t is given by*

$$f'(t) = -\sqrt{\pi^*(t)^\top \widehat{\Omega}\, \pi^*(t)}\,,$$

where $\pi^*(t) \in \arg\max_{\pi \in \Pi} \left(\pi^\top z - t\sqrt{\pi^\top \widehat{\Omega}\, \pi} \right)$. Consequently, $f(t)$ is strictly decreasing.

Proof. **1. Second-Order Conic Reformulation.** Fix any $t \geq 0$. Consider

$$f(t) = \max_{\pi \in \Pi} \left(\pi^\top z - t\sqrt{\pi^\top \widehat{\Omega}\, \pi} \right).$$

Introduce an auxiliary variable s such that

$$s = -\sqrt{\pi^\top \widehat{\Omega}\, \pi} \quad \Leftrightarrow \quad \|\widehat{\Omega}^{1/2}\, \pi\|_2 \leq -s, \quad s \leq 0.$$

The optimization problem becomes

$$\begin{aligned} \max_{\pi, s} \quad & \pi^\top z + t\, s \\ \text{subject to} \quad & \|\widehat{\Omega}^{1/2}\, \pi\|_2 \leq -s, \\ & s \leq 0, \\ & \pi \in \Pi. \end{aligned}$$

The constraint $\|\widehat{\Omega}^{1/2}\, \pi\|_2 \leq -s$ is a standard second-order cone constraint. Therefore, the problem is a concave second-order conic program.

2. Existence and Uniqueness of the Root. Existence. Observe that

$$f(0) = \max_{\pi \in \Pi} \pi^\top z,$$

which is finite since Π is bounded and closed. As $t \to \infty$, the term $-t\sqrt{\pi^\top \widehat{\Omega}\, \pi}$ dominates for every $\pi \in \Pi$, leading to $f(t) \to -\infty$. Since $f(t)$ is continuous (as the pointwise maximum of continuous functions over a compact set), the Intermediate Value Theorem guarantees the existence of some $t^* \geq 0$ such that $f(t^*) = 0$.

Uniqueness. From part 3, we have

$$f'(t) = -\sqrt{\pi^*(t)^\top \widehat{\Omega}\, \pi^*(t)} < 0,$$

which implies that $f(t)$ is strictly decreasing. A strictly decreasing continuous function can cross zero at most once. Therefore, the root t^* satisfying $f(t^*) = 0$ is unique.

3. Derivative of $f(t)$. Define

$$F(\pi, t) = \pi^\top z - t\sqrt{\pi^\top \widehat{\Omega}\, \pi}.$$

Then

$$f(t) = \max_{\pi \in \Pi} F(\pi, t).$$

By the envelope theorem, the derivative of $f(t)$ with respect to t is

$$f'(t) = \frac{\partial}{\partial t} F(\pi^*(t), t) = -\sqrt{\pi^*(t)^\top \widehat{\Omega}\, \pi^*(t)}.$$

Since $\widehat{\Omega}$ is positive-definite and $\pi^*(t) \in \Pi$ with Π bounded, it follows that $\pi^*(t)^\top \widehat{\Omega}\, \pi^*(t) > 0$. Therefore, $f'(t) < 0$, establishing that $f(t)$ is strictly decreasing. □

Implications for Gaussian Bootstrap. The reformulation of computing $\bar{Z}^*$ as a root-finding problem over a family of concave second-order conic programs significantly enhances computational efficiency. Each evaluation of $f(t)$ involves solving a concave SOC optimization problem, which is well-supported by modern optimization solvers. The uniqueness of the root t^* guarantees that iterative root-finding algorithms, such as the Newton-Raphson method or bisection, will converge reliably to the desired solution. Moreover, the explicit expression for the derivative $f'(t)$ facilitates the use of derivative-based optimization methods, further streamlining the computational process.

Appendix D. Further Details on the MVPF Estimates

For ease of reference, the Policy Impacts Library (Hendren and Sprung-Keyser (2021)) descriptions of the social programs corresponding to each MVPF estimate in Tables 2 and 3 are provided below.

Holistic Wrap-around Services Can Improve Employment Rates. The Rochester-Monroe Anti-Poverty Initiative (RMAPI), in partnership with the New York Governor's State Anti-Poverty Task Force, piloted Bridges to Success (BtS) to increase economic mobility for participants. Espinosa, Evans, Phillips and Spilde (2024) evaluates BtS through a 430-person randomized study of the pilot program.

Padua Pilot. Evans, Kolka, Sullivan and Turner (2023) estimates the MVPF of the Padua Pilot, a holistic and intensive case-management program designed by Catholic Charities Fort Worth (CCFW) to address the diverse barriers faced by families in poverty. The program targets working-age adults who are able and willing to work but face significant barriers to self-sufficiency.

Head Start Impact Study. Kline and Walters (2016) analyzes the Head Start Impact Study (HSIS), a randomized controlled trial from 2002 to 2006 that assigned roughly 5,000 three- and four-year-old children to Head Start programs across the United States. Head

Start is the largest early childhood education program in the country and provides education, health, and nutrition services to disadvantaged children and their families.

Wisconsin Scholar Grant to Low-Income College Students. The Wisconsin Scholar Grant (WSG) is a privately funded grant providing low-income students $3,500 per year for up to five years. Hendren and Sprung-Keyser (2020) reanalyzes Goldrick-Rab, Kelchen, Harris and Benson (2016), who exploited a randomization of WSG offers to study their impact on degree completion and college persistence for the first three cohorts of the program (2008–2010).

Housing Vouchers in Chicago. Jacob and Ludwig (2012) estimates the impact of receiving a housing voucher from the Chicago Public Housing Authority lotteries.

Oregon Health Insurance Experiment (Provided to Single Adults). In 2008, Medicaid expansion in Oregon was oversubscribed and therefore allocated by lottery, creating random variation in Medicaid coverage. Estimates are based on Finkelstein, Taubman, Wright, Bernstein, Gruber, Newhouse, Allen, Baicker and Oregon Health Study Group (2012).

Income Maintenance Experiment in Seattle and Denver. The Seattle-Denver Income Maintenance Experiment was a randomized trial providing lump-sum cash transfers to families in place of traditional welfare programs. Estimates are based on Price, Song et al. (2018).

Paycheck Plus: EITC to Adults without Dependents. Paycheck Plus provides an expanded EITC (known as a "Bonus") to low-income singles not traditionally eligible for large EITC benefits. The credit, worth up to $2,000 per year, was available over three years (covering tax years 2014–2016, with bonuses paid in 2015–2017). Estimates are based on Miller, Katz, Azurdia, Isen and Schultz (2017).

Work Advance. WorkAdvance is a workforce development program offering unemployed, low-income adults occupational skills training aimed at employment in targeted sectors such as IT, environmental remediation, transportation, health care, and manufacturing. The MVPF estimate is based on Hendra, Greenberg, Hamilton, Oppenheim, Pennington, Schaberg and Tessler (2016) and Schaberg (2017), who report results from a randomized evaluation of WorkAdvance conducted by four service providers in New York, Oklahoma, and Ohio from 2011 to 2013.

JobStart. The JOBSTART Demonstration (1985–1988) was patterned after the Job Corps vocational education program but operated in a nonresidential setting. It targeted low-skill, young school dropouts and was primarily funded through the 1982 Job Training Partnership

Act. Hendren and Sprung-Keyser (2020) estimates the MVPF of JOBSTART using the final report from Cave, Bos, Doolittle and Toussaint (1993).

Year Up. Year Up is a national job training program for 18- to 24-year-olds disconnected from work and school. Participants receive six months of full-time training in financial services and IT, followed by a six-month internship, as well as counseling, instructional supports, and a weekly stipend. The MVPF estimate is based on initial results from a randomized evaluation conducted by Abt Associates beginning in 2013, as reported in Fein and Hamadyk (2018).

Job Corps. Job Corps is the largest U.S. vocational education program, established in 1964 and administered by the U.S. Department of Labor. It provides job training and other services to at-risk 16- to 24-year-olds in a residential setting via a network of centers run by local public and private agencies Schochet, Burghardt and McConnell (2008). Hendren and Sprung-Keyser (2020) estimates an MVPF for Job Corps primarily using two papers resulting from the National Job Corps study (a randomized trial of roughly 80,000 applicants between 1994 and 1996): Schochet (2018), which reports earnings impacts 21 years later, and the short-run cost-benefit analysis in Schochet, Burghardt, McConnell et al. (2006).

Job Training Partnership Act, Adults. The National JTPA Study (1987) was a randomized experiment with 20,600 participants (both youth and adults) in 16 local programs funded by the 1982 Job Training Partnership Act (JTPA). Hendren and Sprung-Keyser (2020) derives their adult MVPF estimates from the cost-benefit analyses reported in Bloom, Orr, Bell, Cave, Doolittle, Lin and Bos (1997). Because of the wide variety of programs, the authors compute separate MVPFs for adults and youths, pooling the results for adult men and women.

Job Training Partnership Act, Youth. Same study as above, but focusing on youths.

References

Andrews, Isaiah and Jiafeng Chen, "Certified Decisions," 2025. Working Paper, available at https://arxiv.org/abs/2502.17830.

———, **Toru Kitagawa, and Adam McCloskey**, "Inference on winners," *The Quarterly Journal of Economics*, 2024, *139* (1), 305–358.

Armstrong, Timothy B. and Shu Shen, "Inference on optimal treatment assignment," *The Japanese Economic Review*, 2023, *74* (4), 471–500.

Athey, Susan and Stefan Wager, "Policy Learning with Observational Data," *Econometrica*, 2021, *89* (1), 133–161.

Banerjee, Abhijit, Arun G Chandrasekhar, Suresh Dalpath, Esther Duflo, John Floretta, Matthew O Jackson, Harini Kannan, Francine N Loza, Anirudh Sankar, Anna Schrimpf, and Maheshwor Shrestha, "Selecting the most effective nudge: Evidence from a large-scale experiment on immunization," *Econometrica*, 2024. forthcoming, available at https://www.econometricsociety.org/publications/econometrica/forthcoming-papers.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, and Ying Wei, "Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework," *Annals of statistics*, 2018, *46* (6B), 3643.

Ben-Michael, Eli, D. James Greiner, Kosuke Imai, and Zhichao Jiang, "Safe Policy Learning through Extrapolation: Application to Pre-Trial Risk-Assessment," *Journal of the American Statistical Association*, 2025. forthcoming, available at https://www.tandfonline.com/doi/full/10.1080/01621459.2025.2489135.

Bloom, Howard S, Larry L Orr, Stephen H Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M Bos, "The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study," *Journal of human resources*, 1997, pp. 549–576.

Cave, George, Hans Bos, Fred Doolittle, and Cyril Toussaint, "JOBSTART: Final Report on a Program for School Dropouts," Technical Report, Manpower Demonstration Research Corporation 1993.

Chernozhukov, Victor, Denis Chetverikov, and Yuta Koike, "Nearly optimal central limit theorem and bootstrap approximations in high dimensions," *The Annals of Applied Probability*, 2023, *33* (3), 2374–2425.

_______, _______, Kengo Kato, and Yuta Koike, "Improved central limit theorem and bootstrap approximations in high dimensions," *The Annals of Statistics*, 2022, *50* (5), 2562–2586.

_______, _______, _______, and _______, "High-dimensional data bootstrap," *Annual Review of Statistics and Its Application*, 2023, *10* (1), 427–449.

_______, Sokbae Lee, and Adam M. Rosen, "Intersection Bounds: Estimation and Inference," *Econometrica*, 2013, *81* (2), 667–737.

Christensen, Timothy, Hyungsik Roger Moon, and Frank Schorfheide, "Optimal Decision Rules when Payoffs are Partially Identified," 2023. Working Paper, available at https://arxiv.org/abs/2204.11748.

Dehejia, Rajeev H., "Program evaluation as a decision problem," *Journal of Econometrics*, 2005, *125* (1-2), 141–173.

Demirer, Mert, Vasilis Syrgkanis, Greg Lewis, and Victor Chernozhukov, "Semi-Parametric Efficient Policy Learning with Continuous Actions," in "Advances in Neural Information Processing Systems (NeurIPS)," Vol. 32 2019. https://proceedings.neurips.cc/paper/2019/file/08b7dc6e8b36bcaac15847827b7951a9-Paper.pdf.

Doksum, Kjell A and Albert Y Lo, "Consistent and robust Bayes procedures for location based on partial information," *The Annals of Statistics*, 1990, pp. 443–453.

Dupas, Pascaline and Jonathan Robinson, "Why Don't the Poor Save More? Evidence from Health Savings Experiments," *American Economic Review*, June 2013, *103* (4), 1138–71.

Espinosa, Javier, William N Evans, David C Phillips, and Tim Spilde, "How Do Holistic Wrap-Around Anti-Poverty Programs Affect Employment and Individualized Outcomes?," Technical Report, National Bureau of Economic Research 2024.

Evans, William N, Shawna Kolka, James X Sullivan, and Patrick S Turner, "Fighting poverty one family at a time: Experimental evidence from an intervention with holistic, individualized, wrap-around services," Technical Report, National Bureau of Economic Research 2023.

Fein, David and Jill Hamadyk, "Bridging the Opportunity Divide for Low-Income Youth: Implementation and Early Impacts of the Year Up Program. Pathways for Advancing Careers and Education. OPRE Report 2018-65.," *Office of Planning, Research and Evaluation*, 2018.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group, "The Oregon health insurance experiment: evidence from the first year," *The Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.

Foster, Dylan J. and Vasilis Syrgkanis, "Orthogonal Statistical Learning," *The Annals of Statistics*, 2023, *51* (3), 879–908.

García, Jorge Luis and James Joseph Heckman, "Three criteria for evaluating social programs," *Journal of Benefit-Cost Analysis*, 2022, *13* (3), 281–286.

Goldrick-Rab, Sara, Robert Kelchen, Douglas N Harris, and James Benson, "Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion," *American Journal of Sociology*, 2016, *121* (6), 1762–1817.

Gupta, Shanti S and Klaus J Miescke, *On the problem of finding the largest normal mean under heteroscedasticity*, Purdue University. Department of Statistics, 1986.

______ and ______, "On finding the largest normal mean and estimating the selected mean," *Sankhyā: The Indian Journal of Statistics, Series B*, 1990, pp. 144–157.

Hendra, Richard, David H Greenberg, Gayle Hamilton, Ari Oppenheim, Alexandra Pennington, Kelsey Schaberg, and Betsy L Tessler, "Encouraging evidence on a sector-focused advancement strategy: Two-year impacts from the WorkAdvance demonstration," *New York: MDRC*, 2016.

Hendren, Nathaniel, "Measuring economic efficiency using inverse-optimum weights," *Journal of public Economics*, 2020, *187*, 1–13.

______ **and Ben Sprung-Keyser**, "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.

______ **and** ______ , "The Policy Impacts Library," 2021. `https://policyimpacts.org/policy-impacts-library/` Accessed: 2025-01-29.

Hirano, Keisuke and Jack Porter, "Asymptotic Analysis of Statistical Decision Rules in Econometrics," in Steven Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa Matzkin, eds., *The Handbook of Econometrics*, Vol. 7a, Elsevier, 2020, pp. 283–354.

______ **and Jack R. Porter**, "Asymptotics for Statistical Treatment Rules," *Econometrica*, 2009, *77* (5), 1683–1701.

Ishihara, Takuya and Toru Kitagawa, "Evidence Aggregation for Treatment Choice," 2021. Working Paper, available at `https://arxiv.org/abs/2108.06473`.

Jacob, Brian A and Jens Ludwig, "The effects of housing assistance on labor supply: Evidence from a voucher lottery," *American Economic Review*, 2012, *102* (1), 272–304.

Kido, Daido, "Locally Asymptotically Minimax Statistical Treatment Rules Under Partial Identification," 2023. Working Paper, available at `https://arxiv.org/abs/2311.08958`.

Kitagawa, Toru and Aleksey Tetenov, "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 2018, *86* (2), 591–616.

______**, Sokbae Lee, and Chen Qiu**, "Treatment Choice with Nonlinear Regret," 2022. Working Paper, available at `https://arxiv.org/abs/2205.08586`.

Kline, Patrick and Christopher R Walters, "Evaluating public programs with close substitutes: The case of Head Start," *The Quarterly Journal of Economics*, 2016, *131* (4), 1795–1848.

Liese, F. and K.J. Miescke, *Statistical Decision Theory: Estimation, Testing, and Selection* Springer Series in Statistics, Springer New York, 2008.

Manski, Charles F., "Identification and Decision Problems Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choide," *Journal of Econometrics*, 2000, *95* (2), 415–442.

______ , "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 2004, *72* (4), 1221–1246.

______ , *Public Policy in an Uncertain World*, Cambridge, MA: Harvard University Press, 2013.

———, "Communicating Uncertainty in Policy Analysis," *Proceedings of the National Academy of Sciences*, 2019, *116* (16), 7634–7641.

———, "Econometrics for Decision Making: Building on Foundations Sketched by Wald and Haavelmo," *Econometrica*, 2021, *89* (6), 2827–2853.

———, *Discourse on Social Planning Under Uncertainty*, Cambridge, UK: Cambridge University Press, 2024.

Maurer, Andreas and Massimiliano Pontil, "Empirical Bernstein Bounds and Sample-Variance Penalization," in "Proceedings of the 22nd Annual Conference on Learning Theory (COLT)" 2009.

Mbakop, Eric and Max Tabord-Meehan, "Model Selection for Treatment Choice: Penalized Welfare Maximization," *Econometrica*, 2021, *89* (2), 825–848. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16437.

Miller, Cynthia, Lawrence F Katz, Gilda Azurdia, Adam Isen, and Caroline B Schultz, "Expanding the earned income tax credit for workers without dependent children: Interim findings from the paycheck plus demonstration in new york city," *New York: MDRC, September*, 2017.

Montiel Olea, Jose Luis, Chen Qiu, and Jörg Stoye, "Decision Theory for Treatment Choice Problems with Partial Identification," 2023. Working Paper, available at `https://arxiv.org/abs/2312.17623`.

Moon, Sarah, "Optimal Policy Choices Under Uncertainty," 2025. Working Paper, available at `https://arxiv.org/abs/2503.03910`.

Newey, Whitney K., "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 1994, *62* (6), 1349–1382.

Ponomarev, Kirill and Vira Semenova, "On the Lower Confidence Band for the Optimal Welfare," 2024. Workng Paper, available at `https://arxiv.org/abs/2410.07443`.

Price, David J, Jae Song et al., "The long-term effects of cash assistance," *Industrial Relations Section working paper*, 2018, *621.*

Quintas-Martinez, Victor, "Finite-Sample Guarantees for High-Dimensional DML," *arXiv preprint arXiv:2206.07386*, 2022.

Rai, Yoshiyasu, "Statistical Inference for Treatment Assignment Policies," 2019. Working Paper, available at `https://sites.google.com/site/yoshiyasurai/home`.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf, "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

Russell, Thomas, "Policy Transforms and Learning Optimal Policies," 2020. Working Paper, available at `https://arxiv.org/abs/2012.11046`.

Salisbury, Adam, "How We Plan to Approach Uncertainty in Our Cost-Effectiveness Models," January 2024. available at `https://forum.effectivealtruism.org/posts/NWmBkMe3yF4GQNnai/how-we-plan-to-approach-uncertainty-in-our-cost`, assessed on 2024-11-30.

Schaberg, Kelsey, "Can Sector Strategies Promote Longer-Term Effects? Three-Year Impacts from the WorkAdvance Demonstration.," *MDRC*, 2017.

Schilbach, Frank, "Alcohol and Self-Control: A Field Experiment in India," *American Economic Review*, 2019, *109* (4), 1290–1322.

Schochet, Peter Z, "National job corps study: 20-year follow-up study using tax data," *Princeton, NJ: Mathematica Policy Research*, 2018.

________, **John A Burghardt, Sheena M McConnell et al.**, *National Job Corps study and longer-term follow-up study: impact and benefit-cost findings using survey and summary earnings records data*, US Department of Labor, Employment and Training Administration, 2006.

________, **John Burghardt, and Sheena McConnell**, "Does job corps work? Impact findings from the national job corps study," *American Economic Review*, 2008, *98* (5), 1864–1886.

Stoye, Jörg, "Minimax regret treatment choice with finite samples," *Journal of Econometrics*, 2009, *151* (1), 70–81.

Sun, Liyang, "Empirical Welfare Maximization with Constraints," 2021. Working Paper, available at `https://arxiv.org/abs/2103.15298`.

Swaminathan, Adith and Thorsten Joachims, "Counterfactual Risk Minimization: Learning from Logged Bandit Feedback," in "Proceedings of the 32nd International Conference on Machine Learning (ICML)" PMLR 2015, pp. 814–823.

Tetenov, Aleksey, "Statistical treatment choice based on asymmetric minimax regret criteria," *Journal of Econometrics*, 2012, *166* (1), 157–165.

Yata, Kohei, "Optimal Decision Rules Under Partial Identification," 2021. Working Paper, available at `https://arxiv.org/abs/2111.04926`.