Empirical Bayes Methods: Theory and Application

Jiaying Gu, University of Toronto
Chris Walters, UC Berkeley and NBER

NBER Methods Lectures

July 2022
Empirical Bayes Applications

- Economists are increasingly drilling down to study heterogeneity in fine-grained, unit-specific parameters

  - Returns to a year of education $\rightarrow$ Returns to college selectivity $\rightarrow$ Returns to specific colleges (Card, 1999; Dale and Krueger, 2002, 2014; Mountjoy and Hickman, 2021)

  - Industry wage premia $\rightarrow$ Firm-specific wage premia (Krueger and Summers, 1988; Abowd et al., 1999; Card et al., 2018)

  - Effects of neighborhood characteristics $\rightarrow$ Effects of specific neighborhoods (Kling et al., 2007; Chetty and Hendren, 2018; Chetty et al., 2018)

- In settings with many unit-specific parameters, **empirical Bayes** (EB) methods are useful for

  - Learning about the distribution of parameters across units
  - Improving estimates for individual units (“borrowing strength”)
Today’s Agenda

- Goals for the rest of today:
  - Recap basic EB theory
  - Illustrate through two applications

  - Classic parametric EB

- Application 2: Labor market discrimination among large US employers (Kline, Rose, and Walters, forthcoming)
  - Non-parametric/robust EB
Consider a population of students indexed by $i$, each attending one of $J$ schools in a district.

Let $Y_i(j)$ denote student i’s potential academic achievement if s/he attends school $j \in \{1, \ldots, J\}$.

Simple additive model for potential outcomes:

$$Y_i(j) = \beta_j + \varepsilon_i$$

$\beta_j$ is the value-added of school $j$.

$\varepsilon_i$ represents unobserved student heterogeneity (family background, ability, etc.). Normalize $E[\varepsilon_i] = 0$.

Constant effects model: $\beta_j - \beta_k$ is the effect of moving any student from school $k$ to school $j$. 
Questions About Schools

▶ Several possible questions of interest in this setting
▶ Might be interested in the value-added of a particular school, e.g. $\beta_1$
▶ Might be interested in features of the distribution of $\beta_j$’s across schools
  ▶ How much does school quality vary?
▶ Might be interested in making a decision that depends on the $\beta_j$’s
  ▶ Which school should my child attend? Which school(s) should be closed or expanded?
▶ EB methods are useful for answering each of these questions
Letting $D_{ij}$ indicate attendance at $j$, observed outcome is:

$$Y_i = \sum_j \beta_j D_{ij} + \varepsilon_i$$

Project $\varepsilon_i$ on a vector of covariates $X_i$ (e.g. demographics and lagged achievement):

$$Y_i = \sum_j \beta_j D_{ij} + X_i' \gamma + u_i$$

Here $E[X_i u_i] = 0$ by definition

Suppose we have selection-on-observables: additive control for $X_i$ captures all selection bias, so $E[D_{ij} u_i] = 0 \forall j$

Then ordinary least squares (OLS) regression recovers the parameters of this value-added model (VAM)
VAM Estimates

- VAM estimation yields an estimate for each school along with standard errors: $\{\hat{\beta}_j, s_j\}_{j=1}^J$

- Assume:

  $$\hat{\beta}_j|\beta_j, s_j \sim N(\beta_j, s_j^2)$$

- Think of this as an asymptotic approximation: schools are large enough for estimates to be approximately normal and centered at the truth, with variance $\approx s_j^2$
Introducing $G$

- Second level of the hierarchy describes the cross-school distribution of value-added:

$$\beta_j \sim G(\beta), \ j = 1, \ldots, J$$

- The **mixing distribution** $G$ is a key object in the EB framework
- $G$ is an objective feature of the world, not a subjective prior
- $G$ answers questions about variation in value-added
  - How much does school quality vary? $\sigma_{\beta}^2 = \int (\beta - \mu_{\beta})^2 dG(\beta)$
  - What’s the difference between 75th and 25th percentiles of value-added? $G^{-1}(0.75) - G^{-1}(0.25)$

- EB **deconvolution**: Use noisy estimates $\hat{\beta}_j$ along with standard errors $s_j$ to compute an estimate $\hat{G}$ of $G$
The Philosophy of $G$

- What does it mean to say that value-added parameters are random draws from a distribution $G$?

- “Fixed effects” perspective: There are $J$ schools in the district, with fixed but unknown parameters $\{\beta_j\}_{j=1}^J$

- One (unsatisfying) answer: observed schools are sampled from some larger superpopulation

- “Random effects” perspective can be motivated by analyst’s objectives

- Even with finite population of schools, we can ask how the $\beta_j$’s are distributed in this population

- If our loss function cares about average performance across schools, it’s valuable to incorporate distributional information into estimates for individuals

- Continuous/$iid$ models for $G$ as parsimonious approximations

- Random vs. fixed effects is not about correlation of $\beta_j$’s with VAM $X$’s (c.f. “random effects” vs. “correlated random effects”)

Chris Walters (UC Berkeley) Empirical Bayes Methods
Normal/Normal Model

- Suppose $G$ is normal and independent of $s_j$
- Then we have the hierarchical model

$$
\hat{\beta}_j | \beta_j, s_j \sim N(\beta_j, s_j^2)
$$

$$
\beta_j | s_j \sim N(\mu_\beta, \sigma_\beta^2)
$$

- **Hyperparameters** $\mu_\beta$ and $\sigma_\beta^2$ summarize the value-added distribution
- With this model for $G$, deconvolution just requires estimating these two hyperparameters
Estimating Hyperparameters

- Common estimators for value-added hyperparameters:

\[
\hat{\mu}_\beta = \frac{1}{J} \sum_{j=1}^{J} \hat{\beta}_j
\]

\[
\hat{\sigma}^2_{\beta} = \frac{1}{J} \sum_{j=1}^{J} \left[ (\hat{\beta}_j - \hat{\mu}_\beta)^2 - s_j^2 \right]
\]

- Subtracting \(s_j^2\) is a bias-correction accounting for excess variance in \(\hat{\beta}_j\)'s due to sampling error

- \(\hat{\sigma}^2_{\beta} > 0 \implies \text{overdispersion}\) beyond what we’d expect from noise

- Other approaches: MLE; Kline, Saggio, and Sølvsten (2020) unbiased variance estimator
In normal/normal model, posterior mean for $\beta_j$ given ($\hat{\beta}_j, s_j$) is:

$$
\beta_j^* \equiv E[\beta_j|\hat{\beta}_j, s_j] = \left( \frac{\sigma^2_{\beta}}{\sigma^2_{\beta} + s_j^2} \right) \hat{\beta}_j + \left( \frac{s_j^2}{\sigma^2_{\beta} + s_j^2} \right) \mu_{\beta}
$$

- Posterior mean **shrinks** noisy estimate $\hat{\beta}_j$ toward prior mean based on signal-to-noise ratio.

- Linear shrinkage formula coincides with regression of $\beta_j$ on $\hat{\beta}_j \Rightarrow$ minimum mean squared error (MSE) linear predictor even if $G$ isn’t normal.
EB Posterior Means

- Putting the “E” in “EB” – Empirical Bayes posterior mean $\hat{\beta}_j^*$ plugs in estimated hyperparameters $\hat{\sigma}_\beta^2$ and $\hat{\mu}_\beta$:

$$\hat{\beta}_j^* = \left( \frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\beta^2 + s_j^2} \right) \hat{\beta}_j + \left( \frac{s_j^2}{\hat{\sigma}_\beta^2 + s_j^2} \right) \hat{\mu}_\beta$$

- EB posterior shrinks estimate for school $j$ using hyperparameters estimated with the larger pool of schools

- Reflects general EB approach: Use deconvolution estimate $\hat{G}$ as prior when forming posteriors for individual units

  - “Borrowing strength from the ensemble” (Efron and Morris, 1973; Morris, 1983)
  - “Learning from the experience of others” (Efron, 2012)
Summary: A Three-step EB Recipe

1. **Effect estimation:** Estimate parameter for each unit
   \[ \{\hat{\beta}_j, s_j\}_{j=1}^{J} \]

2. **Deconvolution:** Use \{\hat{\beta}_j, s_j\}_{j=1}^{J} to estimate mixing distribution
   \[ \hat{G} \]

3. **Posterior formation:** Treating \( \hat{G} \) as prior, update with \((\hat{\beta}_j, s_j)\) to form posterior
   \[ \{\hat{\beta}^*_j\}_{j=1}^{J} \]
Should we prefer the shrunk posterior mean to the unbiased estimate $\hat{\beta}_j$? It depends on our goals.

Conditional on the value-added of school $j$, MSE for the two estimators is:

$$E \left[ (\hat{\beta}_j - \beta_j)^2 | \beta_j, s_j \right] = s_j^2$$

$$E \left[ (\beta_j^* - \beta_j)^2 | \beta_j, s_j \right] = \left( \frac{\sigma_\beta^2}{\sigma_\beta^2 + s_j^2} \right)^2 s_j^2 + \left( \frac{s_j^2}{\sigma_\beta^2 + s_j^2} \right)^2 (\beta_j - \mu_\beta)^2$$

If we’re only interested in one school (e.g. $\beta_1$), not clear which is better.

Shrinkage reduces variance, but may introduce substantial bias if the school is very different from average.
When to Shrink?

- Now suppose we’re interested in many schools
- In this case the relevant notion of MSE integrates over $G$:

$$E \left[ (\hat{\beta_j} - \beta_j)^2 | s_j \right] = \int E \left[ (\hat{\beta_j} - \beta)^2 | \beta = \beta_j, s_j \right] dG(\beta) = s_j^2$$

$$E \left[ (\beta_j^* - \beta_j)^2 | s_j \right] = \int E \left[ (\beta_j^* - \beta)^2 | \beta_j = \beta, s_j \right] dG(\beta) = \left( \frac{\sigma^2_{\beta}}{\sigma^2_{\beta} + s_j^2} \right) s_j^2$$

- Linear shrinkage estimate is superior if we want an estimator that performs well on average across schools
  - Holds whether or not $G$ is normal (James/Stein 1961 result)
  - See Armstrong et al. (forthcoming) on robust inference
VAM Standard Deviations for Boston Middle Schools (Sixth Grade Math)

This figure compares standard deviations of school effects from alternative OLS value-added models. The notes to Table III describe the controls included in the lagged score and gains models; the uncontrolled model includes only year effects. The variance of OLS value-added is obtained by subtracting the average squared standard error from the sample variance of value-added estimates. Within-sector variances are obtained by first regressing value-added estimates on charter and pilot dummies, then subtracting the average squared standard error from the sample variance of residuals.

Estimates from Angrist et al. (2017)
Histogram of Lagged Score VAM Estimates for Boston (Sixth Grade Math, 2014)

Std. dev. of estimates: 0.221

Math value-added (std. dev.)

BPS estimates
Charter estimates

Chris Walters  (UC Berkeley)  Empirical Bayes Methods
Prior Distribution Pooling Sectors

Chris Walters (UC Berkeley) Empirical Bayes Methods
Posterior Means Pooling Sectors

Std. dev. of estimates: 0.221
Std. dev. of prior: 0.197
Std. dev. of post. means: 0.179

Chris Walters  (UC Berkeley)  Empirical Bayes Methods
Incorporating Covariates

- It is often natural to build observed covariates into EB estimates
  - Learning from the experience of which others?

- Model for $G$ conditional on a vector of characteristics $C_j$, e.g. charter sector indicator:

  $$\beta_j|s_j, C_j \sim N(C_j^\prime \mu, \sigma_r^2)$$

- Estimate $\mu$ from regression of $\hat{\beta}_j$ on $C_j$; deconvolve residuals $\hat{r}_j = \hat{\beta}_j - C_j^\prime \hat{\mu}$ to estimate $\sigma_r^2$

- Resulting EB posterior shrinks $\hat{\beta}_j$ toward estimated linear index:

  $$\hat{\beta}_j^* = \left( \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + s_j^2} \right) \hat{\beta}_j + \left( \frac{s_j^2}{\hat{\sigma}_r^2 + s_j^2} \right) C_j^\prime \hat{\mu}$$
Prior with Charter Sector Location Shift

Chris Walters (UC Berkeley)

Empirical Bayes Methods
Posteriors Shrinking Toward Sector Means

Std. dev. of estimates: 0.221
Charter effect: 0.293
Resid. std. dev. of prior: 0.139
Std. dev. of post. means: 0.183

Math value-added (std. dev.)

BPS posteriors
BPS estimates
BPS prior
Charter posteriors
Charter estimates
Charter prior

Chris Walters (UC Berkeley)
Empirical Bayes Methods
EB for Bias Correction

- EB framework extends naturally to cases where we have multiple estimates of the same parameter, some possibly biased.

- Changing notation, let $\hat{\alpha}_j$ denote OLS estimate for school $j$, and suppose selection-on-observables fails, represented by bias parameter $b_j$:

$$\hat{\alpha}_j | \beta_j, b_j, s_{j\alpha} \sim N(\beta_j + b_j, s_{j\alpha}^2)$$

- Suppose we also have a noisy but (asymptotically) unbiased estimate $\hat{\beta}_j$, e.g. IV estimate from randomized lottery:

$$\hat{\beta}_j | \beta_j, b_j, s_{j\beta} \sim N(\beta_j, s_{j\beta}^2)$$

- Suppose a Hausman test rejects $\text{OLS} = \text{IV}$. Should we throw away OLS?
EB for Bias Correction

\[
\hat{\alpha}_j | \beta_j, b_j, s_{j\alpha} \sim N(\beta_j + b_j, s_{j\alpha}^2)
\]

\[
\hat{\beta}_j | \beta_j, b_j, s_{j\beta} \sim N(\beta_j, s_{j\beta}^2)
\]

- We can use the ensemble \(\{\hat{\alpha}_j, \hat{\beta}_j\}_{j=1}^J\) to estimate \(G(\beta, b)\), the joint distribution of truth and bias

- EB “hybrid” posterior \(\hat{\beta}^*_j = E_G[\hat{\beta}_j | \hat{\beta}_j, \hat{\alpha}_j]\) trades off bias and variance to minimize MSE:

\[
\hat{\beta}^*_j = \hat{\tau}_\beta \hat{\beta}_j + \hat{\tau}_\alpha (\hat{\alpha}_j - (\hat{\mu}_\alpha - \hat{\mu}_\beta)) + (1 - \hat{\tau}_\beta - \hat{\tau}_\alpha) \hat{\mu}_\beta
\]

- Angrist et al. (2017) generalize to underidentified case; see also Chetty and Hendren (2018)
MSE Improvements from Lottery-based Hybrid Estimates

**FIGURE VI**

Root Mean Squared Error for Value-Added Posterior Predictions

This figure plots root mean squared error (RMSE) for posterior predictions of sixth-grade math value-added. Conventional predictions are posterior means constructed from OLS value-added estimates. Hybrid predictions are posterior modes constructed from OLS and lottery estimates. The total height of each bar indicates RMSE. Dark bars display shares of mean squared error due to bias, and light bars display shares due to variance. RMSE is calculated from 500 simulated samples drawn from the data generating processes implied by the estimates in Table VI. The random coefficients model is reestimated in each simulated sample.
EB Decision Rules

- EB posterior means deliver estimates with low MSE
- We often have goals other than minimizing MSE
- Example: Suppose we want to select schools with value-added below a cutoff $c$
- Loss function for decision $\delta_j \in \{0, 1\}$:
  \[
  L(\beta_j, \delta_j) = \delta_j 1 \{\beta_j > c\} + (1 - \delta_j) 1 \{\beta_j \leq c\} \kappa
  \]
- Cost 1 of mistakenly selecting high-performing school; cost $\kappa$ of failing to select low-performing school
- Risk-minimizing decision rule with $J$ schools:
  \[
  \delta^* = \arg\min_{\delta \in D} \sum_j \int \int L(\beta, \delta(\hat{\beta}, s_j)) \frac{1}{s_j} \phi \left( \frac{\hat{\beta} - \beta}{s_j} \right) d\hat{\beta} \, dG(\beta|s_j)
  \]
EB Decision Rules

Solution is to select schools with sufficiently high posterior probability of value-added below $c$:

$$\delta^*(\hat{\beta}_j, s_j) = 1 \left\{ \Pr_G [\beta_j < c | \hat{\beta}_j, s_j] \geq \frac{1}{1 + \kappa} \right\}$$

This means we should select based on posterior $(1/(1 + \kappa))$ quantile rather than posterior mean. In normal/normal model:

$$\delta^*(\hat{\beta}_j, s_j) = 1 \left\{ \left( \frac{\sigma^2_{\beta}}{\sigma^2_{\beta} + s^2_j} \right) \hat{\beta}_j + \left( \frac{s^2_j}{\sigma^2_{\beta} + s^2_j} \right) \mu_{\beta} + \sqrt{\frac{\sigma^2_{\beta} s^2_j}{\sigma^2_{\beta} + s^2_j}} \Phi^{-1} \left( \frac{1}{1 + \kappa} \right) \leq c \right\}$$

EB decision rule plugs in estimated hyperparameters $(\hat{\mu}_{\beta}, \hat{\sigma}^2_{\beta})$

Different objectives call for using different functionals of posterior for decision-making

See Gu and Koenker (2021) for EB analysis of tail selection problems
EB and Machine Learning

- EB methods are closely related to **machine learning** (ML) approaches
- Parametric normal/normal model with $N$ students per school:

\[
Y_{ij} = \beta_j + \varepsilon_{ij}
\]

\[
\varepsilon_{ij} | \beta_j \sim N(0, \sigma_\varepsilon^2)
\]

\[
\beta_j \sim N(0, \sigma_\beta^2)
\]

- Unbiased estimator $\bar{Y}_j = \frac{1}{N} \sum_i Y_{ij}$, with variance $\text{Var}(\bar{Y}_{ij} | \beta_j) = \sigma_\varepsilon^2 / N$

- Posterior distribution for $\beta_j$ is $N(\beta^*_j, V^*)$ with

\[
\beta^*_j = \left( \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_\varepsilon^2 / N} \right) \bar{Y}_j, \quad V^* = \frac{\sigma_\varepsilon^2 \sigma_\beta^2}{N \sigma_\beta^2 + \sigma_\varepsilon^2}
\]
Posterior density for $\beta_j$:

$$f(\beta_j | Y_{1j}, ...., Y_{Nj}) = \frac{\prod_{i=1}^{N} \frac{1}{\sigma_{\epsilon}} \phi \left( \frac{Y_{ij} - \beta_j}{\sigma_{\epsilon}} \right) \frac{1}{\sigma_{\beta}} \phi \left( \frac{\beta_j}{\sigma_{\beta}} \right)}{\int_{-\infty}^{\infty} \prod_{i=1}^{N} \frac{1}{\sigma_{\epsilon}} \phi \left( \frac{Y_{ij} - \beta}{\sigma_{\epsilon}} \right) \frac{1}{\sigma_{\beta}} \phi \left( \frac{\beta}{\sigma_{\beta}} \right) d\beta}$$

Posterior distribution is normal $\Rightarrow$ posterior mean and mode coincide

This implies posterior means maximize posterior density:

$$(\beta_1^*, ..., \beta_J^*) = \arg \max_{(\beta_1, ..., \beta_J)} \sum_j \log f(\beta_j | Y_{1j}, ...., Y_{Nj})$$

$$= \arg \max_{(\beta_1, ..., \beta_J)} \sum_{j=1}^{J} \sum_{i=1}^{N} \log \phi \left( \frac{Y_{ij} - \beta_j}{\sigma_{\epsilon}} \right) + \sum_{j=1}^{J} \log \phi \left( \frac{\beta_j}{\sigma_{\beta}} \right) + \text{cons}$$

Posterior mode is also known as a maximum a posteriori (MAP) estimate
EB and Machine Learning

- Plugging in normal density yields

\[
(\beta^*_1, ..., \beta^*_J) = \arg\ max_{(\beta_1, ..., \beta_J)} - \sum_{j=1}^{J} \sum_{i=1}^{N} \frac{(Y_{ij} - \beta_j)^2}{2\sigma^2} - \sum_{j=1}^{J} \frac{\beta_j^2}{2\sigma^2}
\]

\[
= \arg\ min_{(\beta_1, ..., \beta_J)} \sum_{j=1}^{J} \sum_{i=1}^{N} (Y_{ij} - \beta_j)^2 + \frac{\sigma^2}{\sigma^2_\beta} \sum_{j=1}^{J} \beta_j^2
\]

\[
= \arg\ min_{(\beta_1, ..., \beta_J)} \sum_{j=1}^{J} \sum_{i=1}^{N} (Y_{ij} - \beta_j)^2 + \lambda p(\beta_1, ..., \beta_J)
\]

- This is regularized least squares with an L2 (quadratic) penalty \( p(\cdot) \), also known as ridge regression

- Empirical Bayes \( \implies \) use the data to choose tuning parameters in penalty function
ML penalization/regularization procedures often have an EB interpretation

- Ridge regression estimates (L2 penalization) can be interpreted as posterior means from a model with normal priors
- LASSO estimates (L1 penalization) can be interpreted as MAP estimates from a model with double exponential (Laplace) priors

- When doing model selection or penalization via ML, useful to think about implicit prior distribution and connection to loss function
- See Abadie and Kasy (2019) for analysis of the relative performance of common regularization approaches under various $G$'s
Application 2: Employer-level Labor Market Discrimination

- Kline, Rose and Walters (forthcoming) apply EB methods to study the distribution of discrimination across large US employers.

- Massive resume correspondence study sending applications to multiple establishments at large employers:
  - 108 Fortune 500 firms
  - Up to 125 jobs per firm, each in a different county
  - 8 applications per job (stratified 4 Black/4 white)

- Following Bertrand and Mullainathan (2004), manipulate employer perceptions of race and sex using distinctive names.
Job-level Estimates

- Let $Y_{ijf}(r) \in \{0, 1\}$ indicate potential callback to applicant $i$ at job $j$ within firm $f$ if assigned race $r \in \{b, w\}$

- Average treatment effect at this job is $\Delta_{jf} \equiv E[Y_{ijf}(w) - Y_{ijf}(b)]$

- Observed outcome is $Y_{ijf} = Y_{ijf}(R_{ijf})$, with $R_{ijf} \in \{b, w\}$

- Black/white difference in callback rates (contact gap):
  \[
  \hat{\Delta}_{jf} = \frac{1}{4} \sum_{i=1}^{8} 1\{R_{ijf} = w\} Y_{ijf} - \frac{1}{4} \sum_{i=1}^{8} 1\{R_{ijf} = b\} Y_{ijf}
  \]

- Random assignment of $R_{ijf} \implies \hat{\Delta}_{jf}$ is an unbiased estimate of $\Delta_{jf}$
Firm-level Estimates

Let $\Delta_f = E_f[\Delta_{jf}]$ denote the average of $\Delta_{jf}$ across all jobs within firm $f$.

Observed average contact gap at firm $f$:

$$\hat{\Delta}_f = \frac{1}{J_f} \sum_{j=1}^{J_f} \hat{\Delta}_{jf}$$

Random sampling of jobs $\implies \hat{\Delta}_f$ is an unbiased estimate of $\Delta_f$.

Unbiased (squared) standard error estimator:

$$s^2_f = \frac{1}{J_f(J_f - 1)} \sum_{j=1}^{J_f} (\hat{\Delta}_{jf} - \hat{\Delta}_f)^2$$

$\{\hat{\Delta}_f, s_f\}_{f=1}^F$ provide building blocks for EB analysis of firm heterogeneity.
The Distribution of Discrimination

- Let $G$ denote the distribution of contact gaps across firms:

  $\Delta_f \sim G(\Delta), \ f = 1, \ldots, F$

- $G$ answers questions about concentration of discrimination

  - Is average white/Black difference in callbacks driven by a small share of severe discriminators?

- Start by estimating mean and variance

- Then use flexible deconvolution methods to estimate other features of $G$
Average Contact Gaps by Race and Gender

- White/Black difference: 0.021 (0.002)
- Male/female difference: -0.001 (0.003)

Chris Walters (UC Berkeley)  Empirical Bayes Methods
Variance Estimation

Estimator for variance of $G$:

\[
\hat{\sigma}_\Delta^2 = \left(\frac{F - 1}{F}\right) \left[ \frac{1}{F - 1} \sum_{f=1}^{F} \left(\hat{\Delta}_f - \bar{\Delta}\right)^2 - \frac{1}{F} \sum_{f=1}^{F} s_f^2 \right]
\]

Special case of unbiased leave-out variance component estimator of Kline, Saggio and Sølvsten (2020)

Unbiased $s_f^2 +$ degrees of freedom correction $\implies$ finite-sample unbiased estimate

Rewrite using cross-products of job-level contact gaps:

\[
\hat{\sigma}_\Delta^2 = \left(\frac{F - 1}{F}\right) \left[ \frac{1}{F} \sum_{f=1}^{F} \frac{2}{J_f (J_f - 1)} \sum_{j=2}^{J_f} \sum_{\ell=1}^{j-1} \hat{\Delta}_{fj} \hat{\Delta}_{f\ell} - \frac{2}{F(F - 1)} \sum_{f=2}^{F} \sum_{k=1}^{f-1} \hat{\Delta}_f \hat{\Delta}_k \right]
\]

Interpretation: $\hat{\sigma}_\Delta^2$ measures covariance between contact gaps across jobs at the same firm
Standard Deviations of $G$: Substantial Variation for Both Race and Gender

<table>
<thead>
<tr>
<th></th>
<th>Mean contact gap</th>
<th>Bias-corrected std. dev. of contact gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Race (White - Black)</td>
<td>0.021</td>
<td>0.0185</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.0031)</td>
</tr>
<tr>
<td>Gender (Male - Female)</td>
<td>-0.001</td>
<td>0.0267</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.0038)</td>
</tr>
</tbody>
</table>

Estimates of firm heterogeneity in race and gender discrimination

Estimates from Kline, Rose, and Walters (forthcoming).

Chris Walters (UC Berkeley) Empirical Bayes Methods
Flexible Deconvolution

► Features of $G$ beyond the mean and variance are also of interest

► Hierarchical model:

$$\hat{\Delta}_f | \Delta_f, s_f \sim N(\Delta_f, s_f^2)$$

$$\Delta_f \sim G(\Delta)$$

► Next, consider flexible deconvolution methods imposing little structure on $G$

► N.B.: Need to account for possible dependence between effect sizes $\Delta_f$ and sampling variance $s_f^2$

► Maybe firms where more jobs were sampled discriminate more/less

► Maybe firms where overall callback rates are higher discriminate more/less

- For now, sidestep precision-dependence by transforming estimates into z-scores

- Let $z_f = \hat{\Delta}_f / s_f$ denote the estimated z-score for firm $f$, and let $\mu_f = \Delta_f / s_f$ denote its population counterpart. Then

$$z_f | \mu_f \sim N(\mu_f, 1)$$

$$\mu_f \sim G_\mu(\mu)$$

- Efron (2016) proposes to approximate $G_\mu$ with distribution in smooth exponential family

  - Parameterize density with flexible spline
  - Estimate spline parameters by penalized maximum likelihood
  - Implemented in deconvolver R package (Narasimhan and Efron, 2020)

  - Requires choosing penalization tuning parameter. Sensible approach: calibrate to match unbiased variance estimate
Flexible Deconvolution: NPMLE

- Alternative approach: Non-parametric maximum likelihood estimator (NPMLE; Robbins, 1950; Kiefer and Wolfowitz, 1956)

- NPMLE picks mixing distribution to maximize likelihood of observed data:

\[
\hat{G}_\mu = \max_{G \in \mathcal{G}} \sum_{f=1}^{F} \log \left( \int \phi (z_f - \mu) \, dG(\mu) \right)
\]

- Solution is a discrete distribution with at most \( F \) mass points

- Koenker and Mizera (2014) develop an approximation that is straightforward to compute with modern convex optimization methods

  - Implemented in REBayes R package (Koenker and Gu, 2017)

- See Koenker (2016) for a comparison of the Efron (2016) and NPMLE approaches
From $z$-scores to Levels

- Suppose we have an estimate $\hat{G}_\mu$ of the distribution of $z$-scores.
- To recover the distribution of $\Delta_f = \mu_f s_f$, need a change of variables.
- Suppose $\mu_f$ is independent of $s_f$, and let $g_\mu$ and $h_s$ denote the densities of $\mu_f$ and $s_f$.
- Density of contact gaps is then

$$g_\Delta(x) = \int \frac{1}{s} g_\mu(x/s) h_s(s) ds$$

- Plug in estimated density $\hat{g}_\mu$ of $z$-scores and empirical distribution of standard errors to compute $\hat{g}_\Delta$.
Histogram of Race Contact Gap Estimates

Chris Walters (UC Berkeley) Empirical Bayes Methods
Deconvolved Distribution of Race Contact Gaps

Implied firm mean gap: 0.0164
Implied between firm SD: 0.0183

Figure A12: Deconvolution of firm-level racial discrimination without support restriction

Notes:
This figure presents non-parametric estimates of the distribution of firm-specific white-Black contact rate differences. The red histogram shows the distribution of estimated firm contact gaps. Blue line shows estimates of the population contact gap distributions. The population distributions are estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to firm-specific z-score estimates, then numerically integrating over the empirical distribution of standard errors to recover the distribution of contact gaps. The penalization parameter in the deconvolution step is calibrated so that the resulting distribution matches the corresponding bias-corrected variance estimate from Table 4.
Deconvolution Imposing Shape Restriction: $\Delta f \geq 0$

Implied firm mean gap: 0.0214
Implied between firm SD: 0.0183

Observed gaps
Deconvolved density

Firm white–black contact rate gap

Density

Firm male–female contact rate gap

Density

Implied firm mean gap: −0.0013
Implied between firm SD: 0.0264

Observed gaps
Deconvolved density

Chris Walters (UC Berkeley)

Empirical Bayes Methods
Figure E7: NPMLE estimates of marginal distributions of firm-level discrimination

a) Race b) Gender

Implied firm mean gap: 0.0189
Implied between firm SD: 0.0183

Notes:
This figure presents non-parametric maximum likelihood estimates of the distribution of firm-specific contact gaps estimated using the approach in Koenker and Gu (2017). Panel (a) presents estimates for white-Black contact rate differences, where we impose the restriction that all contact gaps are weakly positive, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines show estimates of population contact gap distributions. Population distributions are estimated allowing a non-parametric bivariate distribution for the mixing distribution of contact gaps and standard errors. The figures plot the marginal distribution of contact gaps. Since the distribution is discrete, the blue lines plot the probability mass function in below, while the histogram reports the share of sample firms in each bin.
Histogram of Gender Contact Gap Estimates

Chris Walters (UC Berkeley)  Empirical Bayes Methods
Deconvolved Distribution of Gender Contact Gaps

Implied firm mean gap: -0.0013
Implied between firm SD: 0.0264

Chris Walters (UC Berkeley)
Empirical Bayes Methods
NPMLE Estimates for Gender

Figure E7: NPMLE estimates of marginal distributions of firm-level discrimination

(a) Race  
(b) Gender

Implied firm mean gap: 0.00216
Implied between firm SD: 0.0292

Notes:
This figure presents non-parametric maximum likelihood estimates of the distribution of firm-specific contact gaps estimated using the approach in Koenker and Gu (2017). Panel (a) presents estimates for white-Black contact rate differences, where we impose the restriction that all contact gaps are weakly positive, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines shows estimates of population contact gap distributions. Population distributions are estimated allowing a non-parametric bivariate distribution for the mixing distribution of contact gaps and standard errors. The figures plot the marginal distribution of contact gaps. Since the distribution is discrete, the blue lines plot the probability mass function in below, while the histogram reports the share of sample firms in each bin.
Lorenz Curves Derived from Efron (2016) \( \hat{G}'s \)

- **Share of firms**
  - Top 20%: 0.46 (0.027)
  - Top 20%: 0.56 (0.019)

- **Gender, Gini**: 0.536 (0.0192)
- **Race, Gini**: 0.394 (0.0339)

The graph illustrates the distribution of firms and lost contacts, showing that the top 20% of firms explain approximately 50-60% of lost contacts. The Lorenz curves indicate different levels of inequality, with gender and race showing distinct Gini coefficients.
Accounting for Precision-Dependence

- Note: if $\mu_f$ is independent of $s_f$, then effect sizes are increasing in standard errors
  
  - $\Delta_f = \mu_f s_f$, so $E[\Delta_f | s_f] = \bar{\mu} s_f$
  
  - Can test whether this approximation is reasonable

- Other approaches to dealing with dependence:
  
  - Treat $s_f$ as a covariate that shifts location and/or scale of $G$
  
  - Variance-stabilizing transformation: Find function $t(\cdot)$ such that $\text{Var}(t(\hat{\Delta}_f) | \Delta_f)$ is approximately constant (e.g. Brown, 2008)
  
  - Estimate bivariate distribution of $(\Delta_f, s_f)$, e.g. with NPMLE
Separate Deconvolutions for Low vs. High $s_f$

Figure E4: Conditional deconvolutions of firm-level discrimination distributions

(a) Race b) Gender

Notes:
This figure presents non-parametric estimates of the distribution of firm-specific contact gaps estimated separately for firms with above/below median standard errors. Panel (a) presents estimates for white-Black contact rate differences, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps in each group. Blue lines show estimates of population contact gap distributions for each group. The population distributions are estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to firm-specific $z$-score estimates within group, then numerically integrating over the group's empirical distribution of standard errors. A common penalization parameter is used in the deconvolution step for both groups and calibrated so that the resulting marginal distribution matches the corresponding bias-corrected variance estimate from Table 4. In panel (a), the density of population $z$-scores is constrained to be weakly positive in each group.
Marginal Distribution from Separate Deconvolutions

Figure E5: Marginal distributions of firm-level discrimination from conditional approach

(a) Race  
(b) Gender

Implied firm mean gap: 0.0204
Implied between firm SD: 0.0183

Notes:
This figure presents non-parametric estimates of the marginal distribution of firm-specific contact gaps corresponding to the group-specific estimates in Figure E4. Panel (a) presents estimates for white-Black contact rate differences, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines shows estimates of population contact gap distributions. The marginal density is computed as the average of the group-specific densities in Figure E4.

Chris Walters (UC Berkeley)  
Empirical Bayes Methods
Firm-level Posteriors

With an estimate of the mixing distribution \( \hat{G} \) in hand, move on to EB step 3: posterior estimates of firm-level discrimination

EB posterior mean for \( \Delta_f \):

\[
\hat{\Delta}_f^* = s_f \times \frac{\int x \phi(z_f - x) \hat{g}_\mu(x) dx}{\int \phi(z_f - x) \hat{g}_\mu(x) dx}
\]

Compare distributions of:

- Unbiased estimates \( \hat{\Delta}_f \)
- Contact gaps \( \Delta_f \), as implied by Efron (2016) \( \hat{G} \) estimate
- EB posterior means \( \hat{\Delta}_f^* \)
Distribution of Race Contact Gaps

Implied firm mean gap: 0.0214
Implied between firm SD: 0.0183

Observed gaps
Deconvolved density

Firm white–black contact rate gap

Implied firm mean gap: −0.0013
Implied between firm SD: 0.0264
Observed gaps
Deconvolved density

Firm male–female contact rate gap

Discrimination deconvolved
a) Race b) Gender

Chris Walters  (UC Berkeley)  Empirical Bayes Methods
Histogram of Posterior Means

EB approach: Treat deconvolved density as prior to form posterior means

a) Race b) Gender

Implied firm mean gap: 0.0214
Implied between firm SD: 0.0183

Observed gaps
Deconvolved density
Posterior means

Chris Walters (UC Berkeley) Empirical Bayes Methods
As with schools, we may have objectives other than minimizing MSE of discrimination estimates.

May want to make decisions about how to classify specific firms:

- Which firms are discriminating at all ($\Delta_f \neq 0$)?
- Which firms are in the top quintile of discrimination ($\Delta_f > G^{-1}(0.8)$)?

Such decisions are closely related to multiple-testing problems ("large-scale inference;" Efron, 2012).

Next, consider robust EB methods for classifying discriminators.
Multiple Testing

- Suppose we conduct a hypothesis test for each firm, yielding a list of \( p \)-values \( \{p_f\}_{f=1}^F \).

- Example: one-tailed \( t \)-test of \( H_0 : \Delta_f = 0 \) vs. \( H_A : \Delta_f > 0 \)

  - Test statistic: \( z_f = \hat{\Delta}_f / s_f \)

  - \( P \)-value: \( p_f = 1 - \Phi (z_f) \)

- Decision rule: reject all hypotheses with \( p \)-values less than \( \bar{p} \)

- How many mistakes do we expect to make?
By Bayes rule, the expected share of non-discriminators among firms with 
*p*-values below \( \bar{p} \) is:

\[
\Pr [\Delta_f = 0 | p_f \leq \bar{p}] = \frac{\Pr [p_f \leq \bar{p} | \Delta_f = 0] \Pr [\Delta_f = 0]}{\Pr [p_f \leq \bar{p}]}
\]

\[
= \frac{\bar{p} \pi_0}{F_p(\bar{p})}
\]

This quantity is the **False Discovery Rate** (FDR) for our decision rule 
(Benjamini and Hochberg, 1995)

If we can limit FDR to \( \bar{q} \), we should expect 100\( \bar{q} \)% of firms classified as 
discriminators to have \( \Delta_f = 0 \)
Estimating \textit{FDR}

\[
FDR(\bar{p}) = \frac{\bar{p}\pi_0}{F_p(\bar{p})}
\]

\begin{itemize}
\item \textit{P}-values are uniformly distributed under the null, so
\[\Pr[p_f \leq \bar{p} | \Delta_f = 0] = \bar{p}\]
\item Denominator is marginal CDF of \textit{p}-values, estimable from empirical share below \(\bar{p}\)
\item Difficulty is estimating \(\pi_0 = \Pr[\Delta_f = 0]\), the population share of true nulls
\begin{itemize}
\item \(\pi_0\) is a feature of \(G\): \(\pi_0 = \int 1[\Delta = 0]dG(\Delta)\)
\item \(\pi_0\) is not point-identified: can’t tell the difference between worlds where a mass of firms have \(\Delta_f\) exactly 0 vs. vanishingly small
\item Efron (2016) continuous approximation automatically implies \(\hat{\pi}_0 = 0\)
\end{itemize}
\end{itemize}
Bounding $\pi_0$

$$FDR(\bar{p}) = \frac{\bar{p}\pi_0}{F_p(\bar{p})}$$

- Conservative approach: plug in $\pi_0 = 1$ (Benjamini and Hochberg, 1995)
  - Still implies low FDR if many $p$-values close to 0 ($F_p(\bar{p}) \gg \bar{p}$)
- But we can do better
  - Logically inconsistent to have $\pi_0 = 1$ but $F_p(\bar{p}) \gg \bar{p}$
  - $\pi_0$ can’t be 1 if mean or variance of $G \neq 0$
  - We can borrow strength from the ensemble of tests to bound $\pi_0$
Bounding $\pi_0$

- At any point $u$, density of $p$-values is mixture of true nulls (uniform) and false nulls (something else):

$$f_p(u) = \pi_0 + (1 - \pi_0)f_1(u)$$

- Since $f_1(u) \geq 0$, we have $\pi_0 \leq f_p(u)$ for any $u$, so minimum density of $p$-values bounds $\pi_0$ (Efron et al., 2001):

$$\pi_0 \leq \min_u f_p(u)$$

- We expect density of false nulls to be concentrated toward zero $\implies$ tightest bound near 1. Storey (2002) proposes tail-density estimator:

$$\hat{\pi}_0 = \frac{\sum_{f=1}^{F} 1\{p_f > \lambda\}p_f}{(1 - \lambda)F}$$

- Higher $\lambda$ means tighter bound but noiser estimate – Storey et al. (2004) propose bootstrap procedure to select $\lambda$

- Armstrong (2015) provides confidence interval for $\pi_0$


$q$-values for FDR Control

- Given estimated bound $\hat{\pi}_0$, control FDR using $q$-values (Storey, 2003):

\[
q_f = \widehat{FDR}(p_f) = \frac{p_f \hat{\pi}_0}{\hat{F}_p(p_f)}
\]

- $q$-value $\approx$ EB equivalent of $p$-value

- Rather than controlling $\Pr[\text{Reject}_f = 1|\Delta_f = 0]$, use Bayes rule + ensemble of tests to control $\Pr[\Delta_f = 0|\text{Reject}_f = 1]$

- If firm $f$’s $q$-val is $q_f$ and we reject all hypotheses with $p$-vals lower than $p_f$, we should expect at most $100q_f\%$ of rejections to be mistakes
$P$-value Histogram from One-Tailed Tests of $H_0 : \Delta_f \leq 0$

Multiple testing: Goal is to control False Discovery Rate

FDR($p$) = $\Pr(\Delta_f = 0 | \hat{p}_f < p)$ = $p \pi_0 \hat{F}(p)$

Base decisions on $\hat{q}_f = \hat{FDR}(\hat{p}_f)$

e.g., if $\hat{q}_f = 0.05$ then we expect at least 19 out of every 20 firms with $p$-values below $\hat{p}_f$ to have $\Delta_f \neq 0$. 
$\hat{\pi}_0 = 0.39 \implies \text{At Least 61\% of Firms Discriminate Against Black Applicants}$
23 of 108 Firms Have $q_f \leq 0.05$

<table>
<thead>
<tr>
<th>Firm</th>
<th>Industry</th>
<th>Contact gap estimate</th>
<th>Std. err.</th>
<th>$p$ -value</th>
<th>$q$ -value</th>
<th>Posterior mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Auto dealers/services</td>
<td>0.0952</td>
<td>0.0197</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0835</td>
</tr>
<tr>
<td>2</td>
<td>Auto dealers/services</td>
<td>0.0507</td>
<td>0.0143</td>
<td>0.0003</td>
<td>0.0061</td>
<td>0.0354</td>
</tr>
<tr>
<td>3</td>
<td>Auto dealers/services</td>
<td>0.0738</td>
<td>0.0220</td>
<td>0.0005</td>
<td>0.0073</td>
<td>0.0489</td>
</tr>
<tr>
<td>4</td>
<td>Auto dealers/services</td>
<td>0.0787</td>
<td>0.0249</td>
<td>0.0010</td>
<td>0.0103</td>
<td>0.0498</td>
</tr>
<tr>
<td>5</td>
<td>Apparel stores</td>
<td>0.0733</td>
<td>0.0250</td>
<td>0.0022</td>
<td>0.0158</td>
<td>0.0448</td>
</tr>
<tr>
<td>6</td>
<td>Other retail</td>
<td>0.0469</td>
<td>0.0159</td>
<td>0.0020</td>
<td>0.0158</td>
<td>0.0286</td>
</tr>
<tr>
<td>7</td>
<td>Other retail</td>
<td>0.0605</td>
<td>0.0219</td>
<td>0.0033</td>
<td>0.0176</td>
<td>0.0365</td>
</tr>
<tr>
<td>8</td>
<td>General merchandise</td>
<td>0.0520</td>
<td>0.0187</td>
<td>0.0031</td>
<td>0.0176</td>
<td>0.0314</td>
</tr>
<tr>
<td>9</td>
<td>Auto dealers/services</td>
<td>0.0613</td>
<td>0.0240</td>
<td>0.0060</td>
<td>0.0194</td>
<td>0.0370</td>
</tr>
<tr>
<td>10</td>
<td>Other retail</td>
<td>0.0560</td>
<td>0.0214</td>
<td>0.0050</td>
<td>0.0194</td>
<td>0.0337</td>
</tr>
<tr>
<td>11</td>
<td>Eating/drinking</td>
<td>0.0560</td>
<td>0.0222</td>
<td>0.0064</td>
<td>0.0194</td>
<td>0.0339</td>
</tr>
<tr>
<td>12</td>
<td>Auto dealers/services</td>
<td>0.0540</td>
<td>0.0215</td>
<td>0.0068</td>
<td>0.0194</td>
<td>0.0327</td>
</tr>
<tr>
<td>13</td>
<td>Food stores</td>
<td>0.0511</td>
<td>0.0204</td>
<td>0.0069</td>
<td>0.0194</td>
<td>0.0310</td>
</tr>
<tr>
<td>14</td>
<td>General merchandise</td>
<td>0.0427</td>
<td>0.0170</td>
<td>0.0068</td>
<td>0.0194</td>
<td>0.0259</td>
</tr>
<tr>
<td>15</td>
<td>Furnishing stores</td>
<td>0.0400</td>
<td>0.0159</td>
<td>0.0066</td>
<td>0.0194</td>
<td>0.0242</td>
</tr>
<tr>
<td>16</td>
<td>Wholesale nondurable</td>
<td>0.0386</td>
<td>0.0158</td>
<td>0.0080</td>
<td>0.0199</td>
<td>0.0235</td>
</tr>
<tr>
<td>17</td>
<td>Apparel manufacturing</td>
<td>0.0350</td>
<td>0.0142</td>
<td>0.0078</td>
<td>0.0199</td>
<td>0.0213</td>
</tr>
<tr>
<td>18</td>
<td>Building materials</td>
<td>0.0373</td>
<td>0.0157</td>
<td>0.0093</td>
<td>0.0218</td>
<td>0.0229</td>
</tr>
<tr>
<td>19</td>
<td>Health services</td>
<td>0.0544</td>
<td>0.0240</td>
<td>0.0132</td>
<td>0.0292</td>
<td>0.0339</td>
</tr>
<tr>
<td>20</td>
<td>Furnishing stores</td>
<td>0.0400</td>
<td>0.0183</td>
<td>0.0152</td>
<td>0.0322</td>
<td>0.0252</td>
</tr>
<tr>
<td>21</td>
<td>Eating/drinking</td>
<td>0.0340</td>
<td>0.0159</td>
<td>0.0172</td>
<td>0.0346</td>
<td>0.0217</td>
</tr>
<tr>
<td>22</td>
<td>General merchandise</td>
<td>0.0423</td>
<td>0.0210</td>
<td>0.0229</td>
<td>0.0439</td>
<td>0.0277</td>
</tr>
<tr>
<td>23</td>
<td>Insurance/real estate</td>
<td>0.0278</td>
<td>0.0140</td>
<td>0.0257</td>
<td>0.0472</td>
<td>0.0183</td>
</tr>
</tbody>
</table>
EB for Decision-Making

▷ What feature of posterior should we use for decisions? As usual, depends on our objectives

▷ Suppose an auditor is interested in investigating discriminators, with utility function

\[
U(\delta) = \sum_{f=1}^{F} \delta_f \left( \Delta_f^{1/\rho} - c \right)
\]

▷ \(\delta_f \in \{0, 1\}\) is investigation indicator, \(c\) is investigation cost, \(\rho \geq 1\) indexes risk aversion

▷ With prior \(G\) and evidence \(\mathcal{E} = \{\hat{\Delta}_f, s_f\}_{f=1}^{F}\), expected-utility maximizing rule is:

\[
\delta_f^* = 1 \left\{ E_G \left[ \Delta_f^{1/\rho} | \mathcal{E} \right] > c \right\}
\]
EB for Decision-Making

- When $\rho = 1$, $\delta_f^* = 1 \{\Delta_f^* > c\}$
  - Risk-neutral auditor investigates based on posterior mean

- When $\rho \to \infty$, $\delta_f^* = 1 \{\Pr_G [\Delta_f = 0|\mathcal{E}] < 1 - c\}$
  - Risk-averse auditor investigates based on local false discovery rate – motivates FDR cutoff rule
  - $q$-value decision rule motivated by optimizing against least-favorable $G$ (highest $\pi_0$) in identified set

- See Kline and Walters (2021) for minimax approach to job-level discrimination with partial identification of $G$
Notes:
This figure illustrates the expected number of contacts per thousand Black applications sent that would be saved if discrimination were eliminated at all firms below a ranking threshold. We consider four rankings: infeasible ranking by true contact gaps ($\Delta f$), ranking by posterior means ($\bar{\Delta f}$), ranking by linear shrinkage estimates ($\tilde{\Delta f}$), and ranking by $q$-values ($\hat{q}_f$). The dashed black line shows the results of ranking firms randomly.
Thanks

▶ Feel free to contact us with questions or issues:

▶ Jiaying: jiaying.gu@utoronto.ca
▶ Chris: crwalters@econ.berkeley.edu

▶ Data and code for employment discrimination application available online:

▶ https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HL04XC
▶ Try it out yourself!
References


References


References

References