Estimating the Value of Evidence-Based Decision Making

Alberto Abadie
MIT

Anish Agarwal
Amazon.com

Guido Imbens
Stanford

Siwei Jia
Amazon.com

James McQueen
Amazon.com

Serguei Stepaniants
Amazon.com

September 9, 2023

Abstract

Business/policy decisions are often based on evidence from randomized experiments and observational studies. In this article we propose an empirical framework to estimate the value of evidence-based decision making (EBDM) and the return on the investment in statistical precision.

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 so many experiments suggests that data evidence provides a value for guiding business/policy decisions. However, we are unaware of the existence of empirical tools that organizations can use to assess the value of their EBDM practices. In the absence of such tools, it is difficult to assess whether too much experimentation is being done, or too little, whether experiments are too large or too small, and whether the right experiments are done. Part of the challenge in evaluating the value of EBDM is that it requires a description of the role of evidence in the business/policy decision

We are grateful to Brad Efron for making a subset of the Cochrane data available to us.

process. In other words, it requires assumptions on what organizations will do with and without various amounts of evidence that they can choose to generate at some cost.

In this article, we propose an empirical Bayes estimator of the value of EBDM. We study the problem of a decision maker deciding 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 effect of the intervention. At the stage where the decision maker decides to implement the intervention or not, she does so to maximize her utility given the available information. We derive expressions for the value of the additional information and show how to estimate this value using meta data on estimates of the effects of business/policy interventions and their standard errors.

Our framework allows decision makers to assess the value of experimental and nonexperimental studies, and how this value changes with the precision of the studies. Currently, many organizations decide on the precision of their studies based on power calculations, which do not take into account the cost and benefits of EBDM.

2. The value of EBDM

We use the notation $X \sim (\theta, \sigma^2)$ to indicate that the random variable X has mean θ and variance σ^2. We use $X \sim N(\theta, \sigma^2)$ to indicate that the distribution of X is Gaussian with mean and variance (θ, σ^2). We use $f_X(\cdot)$ for the density of a random variable, X, and $f_{X|W}(\cdot|w)$ for the conditional density of X given $W = w$. $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and the cumulative distribution function of the standard Gaussian distribution.

2.1. Setup

Consider the problem of a decision maker who has to choose whether to adopt a particular policy on a population of units. The per-unit payoff of the policy, τ, follows a distribution with probability density function $f_\tau(\cdot)$ and mean μ. We assume that the agent has risk-

neutral preferences. As a result, in the absence of additional information, the agent launches the policy as long as 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 $\max\{\mu - c_L, 0\}$.

Suppose, however, that the agent can choose to obtain additional information on the policy payoff, at some cost. In particular, the agent can obtain a signal, $\widehat{\tau}$, with distribution

$$\widehat{\tau} \,|\, \tau \sim N(\tau, \sigma^2), \tag{1}$$

at cost $c_F + c(\sigma^2)$, with $c_F \geqslant 0$, $c(\cdot) \geqslant 0$ and $c'(\cdot) \leqslant 0$. This aspect of the model aims to capture information obtained from studies that estimate policy effects based on experimental or observational data. The constant c_F reflects the fixed cost of a data-driven policy evaluation, and the function $c(\cdot)$ measures the costs of precision, which depends in part on the sample size of the study. The restriction on the derivative of $c(\cdot)$ conveys the notion that more precise information is weakly more costly.

The Gaussianity assumption on the distribution of $\widehat{\tau} \,|\, \tau$ is motivated by approximate Gaussianity of the large sample distributions of many commonly used estimators of treatment effects. After observing the signal $\widehat{\tau}$, the expected payoff of the policy is

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

In this case, the agent launches the policy if

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

For any scalar, x, let $I_{(0,\infty)}(x)$ be the function that takes value one if x is greater than zero and value zero otherwise. The expected payoff with EBDM is

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

$$
\begin{aligned}
&= E\left[I_{(0,\infty)}(E[\tau|\widehat{\tau}] - c_L)E[\tau - c_L|\widehat{\tau}]\right] \\
&= E\left[I_{(0,\infty)}(E[\tau|\widehat{\tau}] - c_L)(E[\tau|\widehat{\tau}] - c_L)\right] \\
&= E\left[\max\left\{E[\tau|\widehat{\tau}] - c_L, 0\right\}\right].
\end{aligned}
$$

Let

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

which is the expected payoff under no additional information besides the information in the distribution of τ. 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 EBDM 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:

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

A second version of VoE, which we term VoID is obtained when, in the absence of additional information about the effect of the intervention, the intervention is always deployed:

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

$\mathsf{VoID}(\sigma^2)$ 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

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

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

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

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 parameters close to zero, as is common in some online experimentation settings, we can approximate

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

2.3. A Gaussian distribution for τ

In this section, we obtain simple closed-form formulas for the case when the distribution of τ can be approximated with a Gaussian distribution,

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

While equation (1) is supported by the Central Limit Theorem for large sample studies, equation (2) imposes two important restrictions. First, τ is Gaussian. Second, this distribution is independent of σ^2. The first is a strong parametric restriction. Gaussianity of τ could be a suitable approximation in some settings, but not in others. The second restriction could be violated, for example, if experimenters have some additional information about τ for each treatment, and adapt the power of their experimental or observational studies accordingly through the choice of sample size. 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.

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

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

The expected payoff with EBDM is

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

Let

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

Recall that the marginal distribution of $\widehat{\tau}$ 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,

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

$V(\sigma^2)$ is decreasing in σ^2 and increasing in γ^2 (see appendix). This conforms to our intuition that higher precision of the signal $\widehat{\tau}|\tau$ and higher uncertainty regarding τ increase the payoff from EBDM.

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 $\sigma^2 \to 0$, we learn 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

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

2.4. A Gaussian mixture distribution for τ

In section 4 we use a Gaussian mixture distribution of τ to evaluate the effects of ... 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 $\widehat{\tau} = \tau + \varepsilon$, where ε is independent Gaussian noise with variance σ^2. Conditional on $\tau \sim N(\mu_j, \gamma_j^2)$, we have

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

As a result,

$$\begin{aligned} E[\tau|\widehat{\tau}] &= \sum_{j=1}^{k} E[\tau|\widehat{\tau}, \tau \sim N(\mu_k, \gamma_j^2)] \Pr(\tau \sim N(\mu_j, \gamma_j^2)|\widehat{\tau}) \\ &= \frac{\displaystyle\sum_{j=1}^{k} \left(\frac{\mu_j/\gamma_j^2 + \widehat{\tau}/\sigma^2}{1/\gamma_j^2 + 1/\sigma^2}\right) \frac{1}{\sqrt{\gamma_j^2 + \sigma^2}} \phi\left(\frac{\widehat{\tau} - \mu_j}{\sqrt{\gamma_j^2 + \sigma^2}}\right) p_j}{\displaystyle\sum_{j=1}^{k} \frac{1}{\sqrt{\gamma_j^2 + \sigma^2}} \phi\left(\frac{\widehat{\tau} - \mu_j}{\sqrt{\gamma_j^2 + \sigma^2}}\right) p_j}. \end{aligned} \tag{6}$$

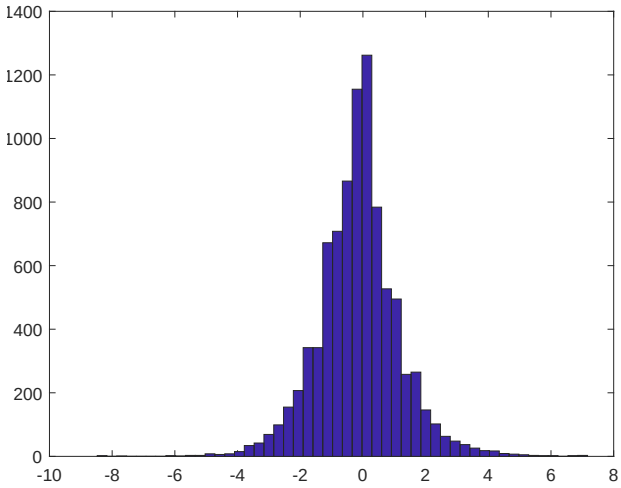


Figure 1: Distribution of estimated treatment effects in the Cochrane data

3. Empirical Bayes estimation

In this section, we consider a setting with multiple realizations from the distribution of $(\tau, \sigma, \widehat{\tau}, \widehat{\sigma})$, where only $\widehat{\tau}$ and $\widehat{\sigma}$ are observed. In this setting, $(\widehat{\tau}, \widehat{\sigma})$ are estimates and the corresponding standard errors for a set of policy evaluations in the data set. We consider the homoskedastic case, with σ^2 constant, and the heteroskedastic case with $\text{var}(\sigma^2) > 0$. In all our calculations we approximate the per-unit launch cost as $c_L \approx 0$. Alternatively we can think of the τ as capturing the net benefits of the treatment after taking out the launch cost.

We illustrate these calculations using a subset of the Cochrane database (Cochrane Collaboration, 2002; Starr et al., 2009) containing information on 8821 randomized experiments. For each experiment we observe the point estimate, $\widehat{\tau}$, as well as the standard error of $\widehat{\tau}$, which can be seen as an estimate of σ. We use the Cochrane dataset as a benchmark, without attempting to interpret the magnitude or direction of the resulting estimates. Figure 1 shows the distribution of $\widehat{\tau}$ across the experiments in the dataset. The average value of $\widehat{\tau}$ is -0.1421 with range $[-8.4763, 7.1663]$. Standard errors have mean 0.7471 and range $[0.0099, 2.1232]$.

3.1. Parametric empirical Bayes

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

$$\text{var}(\widehat{\tau}) = E[\text{var}(\widehat{\tau}|\tau)] + \text{var}(E[\widehat{\tau}|\tau]).$$

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

$$\begin{aligned}\gamma^2 &= \text{var}(\tau) \\ &= \text{var}(E[\widehat{\tau}|\tau]) \\ &= \text{var}(\widehat{\tau}) - E[\text{var}(\widehat{\tau}|\tau)].\end{aligned}$$

As a result, we define $\widehat{\gamma}^2$ as the difference between the variance $\widehat{\tau}$ across experiments in the data minus the mean of the squares of the standard errors. This estimator is not guaranteed to be positive. In the Cochrane dataset, $\widehat{\gamma}^2 = 1.6677 - 0.7641 = 0.9036$.

3.1.1. Homoskedastic case

For the homoskedastic case, we estimate σ^2 as the average of the squares of the standard deviations of $\widehat{\tau}$ across studies. In the Cochrane data, this estimate is 0.7641. Plugging in this value in (4) along with estimates of μ and γ^2, we obtain $V(0.7641) = 0.2138$.

3.1.2. Heteroskedastic case

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 $\widehat{\tau}$ to approximate σ^2, and estimate the value of each study separately, and *(ii)* take the average over all studies in the sample. Doing this for the Cochrane data, we obtain $V = 0.2351$.

3.2. Nonparametric empirical Bayes

We now relax the parametric restriction $\tau \sim N(\mu, \gamma^2)$ of section 3.1 and consider a nonparametric distribution.

3.2.1. Homoskedastic case

Suppose we have $\tau/\sigma|\sigma \sim G_\sigma$, where G_σ is some unspecified distribution. Then,

$$\widehat{\tau}/\sigma \big| \tau/\sigma, \sigma \sim N(\tau/\sigma, 1).$$

Therefore, the marginal density of $\widehat{\tau}/\sigma$ is

$$f_{\hat{\tau}/\sigma}(z) = \int \left(\int \phi(z-t) dG_s(t) \right) dP(s),$$

where P is the distribution of σ. If $\tau/\sigma|\sigma \sim G$, independent of σ (a very strong assumption), then

$$f_{\hat{\tau}/\sigma}(z) = \int \phi(z-t) dG(t),$$

and the distribution G can be estimated via nonparametric empirical Bayes methods (NPEB). Suppose we have n experiments, indexed $i = 1, \ldots, n$. Following Jiang and Zhang (2009), we can estimate G as

$$\max_{G \in \mathcal{G}} \sum_{i=1}^{n} \log \int \phi(\widehat{\tau}/\sigma - t) dG(t). \tag{7}$$

Jiang and Zhang (2009) uses fixed-point methods to solve the nonparametric maximum likelihood problem in (7). The solution is computed over a grid $u_1, \ldots, u_m$ representing points of mass of G, with respective probabilities, $\widehat{f}_1, \ldots, \widehat{f}_m$. Notice that,

$$E[\tau/\sigma|\hat{\tau}/\sigma = z, \sigma = s] = \frac{\int t\, \phi(z-t) dG(t)}{\int \phi(z-t) dG(t)}.$$

Table 1: Expected payoff of EBDM

distribution of τ:		true value	empirical Bayes parametric	empirical Bayes nonparametric
Gaussian	expected payoff	0.3970	0.3970	0.3970
	95% interval		[0.3429, 0.4510]	[0.3423, 0.4517]
mixture	expected payoff	0.3225	0.3942	0.3227
	95% interval		[0.3168, 0.4717]	[0.2685, 0.3768]

The sample analog of $E[\tau|\hat{\tau}/\sigma = z, \sigma = s]$ is

$$\sigma_i \frac{\sum_{j=1}^{m} u_j \phi(\widehat{\tau}_i/\sigma_i - u_j)\widehat{f}_j}{\sum_{j=1}^{m} \phi(\widehat{\tau}_i/\sigma_i - u_j)\widehat{f}_j}.$$

Koenker and Mizera (2014) propose an algorithm to solve (7) based on convex optimization procedures. For the Cocharne data set, using the standard errors of the estimates to approximate $\sigma_1, \ldots, \sigma_n$ and the algorithm in Koenker and Mizera (2014) to estimate G, we obtain the estimate $\widehat{V} = 0.2226$.

3.2.2. Heteroskedastic case

We relax the strong assumption that τ/σ is independent of σ by dividing the support of σ into 5 intervals and doing the NPEB calculations of the previous section interval by interval. Applying this method to the Cochrane data set, we obtain the estimate $\widehat{V} = 0.2897$.

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 $\widehat{\tau}$ as $\widehat{\tau} = \tau + \sigma u$, 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 $E[\tau|\widehat{\tau}]$ over the $n \times 1000 = 500,000$ realizations of $\widehat{\tau}$ in the simulations, and report the average of $\max\{E[\tau|\widehat{\tau}], 0\}$. 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 only 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 as known.

Table 1 reports the true values of the expected payoff of EBDM along with means and 95% intervals for the distribution of the estimates across simulations. When the distribution of τ is Gaussian, the distribution of the parametric and nonparametric empirical Bayes estimates across simulations are both centered at 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% interval for the nonparametric estimator is only 1.1% wider than the interval for the parametric estimator.

For the case when the distribution of τ is a mixture, the results for 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% interval for the nonparametric estimator is 30.1% narrower than the interval for the parametric estimator.

Table 2: Results for the Cochrane data

Distribution of $\widehat{\tau}$:

$$\widehat{\mu} = -0.1421$$

Value of EBDM:

		parametric	nonparametric	
			homoskedastic	heteroskedastic
VoE	estimate	0.2351	0.2226	0.2055
	95% interval	[0.2164, 0.2537]	[0.2060, 0.2392]	[0.1890, 0.2220]
VoID	estimate	0.3772	0.3647	0.3477
	95% interval	[0.3572, 0.3972]	[0.3443, 0.3852]	[0.3267, 0.3686]

5. Application to the Cochrane dataset

Table 2 collects parametric and nonparametric empirical Bayes estimates of the value of EBDM in the Cochrane data. For the parametric case, Table 2 reports estimates computed under heteroskedasticity. For the nonparametric case, Table 2 reports estimates computed under homoskedasticity ($\text{var}(\sigma^2) = 0$) and under heteroskedasticity ($\text{var}(\sigma^2) > 0$). Below each of the estimates of the value of EBDM, Table 2 reports 95% intervals computed over 1000 bootstrap draws from the distribution of $(\widehat{\tau}, \widehat{\sigma}^2)$ in the data. In our calculations, we impose $c_L + c_F + c(\sigma^2) = 0$. The average value of $\widehat{\tau}$ across experiments in the Cochrane data is -0.1421 and the variance of $\widehat{\tau}$ across experiments is 1.6677. Because $\widehat{\tau}$ has a negative mean and a variance that is large relative to the mean, this is a setting where we expect to have substantial gains from EBDM. Indeed, both the parametric and the nonparametric empirical Bayes estimates show positive values of about twice the magnitude of $\widehat{\mu}$. For this application, parametric empirical Bayes and nonparametric empirical Bayes under homoskedasticity produce similar estimates. Allowing for heteroskedasticity decreases the magnitudes of nonparametric empirical Bayes estimates by 7.7% for the case of VoE and 4.7 for the case of VoID.

6. Estimation of counterfactual EBDM values

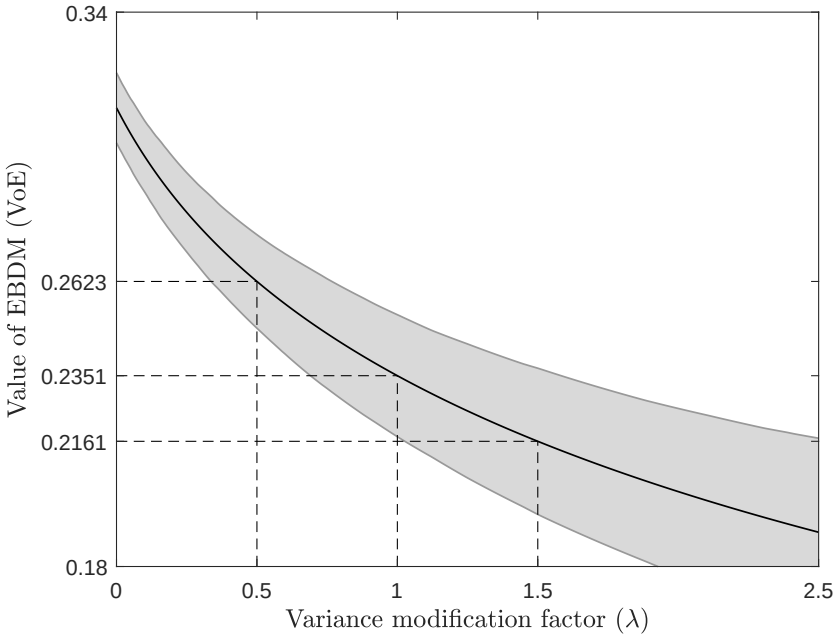
This section provides estimates of the value of EBDM under alternative levels of statistical precision. That is, we estimate how the value of EBDM would change as a result of a change in σ^2. In the resulting counterfactuals, the variance of the estimators is equal to the variance estimates for $\widehat{\tau}_1, \ldots, \widehat{\tau}_n$ in the original sample multiplied by λ. That is, $\lambda = 0.5$ represents a counterfactual scenario where the variances of the estimators are 50% smaller than the variance estimates in the original sample, while for $\lambda = 1.5$ the variances of the estimators are 50% 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 our 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 $\widehat{\tau}_i^* = \tau_i^* + \sigma_i^* U_i$, where $U_1, \ldots, U_n$ are independent draws from the standard Gaussian distribution, and $\sigma_i^* = \sqrt{\lambda}\widehat{\sigma}_i$. We use the new sample $(\widehat{\tau}_1^*, \widehat{\sigma}_1^*), \ldots, (\widehat{\tau}_n^*, \widehat{\sigma}_n^*)$ 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 a empirical counterpart of equation (5) that applies the variance modification factor λ to σ^2.

Figure 2 reports the results obtained from applying the procedure describe above to the Cochrane data. The solid line represents the value of EBDM as a function of the variance modification factor, λ. The shaded area represents 95% intervals from the distribution of EBDM estimates. An investment that reduces estimation variance by half (about a 29.3% decrease in standard errors) leads to an increase in the value of EBDM by 11.6%, from 0.2351 to 0.2623. Conversely, an increase in estimation variance by half (about a 22.5% increase in

Figure 2: Counterfactual values of EBDM



standard errors) decreases the value of EBDM by 8.1%, from 0.2351 to 0.2161.

Appendix

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

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

This conforms to our intuition that more experimental precision should increase the value of experimentation.

The derivative of $V(\sigma^2)$ with respect to γ^2 is

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

This conforms to our intuition that more concentrated priors should reduce the value of experimentation.

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 $V(E[\sigma^2]) \leqslant E[V(\sigma^2)]$.

References

Cochrane Collaboration (2002). The Cochrane library. Database available on disk and CDROM. Oxford, UK, Update Software.

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.

Jiang, W. and Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics*, 37(4):1647–1684.

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.

Starr, M., Chalmers, I., Clarke, M., and Oxman, A. D. (2009). The origins, evolution, and future of The Cochrane Database of Systematic Reviews. *International Journal of Technology Assessment in Health Care*, 25(S1):182–195.