Estimating the Value of Evidence-Based Decision Making

Alberto Abadie   Anish Agarwal   Guido Imbens   Siwei Jia
        MIT   Columbia   Stanford   Amazon.com

James McQueen   Serguei Stepaniants   Santiago Torres
Amazon.com   Amazon.com   MIT

February 8, 2026

Abstract

In an era of data abundance, statistical evidence is increasingly critical for business and policy decisions. Yet, organizations lack empirical tools to assess the value of evidence-based decision making (EBDM), optimize statistical precision, and balance the costs of evidence-gathering strategies against their benefits. To tackle these challenges, this article introduces an empirical framework to estimate the value of EBDM and evaluate the return on investment in statistical precision and project ideation. The framework leverages parametric and nonparametric empirical Bayes methods to account for parameter heterogeneity and measure how statistical precision changes the value of evidence. The value extracted from statistical evidence depends critically on how organizations translate evidence into policy decisions. Commonly used decision rules based on statistical significance can leave substantial value unrealized and, in some cases, generate negative expected value.

1. Introduction

Many organizations use randomized experiments and observational studies to improve their decision making. For example, Gupta et al. (2019) write "Together [Airbnb, Amazon, Booking.com, Facebook, Google, LinkedIn, Lyft, Microsoft, Netflix, Twitter, Uber, Yandex, and Stanford University] have tested more than one hundred thousand experiment treatments last year." The fact that so many organizations conduct such a large number of experiments suggests that these organizations believe that data evidence provides significant value in guiding business and policy decisions. However, we are unaware of empirical tools that organizations can use to assess the actual value of their EBDM practices. In the absence of such tools, it is difficult to determine whether too much experimentation is being conducted or too little, whether experiments are too large or too small, and whether the right experiments are being undertaken. Part of the challenge in evaluating the value of EBDM lies in the need to describe the role of evidence in the business and policy decision-making process. In other words, estimating the value of EBDM requires assumptions about what organizations will do with and without various amounts of evidence, which they can choose to generate at some cost.

In this article, we propose an empirical Bayes estimator for the value of EBDM. We study the problem of a decision maker choosing whether to adopt a particular policy intervention. We use the term "agent" to refer to the decision maker, and "policy," "intervention," and "treatment" interchangeably to refer to the policy intervention under scrutiny. The agent can implement the intervention based on prior information or gather additional information at some cost—for example, by running an experimental or observational evaluation of the intervention's effect. At the stage where the agent decides whether to implement the intervention, they aim to maximize utility based on the available information. We derive expressions for the value of additional information and demonstrate how to estimate this value using metadata on estimates of the effects of business and policy interventions, along with their standard errors.

Our framework allows decision makers to assess in a principled way the value of experi-

mental and non-experimental studies, and how design choices affect that value. Currently, many organizations decide on the precision of their studies based on power calculations. These do not take into account the costs and benefits of EBDM and instead rely on statistical conventions (for example, requiring 80% power for tests at the 5% level). Additionally, our framework enables decision makers to *ex ante* assess whether an experiment is worth conducting based on its cost and expected benefits (with benefits increasing with the decision maker's initial uncertainty about the intervention's effect).

Related literature.—This article builds on foundational work by Blackwell (1951) and Howard (1966) on the value of information. It extends that framework to the applied domain of EBDM, combining costs, precision, and empirical Bayes estimation into a flexible tool for real-world decisions and counterfactual analysis. The resulting methods are particularly suited to data-rich environments, where organizations must balance the benefits of additional information against the costs of generating it.

The empirical setting is also closely related to that of meta-analytic studies (Hedges and Olkin, 1985; Higgins et al., 2019) in that it leverages information from many individual studies. However, unlike meta-analysis, the goal of this article is not to assess the effectiveness of a set of policies but to quantify the value brought by empirical evidence in guiding better policy decisions.

A key component of the EBDM estimand is the expected value of the positive part of the predicted policy payoff, that is, the expectation of the maximum of the predicted policy payoff and zero. A related but distinct object is considered in Semenova (2023) in the context of estimating the size of a latent population whose outcomes are observed regardless of treatment exposure. The estimand in Semenova (2023) targets the distribution of an expectation given observed covariates, and does so in a single study setting. In contrast, our estimand pertains to the distribution of predicted policy payoffs across many studies.

This article also contributes to the growing literature on empirical Bayes methods (see Morris, 1983; Efron, 2010, for foundational work on empirical Bayes). Empirical Bayes and related shrinkage techniques play a central role in applied economics, informing research

on teacher and school value-added (Chetty et al., 2014; Angrist et al., 2017), neighborhood effects (Chetty and Hendren, 2018), income dynamics (Gu and Koenker, 2017), and racial discrimination (Kline et al., 2024), among other topics. As empirical Bayes methods gain traction in applied economics, a parallel methodological literature has emerged in econometrics (Koenker and Mizera, 2014; Abadie and Kasy, 2019; Fessler and Kasy, 2019; Armstrong et al., 2022; Kwon, 2023; Koenker and Gu, 2024; Chen, 2024). Walters (2024) provides a comprehensive account of empirical Bayes methods and their applications in economics. This article applies both parametric and nonparametric empirical Bayes techniques to estimate the distribution of policy payoffs in settings where the data contain information about the effects of many policies.

2. The value of EBDM

The notation $X \sim (\theta, \sigma^2)$ indicates that the random variable X has mean θ and variance σ^2. When X follows a Gaussian distribution with mean θ and variance σ^2, it is denoted as $X \sim N(\theta, \sigma^2)$. $f_X(\cdot)$ represents a probability density function of the random variable X, while $f_{X|W}(\cdot|w)$ denotes a conditional probability density function of X given $W = w$. $F_X(\cdot)$ denotes the cumulative distribution function of X, and $F_{X|W}(\cdot \mid w)$ is the cumulative distribution function of X given $W = w$. The functions $\phi(\cdot)$ and $\Phi(\cdot)$ refer to the probability density function and cumulative distribution function of the standard Gaussian distribution, respectively.

2.1. Setup

Consider the problem of a risk-neutral decision maker tasked with choosing whether to adopt a particular policy for a population of units. The *ex ante* unknown per-unit payoff of the policy, τ, follows a distribution with known probability density function $f_\tau(\cdot)$ and mean $\mu = E[\tau]$. In an organization where teams generate new ideas for policies or interventions, $f_\tau(\cdot)$ can be thought of as the distribution of the quality of those ideas. For retrospective estimation tasks, we take this distribution as fixed. However, organizations can shift it, for example, by prioritizing high-risk projects with substantial upside, protecting agents from

failure, or allocating resources to exploratory projects.

In the absence of additional information, the agent launches the policy if the expected payoff from launching is positive,

$$\mu - c_L > 0,$$

where c_L is the cost of launching per-unit. The expected value of this decision is $\max\{\mu - c_L, 0\}$.

Now, suppose the agent has the option to obtain additional information about the policy payoff at some cost. Specifically, the agent can acquire a signal, $\hat{\tau}$, distributed as

$$\hat{\tau} \mid \tau, \sigma^2 \sim N(\tau, \sigma^2), \tag{1}$$

at a cost of $c_F + c(\sigma^2)$, where $c_F \geqslant 0$, $c(\cdot) \geqslant 0$, and $c'(\cdot) \leqslant 0$. This setup captures the information obtained from studies estimating policy effects using experimental or observational data. The constant c_F reflects the fixed cost of conducting a data-driven policy evaluation, while the function $c(\cdot)$ captures the cost of precision, which partly depends on the study's sample size. The restriction on the derivative $c'(\cdot)$ indicates that obtaining more precise information is weakly more expensive. The assumption of Gaussianity for the distribution of $\hat{\tau} \mid \tau, \sigma^2$ is motivated by the approximate Gaussian nature of the large-sample distributions of many commonly used estimators of treatment effects.

After observing the signal $\hat{\tau}$, the expected payoff of the policy is

$$
\begin{aligned}
E[\tau | \hat{\tau} = t] &= \int u f_{\tau | \hat{\tau}}(u|t) du \\
&= \frac{\int \dfrac{u}{\sigma} \phi((t-u)/\sigma) f_\tau(u) du}{\int \dfrac{1}{\sigma} \phi((t-u)/\sigma) f_\tau(u) du}.
\end{aligned}
$$

If a signal is observed, the agent launches the policy if

$$E[\tau | \hat{\tau}] - c_L > 0.$$

For any set $\mathcal{A}$, let $I_{\mathcal{A}}(x)$ be the function that takes value one if $x \in \mathcal{A}$, and value zero otherwise. The expected payoff with EBDM for a fixed value of the variance of the signal

5

σ^2 is

$$V(\sigma^2) = E\left[I_{(0,\infty)}(E[\tau|\hat{\tau}] - c_L)(\tau - c_L)\right]$$
$$= E\left[I_{(0,\infty)}(E[\tau|\hat{\tau}] - c_L)(E[\tau|\hat{\tau}] - c_L)\right]$$
$$= E\left[\max\{E[\tau|\hat{\tau}] - c_L, 0\}\right].$$

Define $V(\infty)$ as the expected payoff with no information beyond the distribution of τ, that is,

$$V(\infty) = \max\{\mu - c_L, 0\}.$$

Because $\max\{x, 0\}$ is a convex function of x, Jensen's inequality implies,

$$V(\sigma^2) \geqslant \max\{\mu - c_L, 0\} = V(\infty).$$

The value of evidence (VoE) is the difference in expected payoffs $V(\sigma^2)$ and $V(\infty)$, which is nonnegative, minus the cost of acquiring the information, which is generally positive:

$$\text{VoE}(\sigma^2) = V(\sigma^2) - V(\infty) - (c_F + c(\sigma^2)).$$

A second version of VoE, which we term VoID (for Value of Information under Default adoption) is obtained when, in the absence of additional information about the effect of the intervention, the intervention is always deployed and so $V(\infty) = \mu - c_L$:

$$\text{VoID}(\sigma^2) = V(\sigma^2) - (\mu - c_L) - (c_F + c(\sigma^2)).$$

$\text{VoID}(\sigma^2)$ is motivated by settings with ex-ante (pre-evaluation) ambiguity on the distribution of τ, and agents who have a bias for action in the presence of such ambiguity.

2.2. A motivating example

A common instance of the setting described above is one where the decision maker obtains experimental evidence on the effect of the policy. Consider an experiment with N units: $i = 1, \ldots, N$. The experimenter assigns N_1 units at random to treatment and the remaining $N_0 = N - N_1$ to control. If unit i is treated, an outcome is drawn

$$Y_i(1) \sim (\theta_1, \sigma_1^2).$$

If unit i is untreated, the outcome is drawn

$$Y_i(0) \sim (\theta_0, \sigma_0^2).$$

Let W_i be an indicator of treatment for unit i. We observe $Y_i = Y_i(1)W_i + Y_i(0)(1 - W_i)$. The average effect of the treatment is

$$\tau = \theta_1 - \theta_0.$$

A simple estimator of τ is the difference in mean outcomes between treated and nontreated,

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} W_i Y_i - \frac{1}{N_0} \sum_{i=1}^{N} (1 - W_i) Y_i.$$

Then, for large N_0 and N_1, equation (1) holds approximately, with

$$\sigma^2 = \frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}.$$

When a fraction $p = N_1/N$ of units are assigned to treatment, and assuming that the only variable cost of the experiment comes from recruiting subjects at a cost κ per subject, the total cost of the experiment is

$$c_F + \frac{\kappa}{\sigma^2} \left(\frac{\sigma_1^2}{p} + \frac{\sigma_0^2}{1 - p} \right),$$

where c_F is the fixed cost of the experiment.

Neyman's allocation rule, $p/(1 - p) = \sigma_1/\sigma_0$, minimizes the variance σ^2 for a fixed total number of experimental subjects, N. When this rule is applied to allocate subjects between a treatment and a control group, the cost of the experiment becomes

$$c_F + \kappa \frac{(\sigma_1 + \sigma_0)^2}{\sigma^2}.$$

In some settings, researchers favor treatment effect parameters free of units of measurement, such as lift $\tau = (\theta_1 - \theta_0)/\theta_0$. Let

$$\hat{\tau} = \frac{\dfrac{1}{N_1} \sum_{i=1}^{N} W_i Y_i - \dfrac{1}{N_0} \sum_{i=1}^{N} (1 - W_i) Y_i}{\dfrac{1}{N_0} \sum_{i=1}^{N} (1 - W_i) Y_i}.$$

7

Then, for large N_0 and N_1, equation (1) holds with

$$\sigma^2 = \frac{1}{\theta_0^2} \left(\frac{\sigma_1^2}{N_1} + (1+\tau)^2 \frac{\sigma_0^2}{N_0} \right).$$

For values of the lift parameter close to zero, as is common in many online experimentation settings, we can approximate

$$\sigma^2 \approx \frac{1}{\theta_0^2} \left(\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0} \right).$$

In this case, Neyman allocation yields,

$$c_F + \kappa \frac{(\sigma_1 + \sigma_0)^2}{\theta_0^2 \sigma^2},$$

2.3. A Gaussian distribution for τ

This section derives a simple closed-form expression for $V(\sigma^2)$ under the assumption that the distribution of τ is Gaussian,

$$\tau \sim N(\mu, \gamma^2). \tag{2}$$

While equation (1) is supported by the Central Limit Theorem in studies with large samples, equation (2) imposes two important restrictions. First, τ is a Gaussian random variable. Second, implicit in the notation is the assumption that the distribution of τ is independent of σ^2. The first is a strong parametric restriction. The Gaussian approximation for τ could be valid in some settings but questionable in others. The second restriction could be violated, for example, if researchers adapt the power of individual studies to take into account prior information about the effect on the treatment. We dispose of these two restrictions later in the article. We adopt them in this section, however, to obtain closed-form formulas for the value of EBDM in a simple setting.

If equations (1) and (2) hold, the marginal distribution of $\hat{\tau}$ is $\hat{\tau} \sim N(\mu, \gamma^2 + \sigma^2)$. The posterior for τ is given by

$$\tau \mid \hat{\tau} \sim N \left(\frac{\mu/\gamma^2 + \hat{\tau}/\sigma^2}{1/\gamma^2 + 1/\sigma^2}, \frac{1}{1/\gamma^2 + 1/\sigma^2} \right).$$

The expected payoff with EBDM is

$$V(\sigma^2) = E\left[\max\left\{\frac{\mu/\gamma^2 + \hat{\tau}/\sigma^2}{1/\gamma^2 + 1/\sigma^2} - c_L, 0\right\}\right].$$

Let

$$Z = \frac{\mu/\gamma^2 + \hat{\tau}/\sigma^2}{1/\gamma^2 + 1/\sigma^2} - c_L.$$

Recall that the marginal distribution of $\hat{\tau}$ is Gaussian with mean μ and variance $\gamma^2 + \sigma^2$. As a result,

$$Z \sim N\left(\mu - c_L, \frac{\gamma^4}{\gamma^2 + \sigma^2}\right). \tag{3}$$

Now, $V(\sigma^2)$ is the the first moment of the Gaussian distribution in (3) censored from below at zero,

$$V(\sigma^2) = (\mu - c_L)\Phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) + \frac{\gamma^2}{\sqrt{\gamma^2 + \sigma^2}}\phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right). \tag{4}$$

The derivatives of $V(\sigma^2)$ with respect to σ^2, γ^2, and μ are

$$\frac{\partial V(\sigma^2)}{\partial \sigma^2} = -\frac{\gamma^2}{2(\gamma^2 + \sigma^2)^{3/2}}\phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) \leqslant 0,$$

$$\frac{\partial V(\sigma^2)}{\partial \gamma^2} = \frac{\gamma^2 + 2\sigma^2}{2(\gamma^2 + \sigma^2)^{3/2}}\phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) \geqslant 0,$$

and

$$\frac{\partial V(\sigma^2)}{\partial \mu} = \Phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) \geqslant 0.$$

Higher precision of the signal $\hat{\tau}|\tau$ increases the expected payoff from EBDM. In addition, the value of experimentation increases when an organization increases the variance of the distribution of true effects—that is, the variance of idea quality. The derivative of V with respect to μ implies

$$\frac{\partial \text{VoE}}{\partial \mu} = \Phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) - I_{(0,\infty)}(\mu - c_L),$$

and

$$\frac{\partial \text{VoID}}{\partial \mu} = \Phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) - 1.$$

9

VoE peaks at $\mu = c_L$ and decreases monotonically with $|\mu - c_L|$. When $|\mu - c_L|$ is large, a simple rule that selects policies based solely on the sign of $\mu - c_L$ gets most decisions right, leaving little room for additional evidence to add value. VoID, the value of EBDM when policies are adopted by default in the absence of additional information, decreases monotonically with μ. VoID is particularly large when the distribution of τ is concentrated on negative values, as additional information winnows out many ineffective or counterproductive policies.

Notice that

$$\lim_{\sigma^2 \to \infty} V(\sigma^2) = \max\{\mu - c_L, 0\} = V(\infty)$$

and

$$\lim_{\sigma^2 \to 0} V(\sigma^2) = (\mu - c_L)\Phi\left(\frac{\mu - c_L}{\gamma}\right) + \gamma\phi\left(\frac{\mu - c_L}{\gamma}\right).$$

As $\sigma^2 \to \infty$, we lose any additional information about the value of τ beyond its distribution, and $V(\sigma^2)$ converges to $V(\infty)$. As $\sigma^2 \to 0$, the information gathering process reveals the value of τ. In this case, $V(\sigma^2)$ converges to $E[\max\{\tau - c_L, 0\}]$, the mean of the distribution of $\tau - c_L$ censored at zero.

So far, we have treated σ^2 as a constant. We now allow σ^2 to have a non-degenerate distribution, independent of τ. In this case, the average payoff of EBDM is

$$\begin{aligned}
V &= E\left[\max\{E[\tau|\hat{\tau}, \sigma] - c_L, 0\}\right] \\
&= E\left[E\left[\max\{E[\tau|\hat{\tau}, \sigma] - c_L, 0\}|\sigma\right]\right] \\
&= E\left[(\mu - c_L)\Phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right) + \frac{\gamma^2}{\sqrt{\gamma^2 + \sigma^2}}\phi\left(\frac{\mu - c_L}{\gamma^2/\sqrt{\gamma^2 + \sigma^2}}\right)\right],
\end{aligned} \tag{5}$$

with the expectation taken over the distribution of σ^2.

2.4. A Gaussian mixture distribution for τ

In Section 4 we use a Gaussian mixture distribution to evaluate the effects of misspecification of the distribution of τ on EBDM value estimates. Suppose τ follows a mixture of k Gaussian distributions with parameters $(\mu_1, \gamma_1^2), \ldots, (\mu_k, \gamma_k^2)$, and mixture probabilities $p_1, \ldots, p_k$. Let

$\hat{\tau} = \tau + \varepsilon$, where ε is independent Gaussian noise with variance σ^2. Conditional on $\tau \sim N(\mu_j, \gamma_j^2)$, we have

$$E[\tau | \hat{\tau}, \tau \sim N(\mu_j, \gamma_j^2)] = \frac{\mu_j/\gamma_j^2 + \hat{\tau}/\sigma^2}{1/\gamma_j^2 + 1/\sigma^2}.$$

As a result,

$$E[\tau | \hat{\tau}] = \sum_{j=1}^{k} E[\tau | \hat{\tau}, \tau \sim N(\mu_j, \gamma_j^2)] \Pr(\tau \sim N(\mu_j, \gamma_j^2) | \hat{\tau})$$

$$= \frac{\sum_{j=1}^{k} \left(\frac{\mu_j/\gamma_j^2 + \hat{\tau}/\sigma^2}{1/\gamma_j^2 + 1/\sigma^2} \right) \frac{1}{\sqrt{\gamma_j^2 + \sigma^2}} \phi \left(\frac{\hat{\tau} - \mu_j}{\sqrt{\gamma_j^2 + \sigma^2}} \right) p_j}{\sum_{j=1}^{k} \frac{1}{\sqrt{\gamma_j^2 + \sigma^2}} \phi \left(\frac{\hat{\tau} - \mu_j}{\sqrt{\gamma_j^2 + \sigma^2}} \right) p_j}. \tag{6}$$

The simulations in Section 4 use the Gaussian mixture model to capture deviations from normality in the distribution of τ.

3. Empirical Bayes estimation

In this section, we analyze a setting with n realizations from the distribution of $(\tau, \sigma, \hat{\tau}, \hat{\sigma})$, where only the estimates $\hat{\tau}$ and $\hat{\sigma}$ are observed. We use the observations on $\hat{\tau}$ and $\hat{\sigma}$ to estimate μ and γ^2 using an Empirical Bayes strategy. $(\hat{\tau}, \hat{\sigma})$ represent point estimates and their corresponding standard errors for a set of policy evaluations in the dataset. We examine both the homoskedastic case, where σ^2 is constant, and the heteroskedastic case, where $\text{var}(\sigma^2) > 0$. Throughout our analysis, we approximate the per-unit launch cost as $c_L \approx 0$. Alternatively, we can interpret τ as representing the net benefits of the treatment after accounting for the launch cost.

We employ a database of thousands of online experiments run by Upworthy to illustrate the applicability of our methods. Upworthy is a U.S. online news and media publisher that built a large following in the 2010s by pairing positive, uplifting stories with optimized headline-and-image packages designed to drive clicks and shares. Upworthy pioneered large-scale A/B testing of these packages, routinely randomizing visitors across alternative

headlines and images for the same article preview and using click-through performance to guide editorial and distribution choices.

The Upworthy Research Archive records thousands of randomized experiments run by Upworthy between January 24, 2013 and April 30, 2015. We restrict our analysis to the Upworthy Exploratory Dataset, which contains outcomes for 4,873 online experiments. Each experiment compared alternative headline-and-image packages for the same article preview. We analyze outcomes for the first two packages deployed in each experiment, labeling the first as the control arm and the second as the treatment arm. For each package, the archive reports impressions and clicks. We remove from the sample all experiments with fewer than 100 impressions in one of the experimental arms, which yields a sample of $n = 4{,}857$ online experiments. We then compute each experimental arm's click rate per thousand impressions and define $\hat{\tau}$ as the treatment–control difference in those rates. We also compute the standard error of $\hat{\tau}$.

Figure 1 shows the distribution of $\hat{\tau}$. The average value of $\hat{\tau}$ is -0.7621 (clicks per thousand impressions), with range $[-54.17, , 45.13]$. The standard errors have mean 2.9727 and range $[0.2202, , 7.942]$. For the remainder of this section, we use the Upworthy data to illustrate empirical Bayes estimation of the value of EBDM. Sections 5 through 7 delve deeper into the EBDM value estimates for the Upworthy dataset.

3.1. Parametric empirical Bayes

In this section, we adopt a Gaussian specification for the distribution of τ. Because $\hat{\tau}$ is unbiased, we can estimate μ, the mean of the distribution of τ, as the mean of $\hat{\tau}$ across evaluations. To estimate γ^2, the variance of the distribution of τ, we deconvolute the distribution of $\hat{\tau}$ as follows. By the Total Law of Variance,

$$\mathrm{var}(\hat{\tau}) = E\big[\mathrm{var}(\hat{\tau}|\tau)\big] + \mathrm{var}(E[\hat{\tau}|\tau]).$$

Unbiasedness of $\hat{\tau}$ conditional on τ implies

$$\gamma^2 = \mathrm{var}(\tau)$$

12

Figure 1: Distribution of estimated treatment effects in the Upworthy data
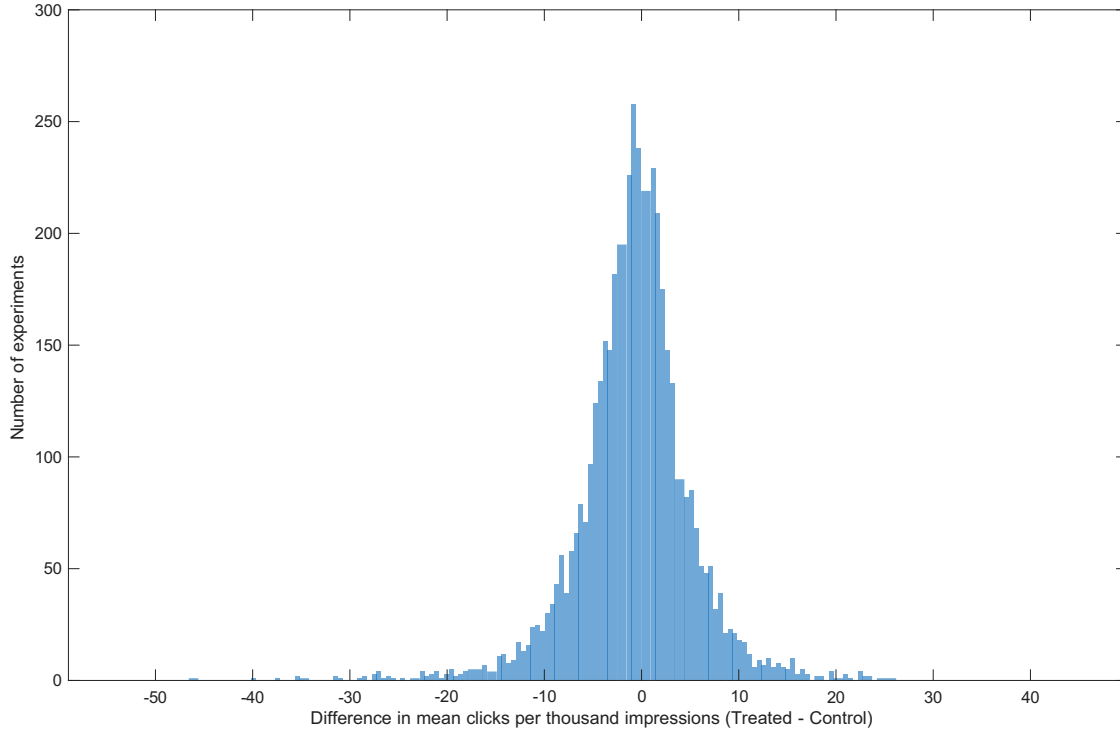
$$= \mathrm{var}(E[\hat{\tau}|\tau])$$

$$= \mathrm{var}(\hat{\tau}) - E[\mathrm{var}(\hat{\tau}|\tau)].$$

As a result, we define $\hat{\gamma}^2$ as the difference between the variance $\hat{\tau}$ across experiments in the data minus the mean of the squares of the standard errors. This estimator is not guaranteed to be non-negative (see Morris, 1983, for a discussion and alternative estimators). In the Upworthy dataset, $\hat{\gamma}^2 = 36.4608 - 10.2437 = 26.2171$.

3.1.1. Homoskedastic case

For the homoskedastic case, we estimate σ^2 as the average of the squares of the standard deviations of $\hat{\tau}$ across studies. In the Upworthy data, this estimate is 10.4919. Plugging in this value in (4) along with estimates of μ and γ^2, we obtain $V(10.2437) = 1.3691$, with the value of information measured in clicks per thousand impressions.

13

We now relax the assumption that σ^2 is constant. It can be shown (see appendix) that $V(\sigma^2)$ is convex. Then, by Jensen's inequality, $V(E[\sigma^2]) \leqslant E[V(\sigma^2)]$. This result implies that the assumption of homoskedasticity may lead to underestimation of the average payoff when σ^2 is not constant. Under heteroskedasticity, we evaluate the expression in (5) plugging in study-specific estimates of σ^2. Relative to the calculations in the previous section, now the value of experimentation is computed for each value of σ^2 and then integrated over the distribution of σ^2. An estimator of V based on a set of policy estimates can be calculated in two steps: *(i)* use the square of the standard error of $\hat{\tau}$ to approximate σ^2, and estimate the value of each study separately, and *(ii)* take the average over all studies in the sample. For the Upworthy data, this procedure yields $V = 1.4057$.

3.2. Nonparametric empirical Bayes

We next relax the parametric restriction $\tau \sim N(\mu, \gamma^2)$ of Section 3.1 and consider a non-parametric distribution for τ.

3.2.1. NPMLE under precision independence

Suppose $\tau \mid \sigma \sim G_\sigma$, where G_σ is an unspecified distribution, and

$$\hat{\tau} \mid (\tau, \sigma) \sim N(\tau, \sigma^2).$$

Under a precision-independence assumption (namely, that τ is independent of σ) we have $\tau \mid \sigma \sim G_\sigma = G$, so the distribution of τ does not depend on σ. It follows that for any $s > 0$, the conditional distribution of τ/σ given $\sigma = s$ coincides with the distribution of τ/s.

Given n studies with observed pairs $\{(\hat{\tau}_i, \sigma_i)\}_{i=1}^n$, we estimate the mixing distribution G using nonparametric empirical Bayes methods. A nonparametric maximum likelihood estimator (NPMLE) of G solves the problem

$$\max_{G \in \mathcal{G}} \sum_{i=1}^n \log \int \phi\left(\frac{\hat{\tau}_i - w}{\sigma_i}\right) dG(w), \tag{7}$$

where $\mathcal{G}$ denotes a class of discrete distributions supported on a fixed grid $u_1, \ldots, u_m$, with

probabilities $g_1, \ldots, g_m$ (see appendix for details). Modern implementations of NPMLE (e.g., Koenker and Gu, 2017) solve (7) efficiently and come with theoretical guarantees (Jiang, 2020; Soloff et al., 2025). The solution is computed over a grid $u_1, \ldots, u_m$ representing support points of G, with corresponding probabilities, $\widehat{g}_1, \ldots, \widehat{g}_m$.

Moreover, as shown in the appendix, the posterior mean satisfies

$$E[\tau | \widehat{\tau} = z, \sigma = s] = \frac{\int w \phi \left(\frac{z - w}{s} \right) dG(w)}{\int \phi \left(\frac{z - w}{s} \right) dG(w)}, \tag{8}$$

from which we derive a sample analog of $E[\tau | \widehat{\tau} = z, \sigma = s]$ as

$$\frac{\sum_{j=1}^{m} u_j \phi \left(\frac{z - u_j}{s} \right) \widehat{g}_j}{\sum_{j=1}^{m} \phi \left(\frac{z - u_j}{s} \right) \widehat{g}_j}.$$

For the Upworthy dataset, we approximate $\sigma_1, \ldots, \sigma_N$ using the reported standard errors, estimate G via the algorithm of Koenker and Mizera (2014) as implemented in Koenker and Gu (2017), and obtain $\widehat{V} = 0.7621$.

3.2.2. Relaxing the precision independence assumption

We relax the precision-independence assumption by partitioning the range of $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_n$ into five intervals and performing the NPEB calculations from the previous section within each interval. We refer to this approach as binning. Applied to the Upworthy data, binning yields $\widehat{V} = 0.9285$.

As an alternative, we use the CLOSE-NPMLE framework of Chen (2024), which models the conditional distribution of the estimates given their standard errors as a flexibly parameterized location–scale family and estimates the mixing distribution nonparametrically via NPMLE. For the Upworthy data, CLOSE-NPMLE yields $\widehat{V} = 0.9560$.

Table 1: Simulation results

distribution of τ:		true value	empirical Bayes parametric	empirical Bayes nonparametric
Gaussian	expected payoff	0.3970	0.3962	0.3963
	95% interval		$[0.3450, 0.4475]$	$[0.3443, 0.4483]$
mixture	expected payoff	0.3230	0.3939	0.3232
	95% interval		$[0.3121, 0.4757]$	$[0.2645, 0.3819]$

4. Simulations

We consider two data generating processes (DGP). In DGP1, the parameters τ have a standard Gaussian distribution $\tau \sim N(0, 1)$, so the parametric empirical Bayes model of Section 2.3 applies. In DGP2, the parameters τ follow the mixture distribution as in Section 2.4. In particular, in DGP2,

$$\tau \sim \begin{cases} N(-5, 1/2) & \text{with prob. } 0.01, \\ N(0, 1/2) & \text{with prob. } 0.98, \\ N(5, 1/2) & \text{with prob. } 0.01. \end{cases}$$

DGP1 and DGP2 both produce a distribution of τ with mean zero and variance one. We generate $\hat{\tau}$ as $\hat{\tau} = \tau + \sigma u$, where u is independent standard Gaussian and $\sigma = 0.1$. To calculate the expected payoff of EBDM, we consider the case of $c_L = 0$.

We run 1000 simulations for DGP1 and DGP2 with $n = 500$. Equation (4) with $\mu = 0$, $\gamma^2 = 1$, $\sigma = 0.1$, and $c_L = 0$ gives the true expected payoff of EBDM under DGP1. To calculate the true expected payoff of EBDM under DGP2, we first use equation (6) to compute $E[\tau|\hat{\tau}]$ over the $n \times 1000 = 500{,}000$ realizations of $\hat{\tau}$ in the simulations, and report the average of $\max\{E[\tau|\hat{\tau}], 0\}$. In each of the simulations, we calculate parametric and nonparametric empirical Bayes estimates of the average payoff of EBDM. The parametric empirical Bayes estimator is the sample analog of equation (5). This estimator is valid under the assumption that the true distribution of τ is Gaussian. The nonparametric empirical Bayes estimator is as in Section 3.2.1. For the simulations in this section, we treat σ^2 as known.

Table 1 reports the true values of the expected payoff of EBDM along with means and

95 percent intervals for the distribution of the estimates across simulations. When τ is Gaussian, the distributions of the parametric and nonparametric empirical Bayes estimates across simulations are both centered near the true value of the expected EBDM payoff. Moreover, there is no evidence of substantial gains from knowledge of the parametric form of the distribution of τ. The 95 percent interval for the nonparametric estimator is only 1.5 percent wider than the interval for the parametric estimator.

For the case when the distribution of τ is a mixture, the results for the parametric estimator reveal a clear bias, while the distribution of the nonparametric estimator remains centered at the true value of the expected payoff. Moreover, the 95 percent interval for the nonparametric estimator is 28.3 percent narrower than the interval for the parametric estimator.

5. Application to the Upworthy dataset

Table 2 reports parametric and nonparametric empirical Bayes estimates of the value of EBDM in the Upworthy data. In the parametric case, the table reports estimates computed under heteroskedasticity (Section 3.1.2). In the nonparametric case, it reports three estimates: the precision-independence NPMLE (Section 3.2.1), and binning and CLOSE-NPMLE estimates (Section 3.2.2) that relax the assumption of precision independence. Below each of the estimates of the value of EBDM, Table 2 reports 95 percent intervals computed over 1,000 bootstrap draws from the distribution of $(\hat{\tau}, \hat{\sigma}^2)$ in the data. In our calculations, we impose $c_L = c_F = c(\sigma^2) = 0$.

Because $\hat{\tau}$ has a negative mean and large dispersion relative to its mean, this is a setting where we expect to have substantial gains from EBDM. Indeed, the parametric model suggests a VoE of about 1.8 times the magnitude of $\hat{\mu}$ (but with a positive sign), while nonparametric empirical Bayes under the most restrictive specification yields a value only slightly larger than $\hat{\mu}$ in magnitude. The most flexible specifications (binning and CLOSE) deliver similar results, implying a VoE that is 25.4 percent larger than $\hat{\mu}$ in magnitude and with a positive sign and a VoID that is 125.4 percent larger than $\hat{\mu}$ in magnitude, again with a positive sign.

Table 2: Value of EBDM for the Upworthy data

Distribution of $\widehat{\tau}$:

$$\widehat{\mu} = -0.7622$$
(standard deviation of $\widehat{\tau}$ is 6.03)

Value of EBDM:

		parametric	nonparametric		
			precision independence	binning	CLOSE-NPMLE
		(1)	(2)	(3)	(4)
VoE	estimate	1.4057	0.7621	0.9285	0.9560
	95% interval	$[1.2792, 1.5321]$	$[0.6743, 0.8499]$	$[0.8290, 1.0280]$	$[0.8519, 1.0601]$
VoID	estimate	2.1678	1.5242	1.6907	1.7181
	95% interval	$[1.9906, 2.3451]$	$[1.3832, 1.6653]$	$[1.5495, 1.8319]$	$[1.5794, 1.8568]$

Note: The table reports estimates of the VoE and VoID for the Upworthy data ($n = 4{,}857$). For each method, the table presents the point estimate of the payoff and a dispersion interval obtained from 1,000 bootstrap samples. The parametric model is computed under a heteroskedasticity assumption.

6. Estimation of counterfactual EBDM values

This section provides estimates of the value of EBDM under alternative levels of statistical precision and under alternative levels of dispersion in the distribution of the effects of the policies.

First, we estimate how the value of EBDM would change as a result of a change in σ^2. In the resulting counterfactuals, the variance of the estimators is equal to the variance of $\widehat{\tau}_1 \,|\, \tau_1, \ldots, \widehat{\tau}_n \,|\, \tau_n$ in the original sample multiplied by λ. That is, $\lambda = 0.5$ represents a counterfactual scenario where the variances of the estimators are 50 percent smaller than the variance estimates in the original sample, while for $\lambda = 1.5$ the variances of the estimators are 50 percent larger than in the original sample.

For simplicity, we consider only counterfactual scenarios such that τ is independent of estimation variance, σ^2, and estimate the value of EBDM using the parametric empirical Bayes estimator of Section 3.1. It is conceptually straightforward to extend this procedure to more general settings (e.g., by modeling the dependence between τ and σ^2 and/or using

nonparametric empirical Bayes estimators).

For each estimate $i = 1, \ldots, n$ in our sample, we draw a value from the empirical Bayes estimate of the distribution of τ. Let $\tau_1^*, \ldots, \tau_n^*$ be the resulting values for the draws. Next, for $i = 1, \ldots, n$, we obtain $\hat{\tau}_i^* = \tau_i^* + \sigma_i^* U_i$, where $U_1, \ldots, U_n$ are independent draws from the standard Gaussian distribution, and $\sigma_i^* = \sqrt{\lambda} \hat{\sigma}_i$. We use the new sample $(\hat{\tau}_1^*, \hat{\sigma}_1^*), \ldots, (\hat{\tau}_n^*, \hat{\sigma}_n^*)$ to compute an estimate of the value of EBDM. We repeat this procedure multiple times to obtain the distribution of EBDM-value estimates for a particular value of λ. The average of this distribution is our estimate of the value of EBDM under variance modification factor λ. For the parametric empirical Bayes case, this average can also be computed directly using an empirical counterpart of equation (5) that applies the variance modification factor λ to σ^2.

Panel A of Figure 2 reports the results obtained from applying the procedure described above to the Upworthy data. The solid line represents the value of EBDM as a function of the variance modification factor, λ. The shaded area represents 95 percent intervals from the distribution of EBDM estimates. An investment that reduces estimation variance by half (about a 29.3 percent decrease in standard errors) leads to an increase in the value of EBDM by 8.33 percent, from 1.4057 to 1.5228. Conversely, an increase in estimation variance by half (about a 22.5 percent increase in standard errors) decreases the value of EBDM by 6.56 percent, from 1.4057 to 1.3133.

We next compute counterfactual EBDM values for different levels of γ^2, the variance of the distribution of τ. For each observation $i = 1, \ldots, N$ in our sample, we draw a value from a distribution with the same mean as the empirical Bayes estimate of the distribution of τ but with variance adjusted by a factor λ. Let $\tau_1^*, \ldots, \tau_n^*$ denote these draws. Next, for each i, we compute $\hat{\tau}_i^* = \tau_i^* + \hat{\sigma}_i U_i$, where $U_1, \ldots, U_n$ are independent standard Gaussian draws. Using the sample $(\hat{\tau}_1^*, \hat{\sigma}_1), \ldots, (\hat{\tau}_n^*, \hat{\sigma}_n)$, we estimate the value of EBDM. Repeating this procedure multiple times yields a distribution of EBDM estimates for a given λ. The mean of this distribution is our estimate of EBDM under variance modification factor λ for γ^2.

Figure 2: Counterfactual values of EBDM
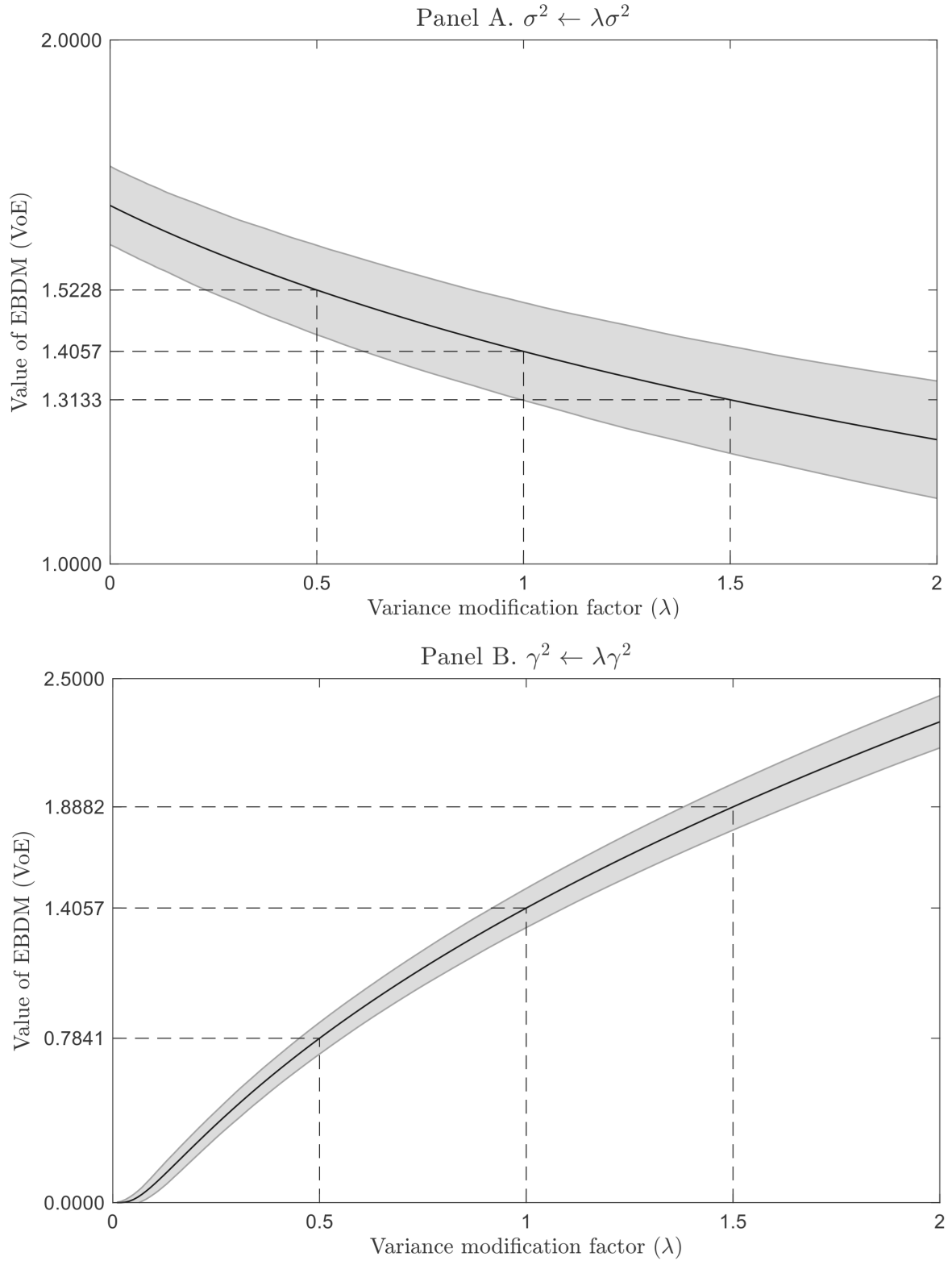
Panel B of Figure 2 shows how changes in the heterogeneity of true policy effects influence the value of EBDM. When the variance modification factor λ is greater than one, meaning that the variance of the true effects increases, the value of EBDM rises. For instance, when γ^2 is increased by 50 percent ($\lambda = 1.5$), the estimated value of EBDM increases from 1.4057 to 1.8882, reflecting a 34.32 percent gain. Conversely, when γ^2 is reduced by half ($\lambda = 0.5$), the value of EBDM declines to 0.7841, representing a 44.22 percent decrease from the baseline.

7. Significance testing reduces the value of information

Organizations commonly implement interventions only when they *(i)* deliver positive estimated treatment effects and *(ii)* attain statistical significance at a prespecified level. This section shows that significance-based decision rules fail to exploit the full informational content of the data and are therefore suboptimal from an EBDM perspective.

Intuitively, statistical significance decision rules prioritize Type I error control rather than maximizing expected payoff. Therefore, they may discard interventions with large but imprecisely estimated effects that could generate substantial rewards, or systematically favor interventions with small but precisely estimated effects, thereby biasing decisions toward low-variance interventions rather than those with high expected payoffs. Moreover, conditioning implementation decisions on statistical significance induces a winner's curse effect: selected interventions are disproportionately likely to be those whose effects were overestimated in the sample due to sampling variation (see Andrews et al., 2023). Empirical Bayes methods address these challenges by shrinking extreme estimates toward the prior distribution of treatment effects, thereby removing substantial noise from signals. This enables decision-makers to better distinguish signal from noise and rank interventions according to their posterior expected value, thus favoring interventions with genuinely positive expected payoffs.

Define the value of EBDM under a statistical significance rule with one-sided significance level α as

$$V_{\text{sig}} = E[(\tau - c_L)I_{(z_{1-\alpha}, \infty)}((\hat{\tau} - c_L)/\sigma)]$$

21

$$= E\big[E[(\tau - c_L)\,|\,\hat{\tau}, \sigma] I_{(z_{1-\alpha}, \infty)}((\hat{\tau} - c_L)/\sigma)\big]$$

where $z_{1-\alpha}$ is the $(1 - \alpha)$-th quantile of a standard normal distribution.

Consider first the case of a fixed value of σ. For the parametric model $\tau \sim N(\mu, \gamma^2)$, the marginal distribution of $\hat{\tau}$ is $\hat{\tau}\,|\,\sigma \sim N(\mu, \gamma^2 + \sigma^2)$. As a result, conditional on σ^2, the value of EBDM under a statistical significance rule at significance level α is

$$
\begin{aligned}
V_{\mathrm{sig}}(\sigma^2) &= E\big[(\tau - c_L) I_{(z_{1-\alpha}, \infty)}((\hat{\tau} - c_L)/\sigma)\,|\,\sigma\big] \\
&= E\left[\left(\frac{\mu/\gamma^2 + \hat{\tau}/\sigma^2}{1/\gamma^2 + 1/\sigma^2} - c_L \right) I_{(z_{1-\alpha}, \infty)}((\hat{\tau} - c_L)/\sigma)\ \middle|\ \sigma \right].
\end{aligned}
$$

Calculations in the appendix show

$$V_{\mathrm{sig}}(\sigma^2) = (\mu - c_L)\Phi\left(\frac{(\mu - c_L) - z_{1-\alpha}\sigma}{\sqrt{\gamma^2 + \sigma^2}} \right) + \frac{\gamma^2}{\sqrt{\gamma^2 + \sigma^2}}\phi\left(\frac{(\mu - c_L) - z_{1-\alpha}\sigma}{\sqrt{\gamma^2 + \sigma^2}} \right). \tag{9}$$

Compare this expression to $V(\sigma^2)$ in (4). It holds that

$$V_{\mathrm{sig}}(\sigma^2) \leqslant V(\sigma^2), \tag{10}$$

provided $\gamma^2 > 0$, with strict inequality except for the case in which the arguments of the functions $\Phi(\cdot)$ and $\phi(\cdot)$ coincide in the expression of $V(\sigma)$ and $V_{\mathrm{sig}}(\sigma)$. The appendix contains a detailed comparison of the payoffs $V(\sigma)$ and $V_{\mathrm{sig}}(\sigma)$.

A plug-in procedure yields the estimator of V_{sig}

$$\frac{1}{n} \sum_{i=1}^{n} \left[(\hat{\mu} - c_L)\Phi\left(\frac{\hat{\mu} - c_L - z_{1-\alpha}\sigma_i}{\sqrt{\hat{\gamma}^2 + \sigma_i^2}} \right) + \frac{\hat{\gamma}^2}{\sqrt{\hat{\gamma}^2 + \sigma_i^2}}\phi\left(\frac{\hat{\mu} - c_L - z_{1-\alpha}\sigma_i}{\sqrt{\hat{\gamma}^2 + \sigma_i^2}} \right) \right].$$

As an alternative, for any estimator of the posterior mean $\hat{E}[\tau\,|\,\hat{\tau}_i, \sigma_i]$ (parametric or nonparametric) an estimator of the value of EBDM under a statistical significance decision rule is given by

$$\frac{1}{n} \sum_{i=1}^{n} \left(\hat{E}[\tau\,|\,\hat{\tau}_i, \sigma_i] - c_L \right) I_{(z_{1-\alpha}, \infty)}((\hat{\tau}_i - c_L)/\sigma_i).$$

Table 3 compares the value delivered by significance-based EBDM to the VoE for the Upworthy data. Across specifications, the empirical Bayes procedures in Section 3 deliver

22

Table 3: The value of optimal vs. significance decision rules in the Upworthy data

	parametric	nonparametric		
		precision independence	binning	CLOSE-NPMLE
	(1)	(2)	(3)	(4)
VoE	1.4057	0.7621	0.9285	0.9560
Significance rule ($\alpha = 0.05$)	1.0287	0.5910	0.6510	0.6609

Note: The table reports estimates of the VoE and the value under a 5 percent significance decision rule for the Upworthy data ($n = 4{,}857$). The parametric estimates are for the heteroskedastic case (Section 3.1.2).

substantially higher value than significance-based decision making. Under the parametric heteroskedastic model, EBDM yields 1.4057, compared with 1.0287 under the 5 percent significance rule—a reduction of about 27 percent. The gap is larger under nonparametric methods: binning and CLOSE yield values between 0.93 and 0.95, whereas the corresponding significance-rule values cluster around 0.66–0.67, implying that significance screening discards roughly 30 percent of attainable value. Overall, these results show that significance-oriented decision rules systematically underperform value-based policies, especially in settings with substantial heterogeneity and estimation noise, where empirical Bayes methods can extract value from interventions that significance tests would discard.

8. Conclusions

This article develops an empirical framework to quantify the value of evidence-based decision making and to examine how statistical precision and heterogeneity in policy effects moderate that value. Using both parametric and nonparametric empirical Bayes methods, we estimate the benefit of incorporating data-driven evidence into decision-making processes by balancing the trade-off between the costs of acquiring information and the expected improvements in outcomes. Higher statistical precision (i.e., lower σ^2) and greater heterogeneity in policy effects (i.e., higher γ^2) increase the value of evidence-based decision making.

Our framework provides a principled approach for organizations to evaluate whether in-

vesting in additional data collection or policy exploration is worthwhile given the expected gains in decision quality. The proposed methods are particularly relevant for organizations that frequently conduct experiments to optimize policies and business strategies. The availability of many experimental evaluations—that is, the availability of many instances of $(\hat{\tau}, \hat{\sigma}^2)$—makes it possible to estimate the distribution of policy effects.

Future research could extend our framework by incorporating more flexible empirical Bayes estimators and exploring settings where estimation variance and treatment effect heterogeneity are determined endogenously by prior information. Additionally, applying our approach to firm-level, governmental, and healthcare decision-making contexts could further validate its usefulness in diverse policy environments.

Currently, many organizations rely on power calculations to guide their study designs, without taking into account the cost-benefit trade-offs associated with evidence-based decision making. As organizations continue to expand their reliance on data for policy decisions, our proposed methods offer a practical procedure for optimizing information acquisition strategies.

Appendix

Convexity of $V(\sigma)$: The second derivative of $V(\sigma^2)$ with respect to σ^2 is

$$\frac{\partial^2 V(\sigma^2)}{\partial \sigma^2 \partial \sigma^2} = \frac{3\gamma^4 + (\mu - c_L)^2 (\gamma^2 + \sigma^2)}{4\gamma^2 (\gamma^2 + \sigma^2)^{5/2}} \phi\left(\frac{\mu - c_L}{\gamma^2 / \sqrt{\gamma^2 + \sigma^2}}\right) \geqslant 0.$$

Now, Jensen's inequality implies $V(E[\sigma^2]) \leqslant E[V(\sigma^2)]$.

Derivation of (7): For any integrable function h and any $s > 0$,

$$\int h(t) \, dF_{\tau/\sigma|\sigma}(t \mid s) = E[h(\tau/\sigma) \mid \sigma = s] = E[h(\tau/s)] = \int h(w/s) \, dG(w).$$

By the law of total probability,

$$f_{\hat{\tau}/\sigma|\sigma}(r \mid s) = \int f_{\hat{\tau}/\sigma|\tau/\sigma,\sigma}(r \mid t, s) \, dF_{\tau/\sigma|\sigma}(t \mid s)$$

$$= \int \phi(r - t) \, dF_{\tau/\sigma|\sigma}(t \mid s)$$

$$= \int \phi\left(r - \frac{w}{s}\right) dG(w).$$

Now consider n studies indexed by $i = 1, \ldots, n$, with observed pairs $\{(\widehat{\tau}_i, \sigma_i)\}_{i=1}^n$. Then, the likelihood of the data $\{(\widehat{\tau}_i, \sigma_i)\}_{i=1}^n$ is

$$\prod_{i=1}^n f_{\widehat{\tau}/\sigma|\sigma}\left(\frac{\widehat{\tau}_i}{\sigma_i} \,\Big|\, \sigma_i\right) = \prod_{i=1}^n \int_{\mathbb{R}} \phi\left(\frac{\widehat{\tau}_i - w}{\sigma_i}\right) dG(w).$$

justifying the nonparametric Maximum Likelihood procedure.

Proof of (8): Since $\widehat{\tau} \mid (\tau = w, \sigma = s) \sim N(w, s^2)$, it follows that $f_{\widehat{\tau}|\tau,\sigma}(z \mid w, s) = \phi((z - w)/s)/s$. Now, the law of total probability implies

$$f_{\widehat{\tau}|\sigma}(z \mid s) = \int f_{\widehat{\tau}|\tau,\sigma}(z \mid w, s) dG(w) = \frac{1}{s} \int \phi\left(\frac{z - w}{s}\right) dG(w).$$

Moreover, it follows from Bayes' rule that

$$dF_{\tau|\widehat{\tau},\sigma}(w \mid z, s) = \frac{f_{\widehat{\tau}|\tau,\sigma}(z \mid w, s) dG(w)}{f_{\widehat{\tau}|\sigma}(z \mid s)}$$

$$= \frac{\phi\left(\dfrac{z - w}{s}\right) dG(w)}{\int \phi\left(\dfrac{z - u}{s}\right) dG(u)}$$

whereby,

$$E[\tau|\widehat{\tau} = z, \sigma = s] = \int w \, dF_{\tau|\widehat{\tau},\sigma}(w \mid z, s) = \frac{\int w \phi\left(\dfrac{z - w}{s}\right) dG(w)}{\int \phi\left(\dfrac{z - w}{s}\right) dG(w)}.$$

Proof of (9): For the parametric model $\tau \sim N(\mu, \gamma^2)$, the marginal distribution of $\widehat{\tau}$ is $\widehat{\tau} \mid \sigma \sim N(\mu, \gamma^2 + \sigma^2)$. As a result, the value of EBDM under a statistical significance rule at significance level α is

$$V_{\text{sig}}(\sigma^2) = E\left[\left(\frac{\mu/\gamma^2 + \widehat{\tau}/\sigma^2}{1/\gamma^2 + 1/\sigma^2} - c_L\right) I_{(z_{1-\alpha}, \infty)}((\widehat{\tau} - c_L)/\sigma) \,\Big|\, \sigma\right].$$

Let $\hat{t} = (\hat{\tau} - c_L)/\sigma$, so $\hat{t} \mid \sigma \sim N((\mu - c_L)/\sigma, (\gamma^2 + \sigma^2)/\sigma^2)$. Then,

$$V_{\text{sig}}(\sigma) = (\mu - c_L) \Pr(\hat{t} > z_{1-\alpha} \mid \sigma) + \left(\frac{\gamma^2}{\gamma^2 + \sigma^2} \right) E[(\hat{\tau} - \mu) I_{(z_{1-\alpha}, \infty)}(\hat{t}) \mid \sigma].$$

Note that

$$\Pr(\hat{t} > z_{1-\alpha} \mid \sigma) = \Pr\left(\frac{\hat{t} - (\mu - c_L)/\sigma}{\sqrt{(\gamma^2 + \sigma^2)/\sigma^2}} > \frac{z_{1-\alpha} - (\mu - c_L)/\sigma}{\sqrt{(\gamma^2 + \sigma^2)/\sigma^2}} \ \middle| \ \sigma \right)$$

$$= 1 - \Phi\left(\frac{z_{1-\alpha} - (\mu - c_L)/\sigma}{\sqrt{(\gamma^2 + \sigma^2)/\sigma^2}} \right).$$

Moreover, it follows from integration by parts that

$$E[(\hat{\tau} - \mu) I_{(z_{1-\alpha}, \infty)}(\hat{t}) \mid \sigma] = \int_{c_L + z_{1-\alpha}\sigma}^{\infty} (t - \mu) \frac{1}{\sqrt{\gamma^2 + \sigma^2}} \phi\left(\frac{t - \mu}{\sqrt{\gamma^2 + \sigma^2}} \right) dt$$

$$= \sqrt{\gamma^2 + \sigma^2} \phi\left(\frac{c_L + z_{1-\alpha}\sigma - \mu}{\sqrt{\gamma^2 + \sigma^2}} \right).$$

Therefore,

$$V_{\text{sig}}(\sigma) = (\mu - c_L) \left(1 - \Phi\left(\frac{z_{1-\alpha} - (\mu - c_L)/\sigma}{\sqrt{(\gamma^2 + \sigma^2)/\sigma^2}} \right) \right) + \frac{\gamma^2}{\sqrt{\gamma^2 + \sigma^2}} \phi\left(\frac{z_{1-\alpha} - (\mu - c_L)/\sigma}{\sqrt{(\gamma^2 + \sigma^2)/\sigma^2}} \right)$$

$$= (\mu - c_L) \Phi\left(\frac{(\mu - c_L) - z_{1-\alpha}\sigma}{\sqrt{\gamma^2 + \sigma^2}} \right) + \frac{\gamma^2}{\sqrt{\gamma^2 + \sigma^2}} \phi\left(\frac{(\mu - c_L) - z_{1-\alpha}\sigma}{\sqrt{\gamma^2 + \sigma^2}} \right).$$

Proof of (10): Consider the function $f(x) = a\Phi(x) + b\phi(x)$, with $b > 0$. Using the fact that $\partial\phi(x)/\partial x = -x\phi(x)$, it follows that

$$\frac{\partial f(x)}{\partial x} = (a - bx)\phi(x).$$

That is, $\partial f(x)/\partial x$ is positive for $x < a/b$, equal to zero for $x = a/b$, and negative for $x > a/b$. As a result, $x^* = a/b$ gives the unique global maximum. Making $a = \mu - c_L$ and $b = \gamma^2/\sqrt{\gamma^2 + \sigma^2}$ proves (10).

Comparison of $V(\sigma)$ and $V_{sig}(\sigma)$

First, notice that $V_{\text{sig}}(\sigma^2)$ and $V(\sigma^2)$ are continuous in σ^2, and $V_{\text{sig}}(0) = V(0)$. That is, the two rules extract similar values in low-noise environments.

As the scale of the noise, σ^2, increases, both $V_{\text{sig}}(\sigma^2)$ and $V(\sigma^2)$ decrease. Indeed, some algebra shows,

$$\frac{\partial V_{\text{sig}}(\sigma^2)}{\partial \sigma^2} = -\left(\frac{\gamma^2(\gamma^2 + \sigma^2) + ((\mu - c_L)\sigma + z_{1-\alpha}\gamma^2)^2}{2(\gamma^2 + \sigma^2)^{5/2}}\right)\phi\left(\frac{(\mu - c_L) - z_{1-\alpha}\sigma}{\sqrt{\gamma^2 + \sigma^2}}\right) < 0.$$

Moreover

$$\lim_{\sigma^2 \to \infty} V_{\text{sig}}(\sigma^2) = (\mu - c_L)\alpha,$$

and

$$\lim_{\sigma^2 \to \infty} V(\sigma^2) = \max\{\mu - c_L, 0\}.$$

That is, as σ^2 increases, significance testing loses power and rejects with probability α. At the same time, as σ^2 increases, $\hat{\tau}$ loses informativeness and $E[\tau \mid \hat{\tau}, \sigma^2]$ converges to μ. This explains the limit of $V_{\text{sig}}(\sigma^2)$. For sufficiently large σ^2, the value of EBDM under a significance decision rule is negative if $\mu - c_L < 0$, and recovers only a small fraction α of $\mu - c_L$ if $\mu - c_L > 0$. In contrast, $V(\sigma^2)$ is always non-negative and always recovers at least the full value of $\mu - c_L$ when $\mu - c_L > 0$.

References

Abadie, A. and Kasy, M. (2019). Choosing among regularized estimators in empirical economics: The risk of machine learning. *The Review of Economics and Statistics*, 101(5):743–762.

Andrews, I., Kitagawa, T., and McCloskey, A. (2023). Inference on winners. *The Quarterly Journal of Economics*, 139(1):305–358.

Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, 132(2):871–919.

Armstrong, T. B., Kolesár, M., and Plagborg-Møller, M. (2022). Robust empirical Bayes confidence intervals. *Econometrica*, 90(6):2567–2602.

Blackwell, D. (1951). Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102.

Chen, J. (2024). Empirical Bayes when estimation precision predicts parameters. arXiv. 2212.14444.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Fessler, P. and Kasy, M. (2019). How to use economic theory to improve estimators: Shrinking toward theoretical restrictions. *The Review of Economics and Statistics*, 101(4):681–698.

Gu, J. and Koenker, R. (2017). Unobserved heterogeneity in income dynamics: An empirical Bayes perspective. *Journal of Business & Economic Statistics*, 35(1):1–16.

Gupta, S., Kohavi, R., Tang, D., Xu, Y., Andersen, R., Bakshy, E., Cardin, N., Chandran, S., Chen, N., Coey, D., et al. (2019). Top challenges from the first practical online controlled experiments summit. *ACM SIGKDD Explorations Newsletter*, 21(1):20–35.

Hedges, L. V. and Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press, Orlando.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A., editors (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Chichester, UK, 2nd edition.

Howard, R. A. (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26.

Jiang, W. (2020). On general maximum likelihood empirical Bayes estimation of heteroscedastic iid normal means. *Electronic Journal of Statistics*, 14:2272–2297.

Kline, P., Rose, E. K., and Walters, C. R. (2024). A discrimination report card. *American Economic Review*, 114(8):2472–2525.

Koenker, R. and Gu, J. (2017). REBayes: an R package for empirical Bayes mixture methods. *Journal of Statistical Software*, 82:1–26.

Koenker, R. and Gu, J. (2024). Empirical Bayes for the reluctant frequentist. arXiv. 2404.03422.

Koenker, R. and Mizera, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *Journal of the American Statistical Association*, 109(506):674–685.

Kwon, S. (2023). Optimal shrinkage estimation of fixed effects in linear panel data models. arXiv. 2308.12485.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

Semenova, V. (2023). Generalized Lee bounds. *Journal of Econometrics*. Forthcoming.

Soloff, J. A., Guntuboyina, A., and Sen, B. (2025). Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(1):1–32.

Walters, C. (2024). Empirical Bayes methods in labor economics. Handbook of Labor Economics, vol. 5, pages 183–260. Elsevier.