Inaccurate Statistical Discrimination: An Identification Problem

Aislinn Bohren, Kareem Haggag, Alex Imas & Devin G. Pope

Discrimination in economics

Literature defines two potential sources of discrimination.

- **Taste- or preference-based:** due to animus against members of a group (Becker 1957).
- **Statistical:** uncertainty about productivity; group identity informative – believe group discriminated against has:
  - Lower avg. productivity (Phelps 1972; Arrow 1973);
  - Higher/lower variance in productivity (Aigner Cain 1977; Bartoš Bauer Chytilová Matějka 2016);
  - Less/more informative or biased signal of productivity.

Methods to distinguish between these sources often rely on assumption that beliefs are accurate.

Inaccurate beliefs

- Evidence that people have systematic biases in beliefs.
  - Inaccurate stereotypes based on representativeness heuristic (Bordalo et al. (2016)).
  - Enormous literature in psychology discussing how stereotypes can be inaccurate (see Fiske (2018) for review).
- Inaccurate beliefs may arise from a lack of information.
  - Inexperience (e.g. aliens arrive from Mars).
  - Subjective information: unsure of others' beliefs + preferences (e.g. interpreting a recommendation letter).
  - Dynamics: accurate priors about population averages but unsure how selection impacts current population (e.g. hiring & promotion criteria).

Inaccurate beliefs affect discrimination

- Evaluators use beliefs to learn about trait of candidate (e.g. productivity).
- Inaccurate beliefs lead to inaccurate inference about the candidate.
  - May persist even in the face of a lot of information about candidate (learning with model misspecification literature).
  - May cause inefficient exit from market.
  - Bohren Imas Rosenberg (2019): dynamics of discrimination with inaccurate beliefs.

Why do inaccurate beliefs matter?

Methodological

- Erroneously assuming accurate beliefs leads to misidentification of source in outcome-based tests — common method used to distinguish source.
- Relaxing assumption of accurate beliefs creates an identification problem.

Policy

- Effective interventions when beliefs are inaccurate differ from those for accurate statistical or taste-based sources.
- Welfare implications also differ.

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%
Discuss accurate versus inaccurate beliefs	11	10.5%

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%
Discuss accurate versus inaccurate beliefs	11	10.5%
Test for inaccurate beliefs	7	6.7%

Literature survey

Most economics papers on discrimination ignore the possibility of inaccurate beliefs.

	# Papers	Percent
Total Papers	105	100.0%
Evidence of Discrimination	102	97.1%
Discuss taste-based versus statistical source	65	61.9%
Test for taste-based versus statistical source	49	46.7%
Discuss accurate versus inaccurate beliefs	11	10.5%
Test for inaccurate beliefs	7	6.7%
Measure beliefs	7	6.7%

Scope: empirical papers published between 1990-2018 in AEJ: Applied, AEJ: Policy, AER, EMA, JEEA, JLE, JPE, ReStud, ReStat, QJE.

A model of discrimination with inaccurate beliefs

Worker:

- Group identity $g \in \{M, F\}$.
- Productivity $a \sim N(\mu_g, 1/\tau_g)$.
- Signal $s = a + \epsilon$, $\epsilon \sim N(0, 1/\eta_g)$.
- Focus on discrimination against group F.

A model of discrimination with inaccurate beliefs

Worker:

- Group identity $g \in \{M, F\}$.
- Productivity $a \sim N(\mu_g, 1/\tau_g)$.
- Signal $s = a + \epsilon$, $\epsilon \sim N(0, 1/\eta_g)$.
- Focus on discrimination against group F.

Evaluator:

- Decides whether to hire worker, $v \in \{0, 1\}$.
- Holds subjective beliefs $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$.
  - Misspecified model when subjective parameters $\neq$ true parameters.
- Observes g and s; uses Bayes rule to updates belief about a.
- Hires worker if subjective posterior $\hat{E}_\theta[a|s, g]$ above group-specific hiring threshold $u_g \in \mathbb{R}$ (reduced form for preferences).
- $\theta = \{u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g\}_{g \in \{M,F\}}$ denotes evaluator's type.

Partiality

Categorize different forms of preferences and beliefs.

- Use partiality to refer to properties of model primitives.
- To distinguish from discrimination – property of behavior and consequence of primitives.

Definition (Preference Partiality)

An evaluator has *preference partiality* if $u_F \neq u_M$.

Definition (Belief Partiality)

An evaluator has *belief partiality* if $(\hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) \neq (\hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M)$. Beliefs are accurate if $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) = (\mu_g, \tau_g, \eta_g)$ for $g \in \{M, F\}$ and otherwise inaccurate.

Discrimination

- Difference between hiring decisions

$$D(s,\theta) \equiv v(s,M,\theta) - v(s,F,\theta)$$

where $v(s,g,\theta) \equiv \mathbb{1}\{\hat{E}_\theta[a|s,g] \geq u_g\}$ denotes optimal hiring decision.

- Discrimination occurs against group F if $\exists s$ s.t. $D(s,\theta) > 0$.
- Interested in when different sets of beliefs and preferences give rise to same discriminatory behavior.

Definition

Two evaluators of types θ and θ' exhibit equivalent discrimination if $D(s,\theta) = D(s,\theta')$ for all $s \in \mathbb{R}$.

Equivalent discrimination

Posterior mean productivity monotonic wrt s

- Represent optimal hiring decision as cut-off rule wrt signal: hire if $s \geq \bar{s}(\theta, g)$.
- Two types have same signal cut-offs $\Rightarrow$ equivalent discrimination.

Equivalent discrimination

Posterior mean productivity monotonic wrt s

- Represent optimal hiring decision as cut-off rule wrt signal: hire if $s \geq \bar{s}(\theta, g)$.
- Two types have same signal cut-offs $\Rightarrow$ equivalent discrimination.

Proposition (Set of types that exhibit equivalent discrimination)

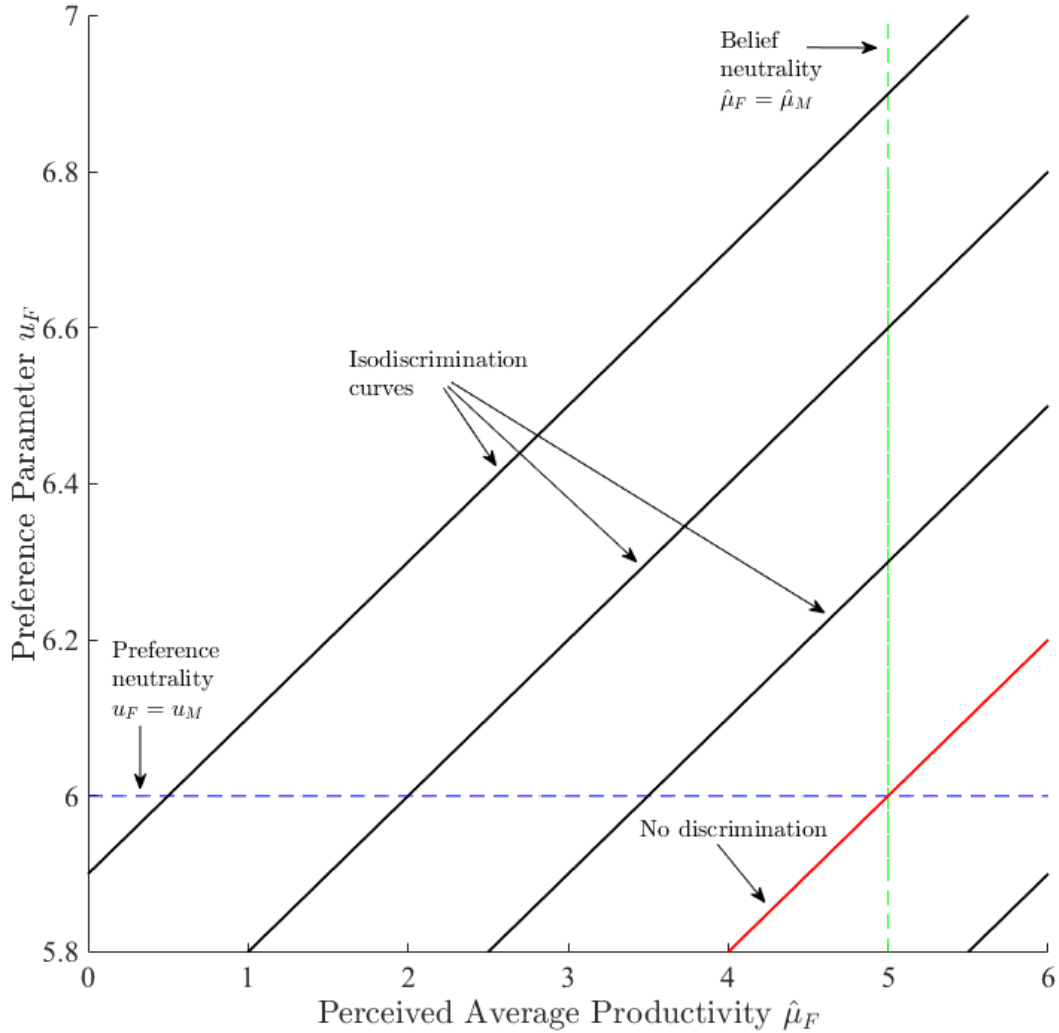
For any constants $(s_M, s_F) \in \mathbb{R}^2$ with $s_M \neq s_F$, equivalent discrimination occurs for the set of types $\{\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)\}_{g \in \{M,F\}}$ s.t.

$$\frac{\hat{\tau}_M + \hat{\eta}_M}{\hat{\eta}_M} u_M - \frac{\hat{\tau}_M}{\hat{\eta}_M} \hat{\mu}_M = s_M \tag{1}$$

$$\frac{\hat{\tau}_F + \hat{\eta}_F}{\hat{\eta}_F} u_F - \frac{\hat{\tau}_F}{\hat{\eta}_F} \hat{\mu}_F = s_F. \tag{2}$$

No discrimination occurs for types that satisfy (1), (2) when $s_M = s_F$.

Isodiscrimination curves



$(u_M, \hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M) = (6, 5, .5, 2)$, $(\hat{\tau}_F, \hat{\eta}_F) = (.5, 2)$

Identifying discrimination

Can property of interest can be backed out from available data?

- When evaluators may have inaccurate beliefs, what can we say about preferences and beliefs in common empirical designs used to study discrimination?

Existence of discrimination

Show there exists an s such that $D(s, \theta) > 0$.

- Assume researcher observes group g and hiring decision v for each worker.
- If researcher also observes s, this is straightforward (direct method).
- Otherwise, can use correspondence or audit study method: create fictitious workers with randomly assigned g and s.
- Provided observe hiring decisions for sufficiently rich set of signals, can also identify isodiscrimination curve.

Source of discrimination

Proceed with assuming researcher can identify isodiscrimination curve (i.e. using correspondence study).

- What form of partiality generates observed discriminatory behavior?
- Observing isodiscrimination curve does not allow researcher to distinguish between preference, accurate belief and inaccurate belief partiality.
- Moreover, not possible to rule out any potential source.

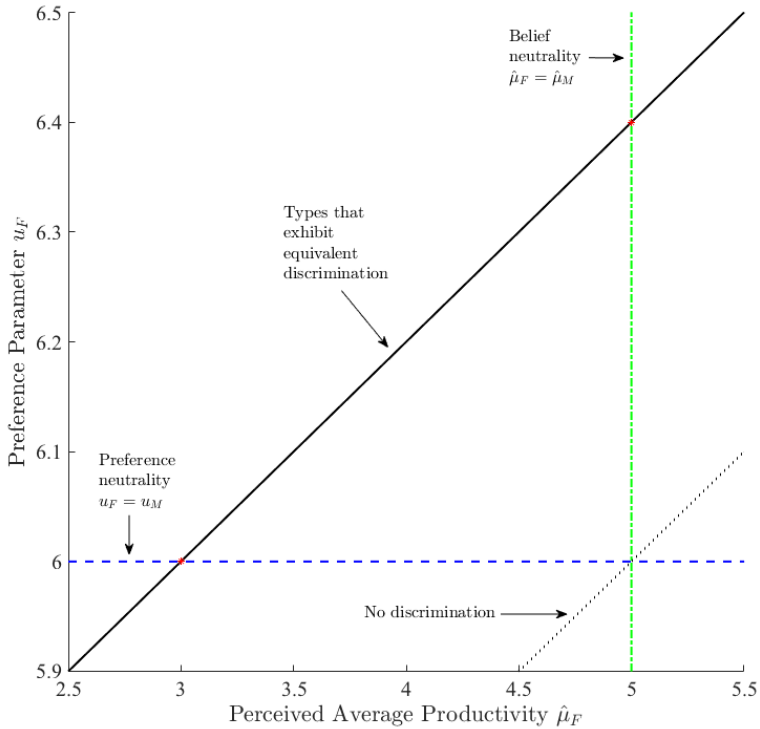
Each form of partiality in isolation can generate given pattern of discrimination

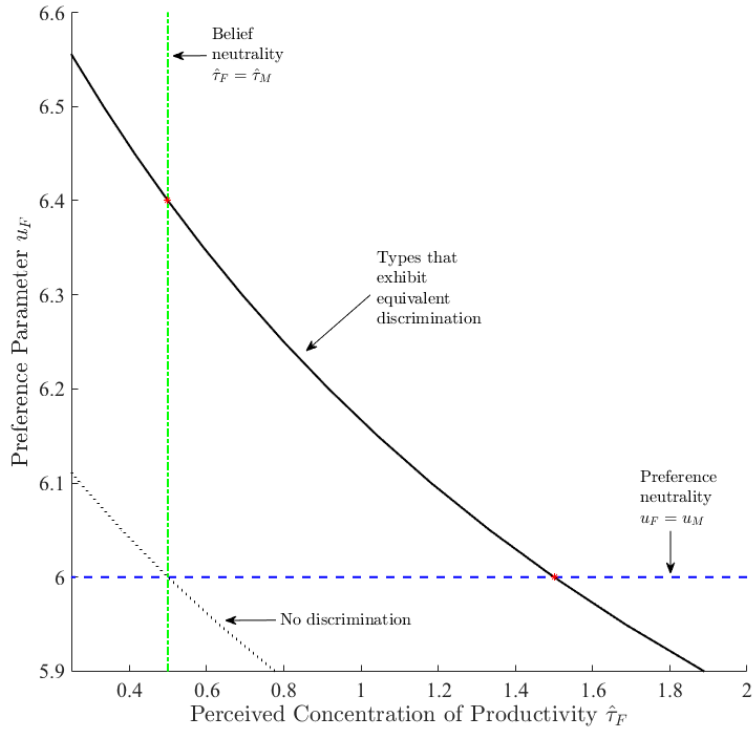
Proposition

For any type θ that discriminates against group F, there are a continuum of types that exhibit equivalent discrimination, including:

1. *A type θ' with preference partiality against group F and belief neutrality, $u'_F > u'_M$;*
2. *A type θ' with preference neutrality and belief partiality due to lower expected productivity, $\hat{\mu}'_F < \hat{\mu}'_M$;*
3. *A type θ' that believes the market is cherry-picking (lemon-dropping) and has preference neutrality and belief partiality due to higher (lower) concentration of productivity, $\hat{\tau}'_F > \hat{\tau}'_M$;*
4. *A type θ' that believes the market is cherry-picking (lemon-dropping) and has preference neutrality and belief partiality due to lower (higher) signal precision, $\hat{\eta}'_F < \hat{\eta}'_M$ ($\hat{\eta}'_F > \hat{\eta}'_M$).*

Each form of partiality in isolation can generate given pattern of discrimination





Isodiscrimination curve $(s_M, s_F) = (6.25, 6.75)$; $(u_M, \hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M) = (6, 5, .5, 2)$

Left figure: $(\hat{\tau}_F, \hat{\eta}_F) = (.5, 2)$; Right figure: $(\hat{\mu}_F, \hat{\eta}_F) = (5, 2)$

Source of discrimination

Corollary

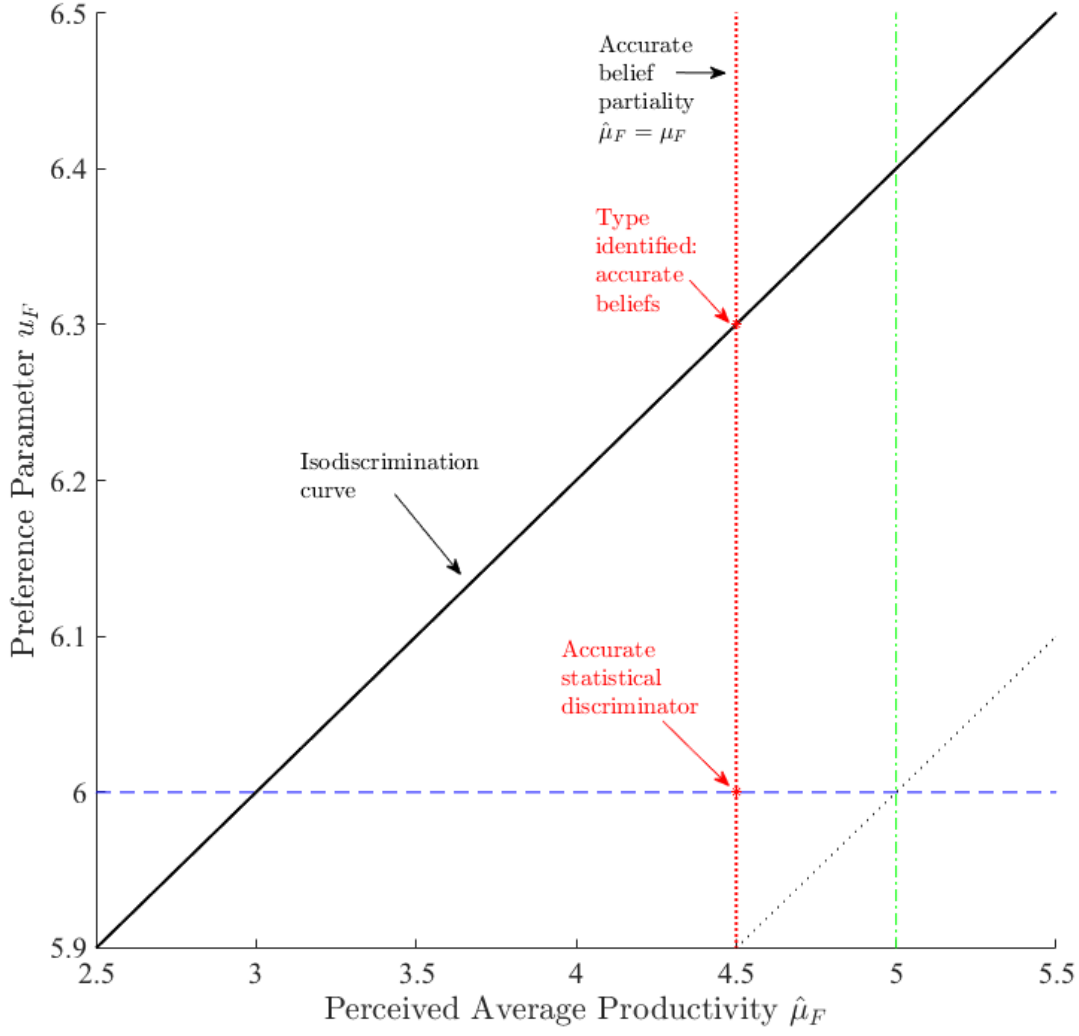
Observing the isodiscrimination curve does not rule out any category of partiality.

Outcomes-based test

Suppose it is also possible to observe the true productivity and signal distributions for each group.

- Common method to identify source of discrimination.
- Under assumption of accurate beliefs, this identifies evaluator's type.

Outcomes-based test: accurate beliefs



Isodiscrimination curve $(s_M, s_F) = (6.25, 6.75)$; true distributions $(\mu_M, \tau_M, \eta_M) = (5, .5, 2)$, $(\mu_F, \tau_F, \eta_F) = (4.5, .5, 2)$; $(u_M, \hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M) = (6, 5, .5, 2)$.

Identification crucially depends on accurate belief assumption

- Type no longer identified when inaccurate beliefs are possible.
- Erroneously assuming accurate beliefs leads researcher to misattribute discrimination from inaccurate beliefs to preferences.
- Inaccurate beliefs + preference partiality can be mistaken for accurate beliefs.

Proposition (Misidentified Preferences)

Suppose a researcher incorrectly assumes an evaluator has accurate beliefs and uses the outcomes-based method to identify the type.

1. *For a generic set of types and true distributions, the researcher misidentifies preferences.*
2. *If inaccurate beliefs increase (decrease) discrimination against group F, then the researcher overestimates (underestimates) the evaluator's preference partiality against group F.*

Dynamics of inaccurate beliefs

Suppose inaccurate beliefs + preference partiality generate equivalent discrimination to accurate statistical discrimination.

- May be tempted to conclude inaccurate beliefs innocuous.
- But inaccurate beliefs may negatively affect worker in future performance evaluations and promotions.
- Example:

Dynamics of inaccurate beliefs

Suppose inaccurate beliefs + preference partiality generate equivalent discrimination to accurate statistical discrimination.

- May be tempted to conclude inaccurate beliefs innocuous.
- But inaccurate beliefs may negatively affect worker in future performance evaluations and promotions.
- Example:
  - Inaccurate beliefs exaggerate true difference in avg. productivity between groups.
  - Preferences somewhat favor disadvantaged group through lower hiring threshold ($u_F < u_M$) for entry-level position.
  - Evaluator only feels compelled to favor disadvantaged group for entry-level hiring.
  - $\Rightarrow$ Inaccurate beliefs lead to persistently lower rates of promotion and advancement.

Sometimes possible to rule out accurate statistical discrimination

- Is pattern of discrimination consistent with accurate belief partiality + preference neutrality?
- Of particular interest, since ruling out establishes discrimination stems from animus towards group or inaccurate beliefs about them.

Sometimes possible to rule out accurate statistical discrimination

- Is pattern of discrimination consistent with accurate belief partiality + preference neutrality?
- Of particular interest, since ruling out establishes discrimination stems from animus towards group or inaccurate beliefs about them.

Proposition (Rejecting accurate statistical discrimination)

Suppose a researcher observes the true productivity and signal distributions for each group and can identify the isodiscrimination curve with thresholds (s_M, s_F)*. If*

$$\frac{\tau_M \mu_M + \eta_M s_M}{\tau_M + \eta_M} \neq \frac{\tau_F \mu_F + \eta_F s_F}{\tau_F + \eta_F}, \tag{3}$$

the evaluator is not an accurate statistical discriminator.

Source of discrimination

So how can we identify the source of discrimination when beliefs may be inaccurate?

Source of discrimination

So how can we identify the source of discrimination when beliefs may be inaccurate?

- Elicit subjective beliefs.
  - Analogous technique to outcomes-method.
  - Identifies preferences.
  - Does not identify whether beliefs are accurate – to do so, must also observe true distributions.

Source of discrimination

So how can we identify the source of discrimination when beliefs may be inaccurate?

- Elicit subjective beliefs.
  - Analogous technique to outcomes-method.
  - Identifies preferences.
  - Does not identify whether beliefs are accurate – to do so, must also observe true distributions.
- Manipulate information.
  - Vary number of signal draws evaluator observes.
  - Important caveat: crucial that multiple signals drawn from same distribution.
  - Online reviews vs. SAT score + GPA.

Manipulating information

Proposition (Set of types that exhibit equivalent discrimination)

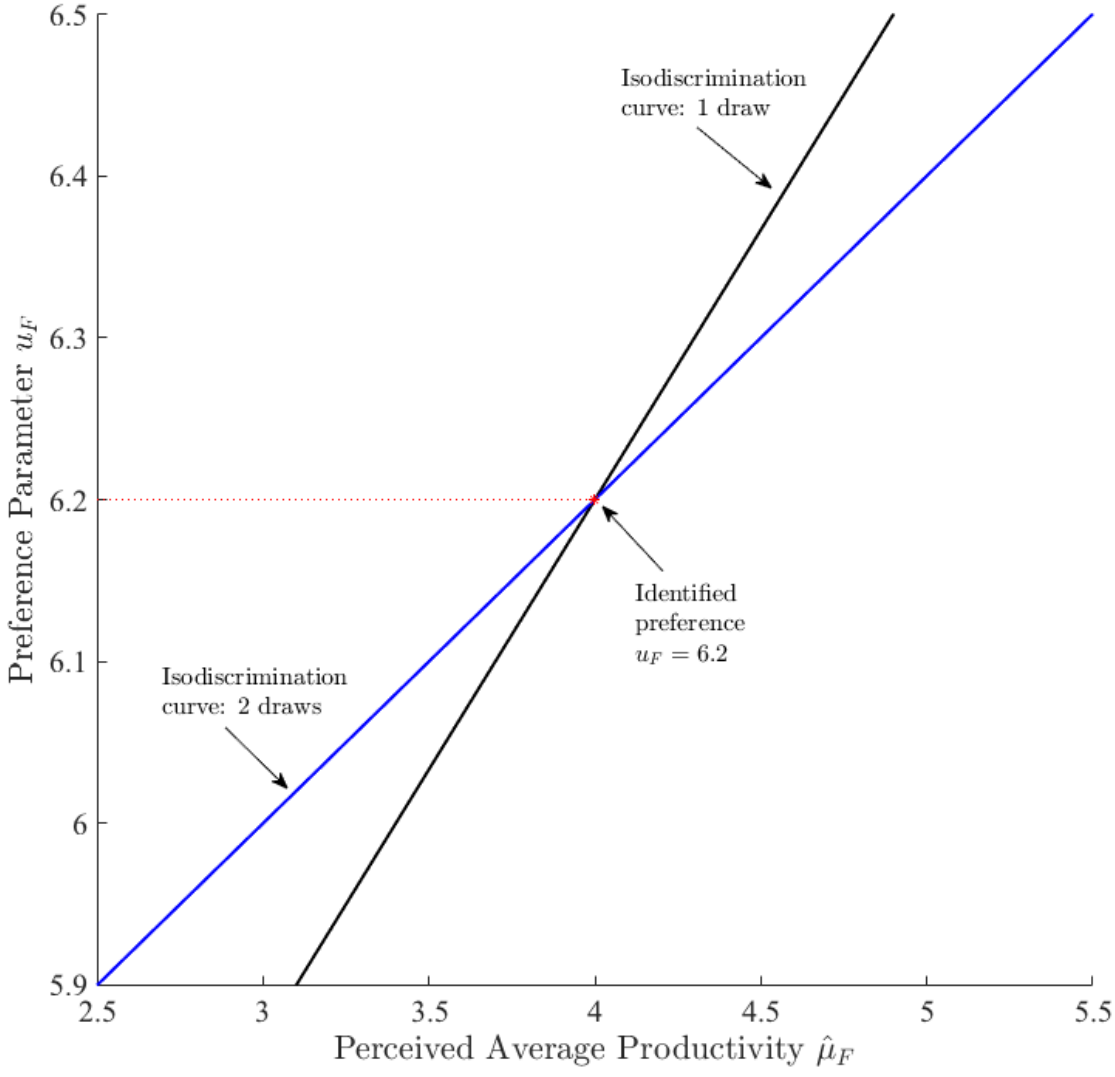
Suppose an evaluator of type $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g \in \{M,F\}}$ *observes either* x_1 *or* $x_2 \neq x_1$ *signal draws. Then the set of types*

$$\left\{ \theta' = (u_g, \hat{\mu}'_g, \hat{\tau}'_g, \hat{\eta}'_g)_{g \in \{M,F\}} \middle| \hat{\mu}'_g = u_g - \frac{\hat{\tau}_g / \hat{\eta}_g}{\hat{\tau}'_g / \hat{\eta}'_g}(u_g - \hat{\mu}_g) \right\}$$

exhibit equivalent discrimination to θ *across both informational treatments. This set of types also exhibit equivalent discrimination for any number* $x \geq 1$ *of signal draws.*

- Unique level of preference partiality yields equivalent discrimination across informational treatments.
- Identifies preferences but not beliefs.

Information manipulation



Isodiscrimination curves for $(u_F, \hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) = (6.2, 4, .5, 1)$

Experiment

Illustrative example

- Demonstrate importance of inaccurate beliefs in how we study discrimination.
- Suggest possible intervention to correct inaccurate beliefs.

Experiment

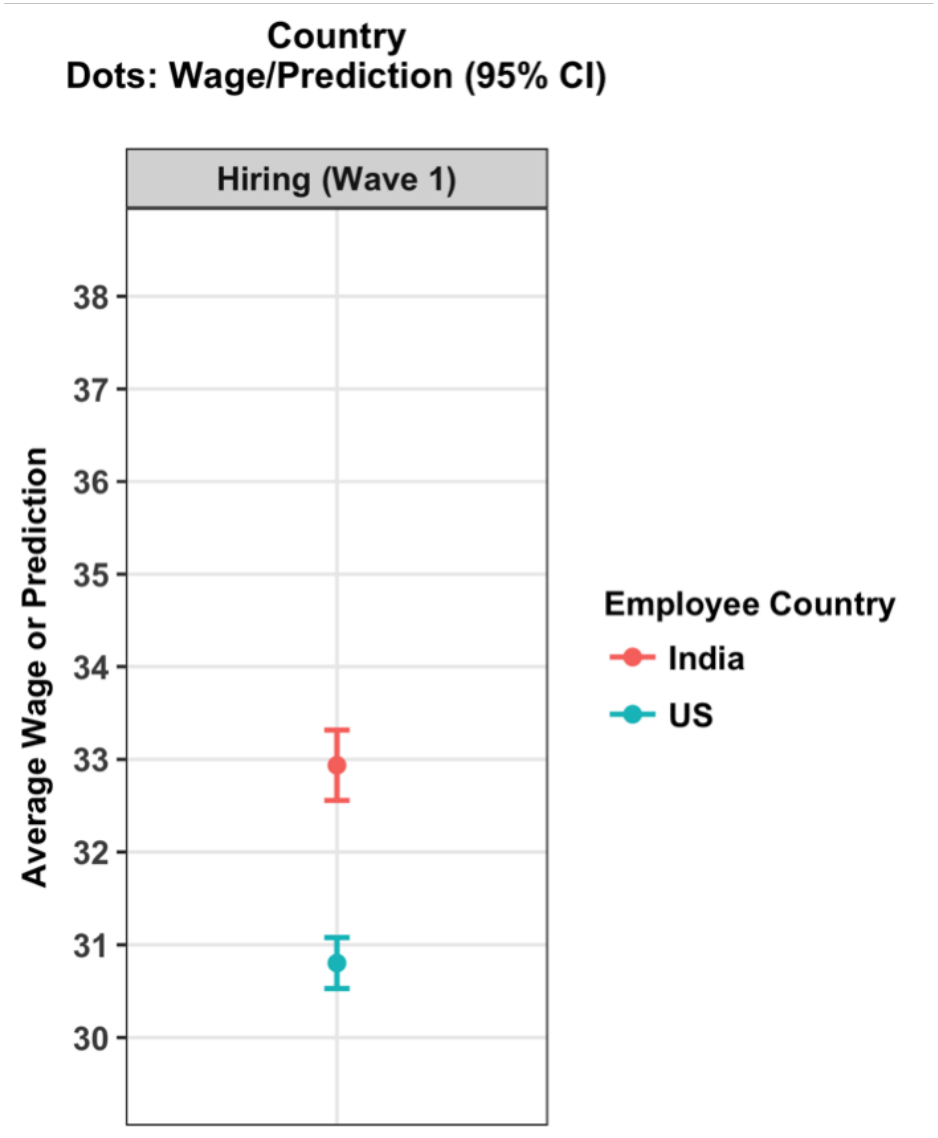
- **Part 1:** 600 Mturkers (400 from US; 200 from India) recruited to be "employees."
  - Completed 50 question math test.
  - Answered 8 questions to develop personal profile.
- **Part 2:** 600 different Mturkers (400 from US; 200 from India) recruited to be "employers."
  - Observed profiles of 20 different employees.
  - Decided maximum wage (wtp) to hire each (0-50 cents).
- **Payment**
  - Drawn random number x between 0 and 50 for each potential employee; if wtp $\geq x$, employer hires employee and pays wage x.
  - Employers earn 1 cent for each question that hired employee answered correctly.
  - Employers told avg # correct answers was 37.
  - Comprehension questions to check understanding.

Step 1: check for differential treatment

Three sets of group identities in experiment:

1. **Indian vs. American**
2. Male vs. Female
3. Old vs. Young

Indians employees receive higher wages than Americans



Summary

Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Given the conventional methods for studying discrimination, how would we test for statistical vs. taste-based discrimination?

Summary

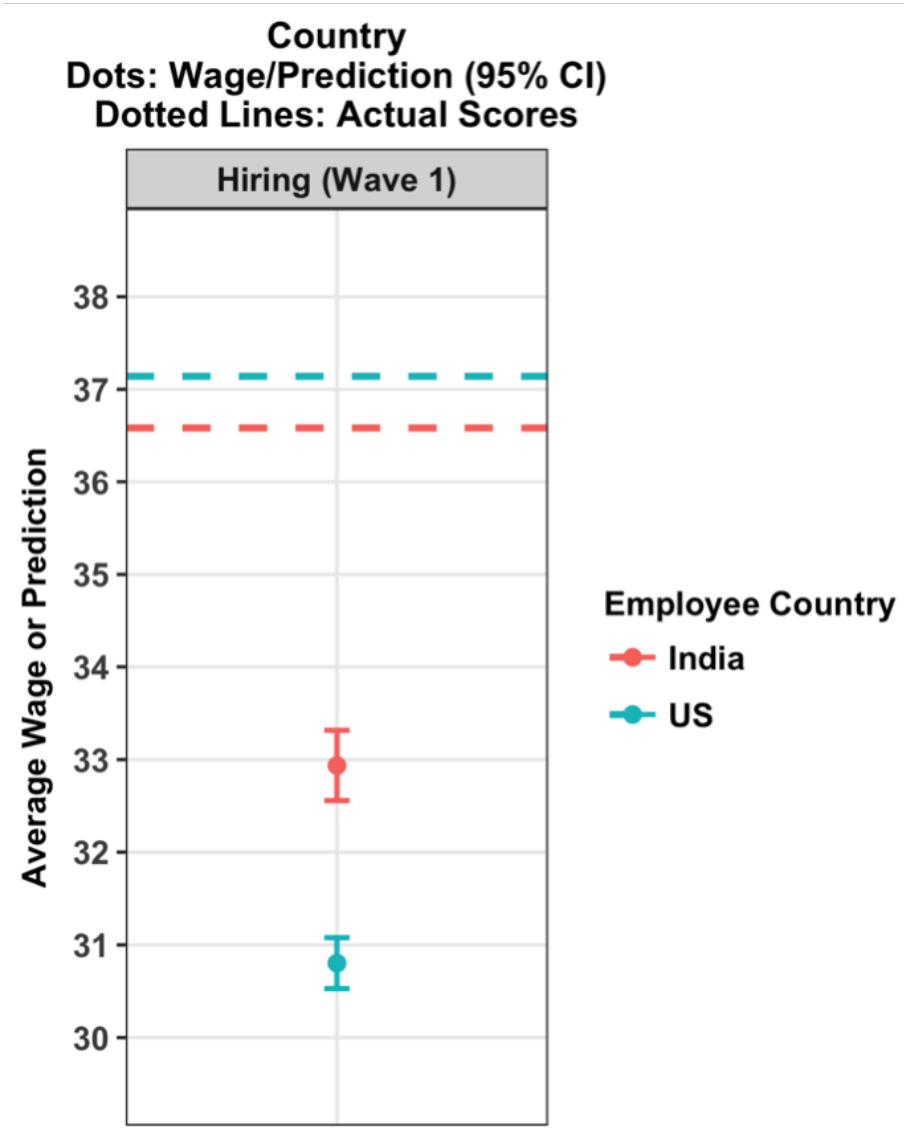
Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Given the conventional methods for studying discrimination, how would we test for statistical vs. taste-based discrimination?

- Compare to true performance distributions.

Taste-based discrimination against Americans.



Summary

Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Part 2: Compare to True Distributions (e.g. standard technique)

- Taste-based discrimination against Americans.

But what if beliefs are wrong?

Part 3: Elicit Average Beliefs

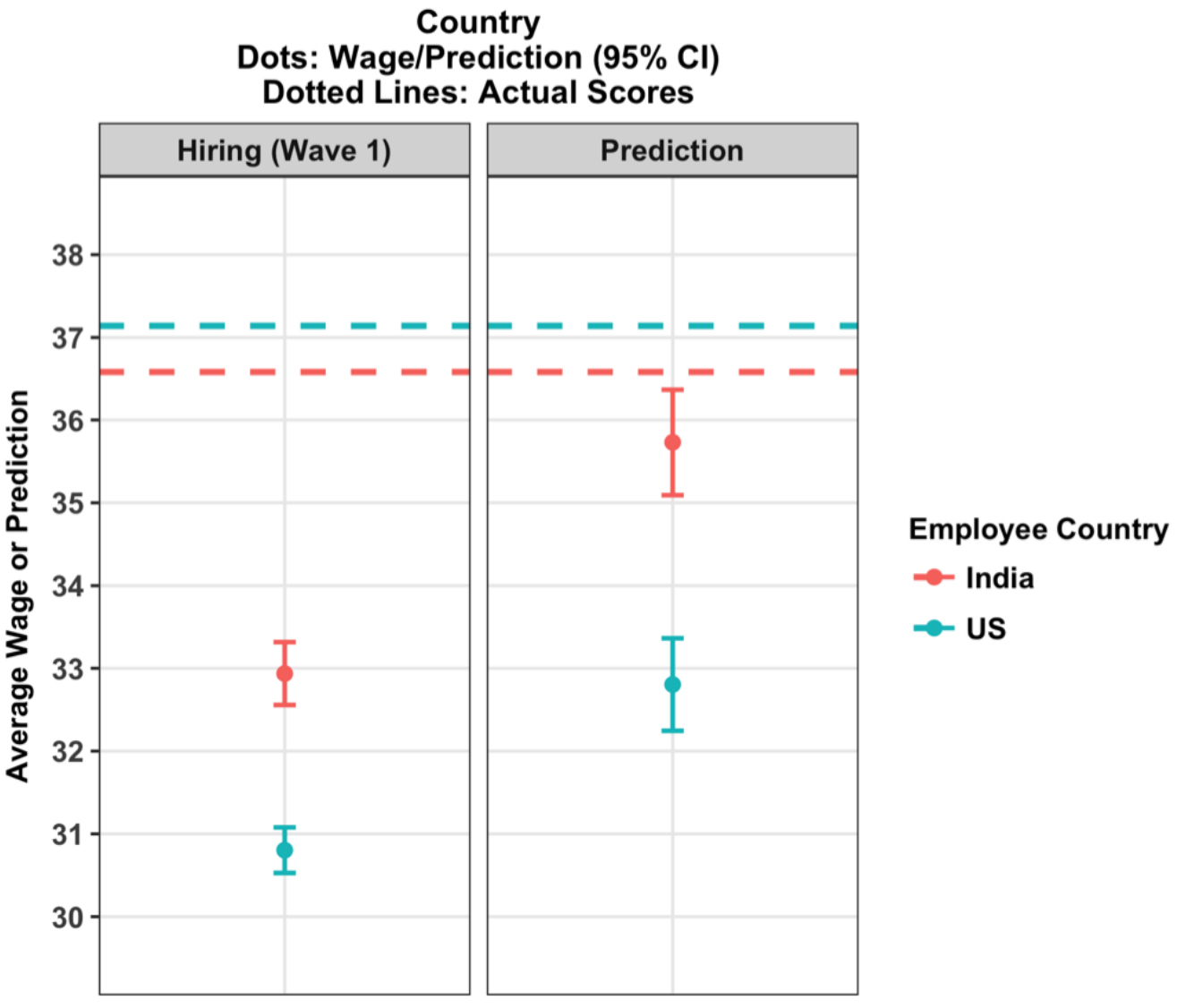
Employers asked following question:

"On average, how many questions do you think g answered correctly?"

where $g \in \{Americans, Indians\}$.

- Provided incentive compatible bonuses for performance on questions.
- Randomized between small and large incentives – no differences between two groups.

Inaccurate statistical discrimination against Americans.



Summary

Part 1: Document Differential Treatment

- Indians favored relative to Americans.

Part 2: Compare to True Distributions (e.g. standard technique)

- Taste-based discrimination against Americans.

Part 3: Compare True and Perceived Distributions

- Inaccurate statistical discrimination against Americans.

Intervention: is it possible to correct beliefs?

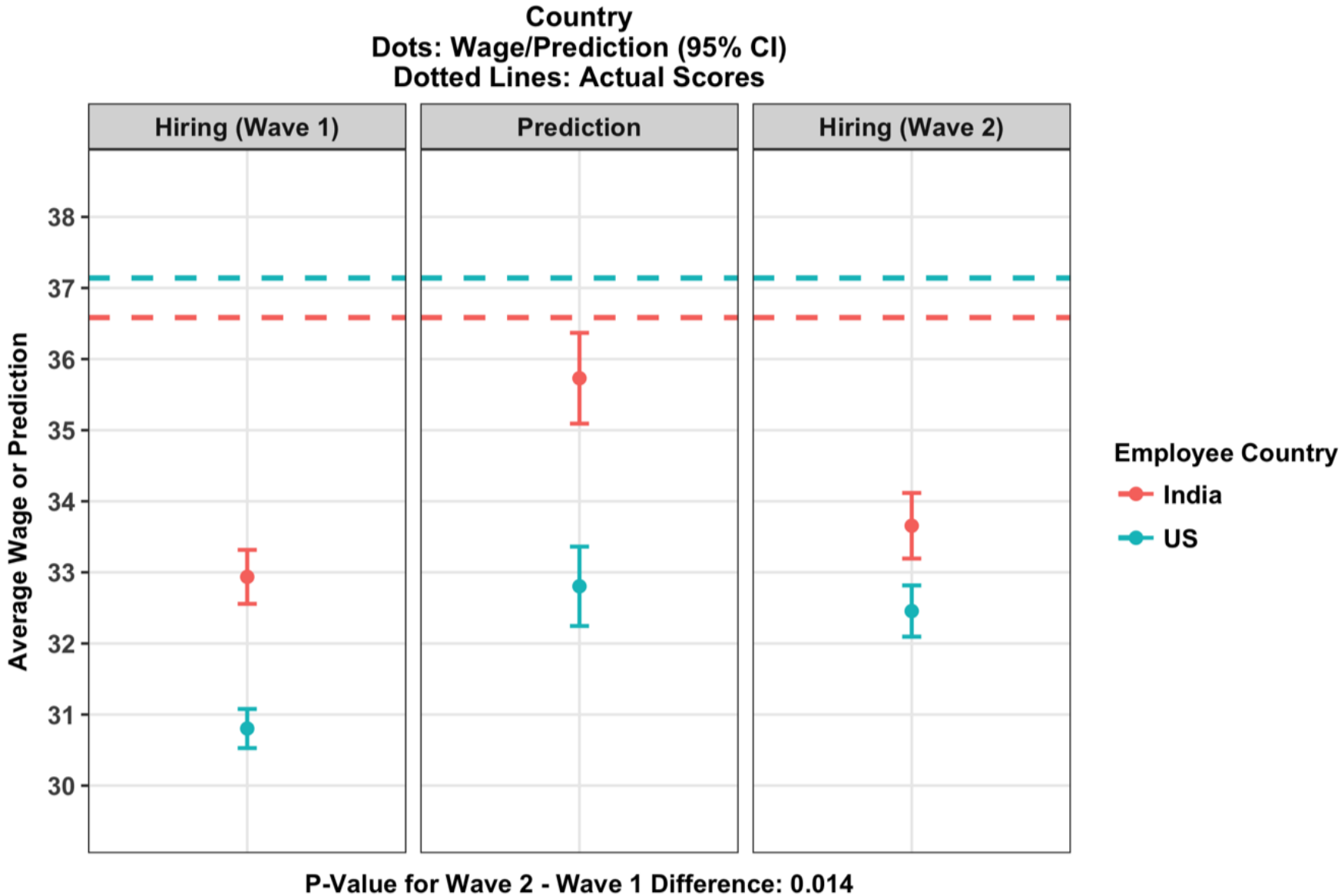
Employers were told:

"Here are the correct answers for the belief questions that you answered."

- Americans answered 37.14 questions correctly on average.
- Indians answered 36.58 questions correctly on average.

Asked to make wage offers based on 10 additional employee profiles (same hiring/payment scheme).

Either taste-based or still some inaccurate statistical discrimination against Americans.



Conclusions

Inaccurate beliefs may be prevalent in many domains currently being studied by discrimination researchers.

- Most tests to separate statistical and taste-based sources rely on strong assumption: people hold accurate beliefs.
- Incorrectly imposing this assumption can lead to mistaken attribution of source, and therefore, misguided policy/welfare analysis.
- Relaxing this assumption generates an identification problem.
- Illustrative example using results from an online experiment demonstrates the importance of inaccurate beliefs in how we study discrimination.

Conclusions: Two Key Forms of Inaccurate Beliefs

Inaccurate priors / beliefs about population.

- Persistent discrimination due to selection effects / exit
- May not "wash out" in the long-run.
- Discrimination reversals can be sign of selection (Bohren Imas Rosenberg (2019)).

Inaccurate beliefs about signals.

- Persistent discrimination due to incorrect updating.
- Could be about exogenous signals (e.g. how predictive is a test) or endogenous signals (e.g. interpreting reference letter).