Identifying Prediction Mistakes in Observational Data

The Quarterly Journal of Economics, August 2024

Ashesh Rambachan

Presenter: Joe Paul

Motivation

- Expert decision-makers make consequential decisions based on predictions of an unknown outcome

Motivation

- Expert decision-makers make consequential decisions based on predictions of an unknown outcome
- Common situations in Economics:
    - Judges deciding pretrial release
        - Must predict whether a defendant will fail to appear in court
    - Doctors making diagnoses
        - Predict if patient suffers from undiagnosed condition
    - Managers making hiring decisions
        - Predict future worker productivity

Core Challenges

Three key identification challenges make this a challenging econometric problem:

1. Decision-makers observe private information relevant to predicting the outcome
2. Unknown preferences that may vary across decisions
3. Missing data - outcomes only selectively observed for some choices

Core Challenges

Three key identification challenges make this a challenging econometric problem:

1. Decision-makers observe private information relevant to predicting the outcome
2. Unknown preferences that may vary across decisions
3. Missing data - outcomes only selectively observed for some choices

Existing research relies on strong assumptions:

- Restrict preferences to be fixed across decisions and DMs
- Observed choices as good as randomly assigned (Assume away private information)
- Parametric models of private information

Framework Overview

A unifying framework to analyse systematic prediction mistakes under weak assumptions on preferences and information sets in general observational settings.

1. Model choices through expected utility maximisation
2. Test if choices consistent with accurate beliefs
3. Allow for:
   - Flexible private information
   - Heterogeneous preferences
   - Missing outcomes
4. Apply to pretrial release in NYC

Model Setup & Identification Challenge

For each decision i:

- Characteristics: X_i (observed by DM and researcher)
- Binary choice: $C_i \in \{0, 1\}$
- Latent outcome: Y_i^* (unknown at time of choice)

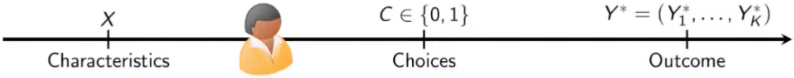
Model Setup & Identification Challenge

For each decision i:

- Characteristics: X_i (observed by DM and researcher)
- Binary choice: $C_i \in \{0, 1\}$
- Latent outcome: Y_i^* (unknown at time of choice)

Key Challenge: Missing Data

- Only observe outcome if $C_i = 1$: $Y_i = C_i \times Y_i^*$
- Observe $P(Y^*|C = 1, X)$ but not $P(Y^*|C = 0, X)$
- $P(Y^*|X)$ only partially identified



Expected Utility Framework

DM's decision problem:

- Observes (X, V) where V is private information
- Forms posterior beliefs about Y^*
- Makes choice C to maximize expected utility

Expected Utility Framework

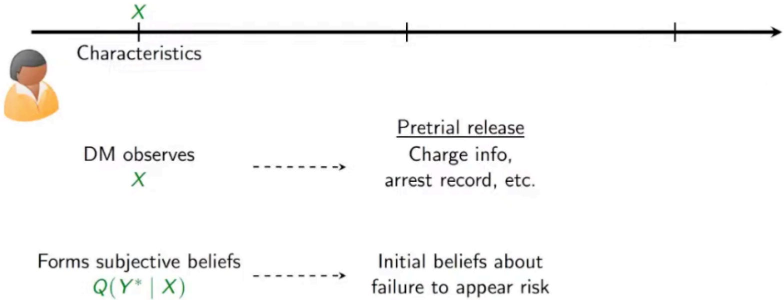
DM's decision problem:

- Observes (X, V) where V is private information
- Forms posterior beliefs about Y^*
- Makes choice C to maximize expected utility

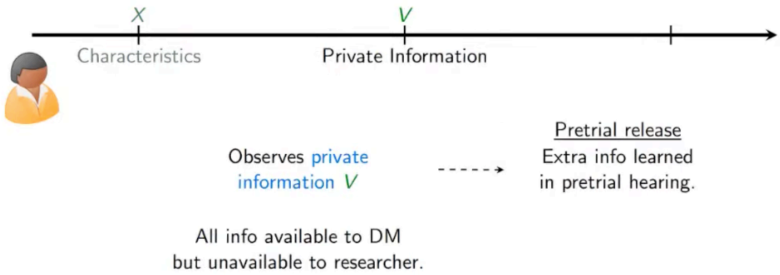
Three unknown components:

1. Utility function $u(c, y^*; x_0)$
2. Subjective beliefs about $Y^*|X$
3. Distribution of private information V

Information and Beliefs

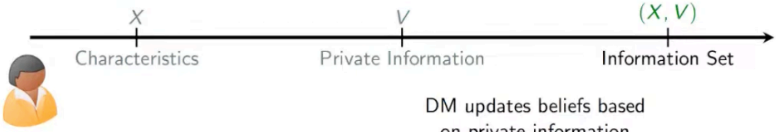


Information and Beliefs



- No assumptions are placed on the distribution of V and we will model non-parametrically

Information and Beliefs

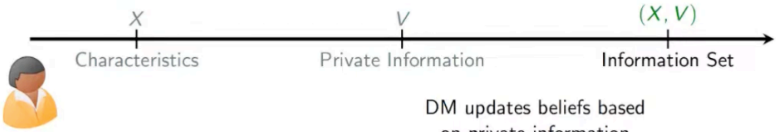


DM updates beliefs based on private information

Applies Bayes rule

Evaluates choices under posterior beliefs $Q(Y^* \mid V, X)$.

Information and Beliefs



DM updates beliefs based
on private information

Applies Bayes rule

Evaluates choices
under posterior beliefs
$Q(Y^* \mid V, X)$.

$$\underbrace{Q(Y^*|V,X)}_{\text{Posterior}} \propto \underbrace{Q(V|Y^*,X)}_{\text{Likelihood}} \; \underbrace{Q(Y^*|X)}_{\text{Subjective Beliefs}}$$

Goal: Infer DM's subjective beliefs $Q(Y^*|X)$ using observed choices $(X, C, Y) \sim \mathcal{P}$ without specifying private information structure

Utility Function & Exclusion Restriction

Suppose we are able to partition characteristics: $X = (X_0, X_1)$

- $u(c, y^*; x_0)$: Payoff depends only on X_0
- Characteristics X_0 directly affect utility function and beliefs, whereas other characterises X_1 and private information V only affect beliefs.

Utility Function & Exclusion Restriction

Suppose we are able to partition characteristics: $X = (X_0, X_1)$

- $u(c, y^*; x_0)$: Payoff depends only on X_0
- Characteristics X_0 directly affect utility function and beliefs, whereas other characterises X_1 and private information V only affect beliefs.

Example (Pretrial Release):

- X_0: Race, age, charge severity
- X_1: Prior record, current charges, etc.

Utility Function & Exclusion Restriction

Suppose we are able to partition characteristics: $X = (X_0, X_1)$

- $u(c, y^*; x_0)$: Payoff depends only on X_0
- Characteristics X_0 directly affect utility function and beliefs, whereas other characterises X_1 and private information V only affect beliefs.

Example (Pretrial Release):

- X_0: Race, age, charge severity
- X_1: Prior record, current charges, etc.

$\implies$ Conditional on X_0, variation in DM's choice probabilities across X_1 only reflect variation in posterior beliefs $Q(Y^* \mid V, X)$ but not variation in utility function $u(c, y^*; x_0)$

Utility Maximisation

- Given utility function, expected utility maximisation yields set of feasible joint distributions that must satisfy the following two outcomes:

$$(X, V, C, Y^*) \sim Q \text{ where } \begin{cases} (i) C \in \arg\max_{c'} \mathbb{E}Q\left[u(c', Y^*; X_0) | X, V\right], \\ (ii) C \perp Y^* \mid X, V. \end{cases}$$

Utility Maximisation

- Given utility function, expected utility maximisation yields set of feasible joint distributions that must satisfy the following two outcomes:

$$(X, V, C, Y^*) \sim Q \text{ where } \begin{cases} (i) C \in \arg\max_{c'} \mathbb{E}Q\left[u(c', Y^*; X_0) | X, V\right], \\ (ii) C \perp Y^* \mid X, V. \end{cases}$$

- We can say that a DM's observed choices are ***consistent with expected utility maximisation*** , at *accurate beliefs*, if there exists a utility function and associated joint distribution Q under expected utility model s.t.:

$$\underbrace{Q(X, C, Y)}_{\text{model}} = \underbrace{P(X, C, Y)}_{\text{data}}, \text{ where } Y = C \cdot Y^*$$

Linear Utility = Threshold Rule

Under linear utility

$$u(c, y^*; x_0) = u_1(x_0)y^*c + u_0(x_0)(1 - y^*)(1 - c)$$

1. DM chooses $C = 1$ when:

$$E[u_1(x_0)Y^*|X, V] \geq E[u_0(x_0)(1 - Y^*)|X, V]$$

2. Simplifies to threshold rule:

$$E[Y^*|X, V] \geq \tau(x_0)$$

Linear Utility = Threshold Rule

Under linear utility

$$u(c, y^*; x_0) = u_1(x_0)y^*c + u_0(x_0)(1 - y^*)(1 - c)$$

1. DM chooses $C = 1$ when:

$$E[u_1(x_0)Y^*|X, V] \geq E[u_0(x_0)(1 - Y^*)|X, V]$$

2. Simplifies to threshold rule:

$$E[Y^*|X, V] \geq \tau(x_0)$$

If choices maximise EU at accurate beliefs:

- When $C = 1$: posterior belief above threshold
- When $C = 0$: posterior belief below threshold

Ordering of Posteriors

This implies ordering of posteriors:

$$\max_{\tilde{x}_1} \underbrace{E[Y^*|C=1, X=(x_0, \tilde{x}_1)]}_{\text{point identified}} \leq \min_{\tilde{x}_1} \underbrace{E[Y^*|C=0, X=(x_0, \tilde{x}_1)]}_{\text{not identified}}$$

Ordering of Posteriors

This implies ordering of posteriors:

$$\max_{\tilde{x}_1} \underbrace{E[Y^*|C=1, X=(x_0, \tilde{x}_1)]}_{\text{point identified}} \leq \min_{\tilde{x}_1} \underbrace{E[Y^*|C=0, X=(x_0, \tilde{x}_1)]}_{\text{not identified}}$$

Key result: Can test accurate beliefs even with missing data

- If inequality holds for **any possible value** of missing outcomes $\rightarrow$ DM could be rational
- If inequality **cannot hold** for any value $\rightarrow$ DM making systematic mistakes

Ordering of Posteriors

This implies ordering of posteriors:

$$\max_{\tilde{x}_1} \underbrace{E[Y^*|C=1, X=(x_0, \tilde{x}_1)]}_{\text{point identified}} \leq \min_{\tilde{x}_1} \underbrace{E[Y^*|C=0, X=(x_0, \tilde{x}_1)]}_{\text{not identified}}$$

Key result: Can test accurate beliefs even with missing data

- If inequality holds for **any possible value** of missing outcomes $\rightarrow$ DM could be rational
- If inequality **cannot hold** for any value $\rightarrow$ DM making systematic mistakes

Why this works:

- Don't need actual outcomes for detained defendants
- We only need to check if we could reproduce the decision makers choices under **any** threshold rule on these conditional expectations

Additional Assumptions Needed

With the assumptions so far, the DM's choices are **always consistent** w/ EU max. at *some* linear utility function and accurate beliefs.

1. Econometric Assumptions for Missing Data:

$$\underline{\mathbb{E}}\left[Y^*|C=0,X\right] \leq \mathbb{E}\left[Y^*|C=0,X\right] \leq \overline{\mathbb{E}}\left[Y^*|C=0,X\right]$$

Additional Assumptions Needed

With the assumptions so far, the DM's choices are **always consistent** w/ EU max. at *some* linear utility function and accurate beliefs.

1. Econometric Assumptions for Missing Data:

$$\underline{\mathbb{E}}\left[Y^*|C=0,X\right] \leq \mathbb{E}\left[Y^*|C=0,X\right] \leq \overline{\mathbb{E}}\left[Y^*|C=0,X\right]$$

2. Behavioral Assumptions:

- Utility exclusion restrictions on X_1
- Private information doesn't directly affect utility

Constructing Bounds: IV Approach

Use quasi-random assignment of judges:

- $Z =$ judge leniency (leave-one-out release rate)
- Key assumption: $Y^* \perp\!\!\!\perp Z | X$

Constructing Bounds: IV Approach

Use quasi-random assignment of judges:

- $Z =$ judge leniency (leave-one-out release rate)
- Key assumption: $Y^* \perp\!\!\!\perp Z | X$

Intuition for bounds:

- Lenient judges release more defendants
- Marginal defendants reveal information
  - Lower bound: Risk of detained defendants must be at least as high as risk of released defendants
  - Upper bound: Risk of detained defendants can't be higher than risk of marginal defendants released by most lenient judges

Implementation of IV Bounds

For any characteristics x and instrument z, define:

- $\pi_1(x, z)$: Probability of release (C=1) given X=x and instrument value z
- $\pi_0(x, z)$: Probability of detention (C=0)
- $\mu_1(x, z)$: Observed failure rate among released defendants
- $\mu_0(x, z)$: Unobserved failure rate among detained defendants

Under standard assumptions on the instrument, we can bound $\mu_0(x, z)$:

$$\underline{\mu}_0(x) = \max\left\{\frac{\underline{\mu}(x) - \mu_1(x, z)\pi_1(x, z)}{\pi_0(x, z)}, 0\right\}$$

$$\overline{\mu}_0(x) = \min\left\{\frac{\bar{\mu}(x) - \mu_1(x, z)\pi_1(x, z)}{\pi_0(x, z)}, 1\right\}$$

where:

- $\underline{\mu}(x) = \max_{\tilde{z}}\{\mu_1(x, \tilde{z})\pi_1(x, \tilde{z})\}$
- $\bar{\mu}(x) = \min_{\tilde{z}}\{K\pi_0(x, \tilde{z}) + \mu_1(x, \tilde{z})\pi_1(x, \tilde{z})\}$

NYC Pretrial Application

Data:

- 570k cases (2008-2013)
- 265 judges (focus on top 25 by volume)
- Outcome: Failure to appear in court

NYC Pretrial Application

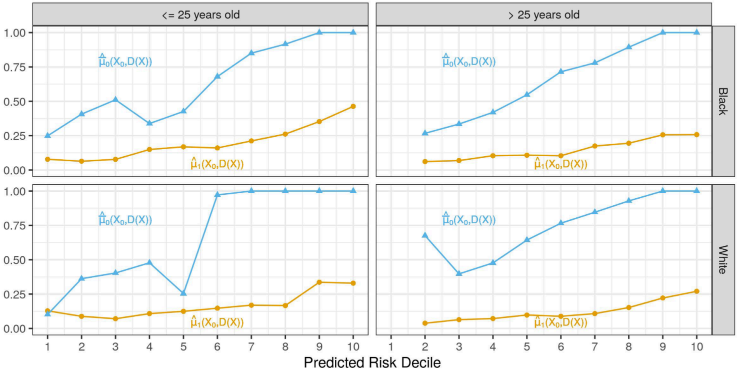
Data:

- 570k cases (2008-2013)
- 265 judges (focus on top 25 by volume)
- Outcome: Failure to appear in court

Testing approach:

1. Construct bounds using judge IV
2. Test for “misrankings” in decisions
3. Compare highest-risk releases vs. lowest-risk detentions

Evidence of Prediction Mistakes



Clear evidence of misrankings:

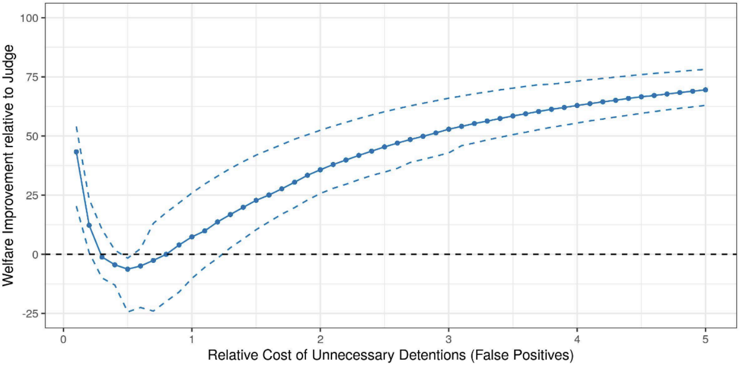
- Orange: Observed failure rates of released defendants
- Blue: Upper bounds for detained defendants
- When orange > blue: Must be prediction mistakes

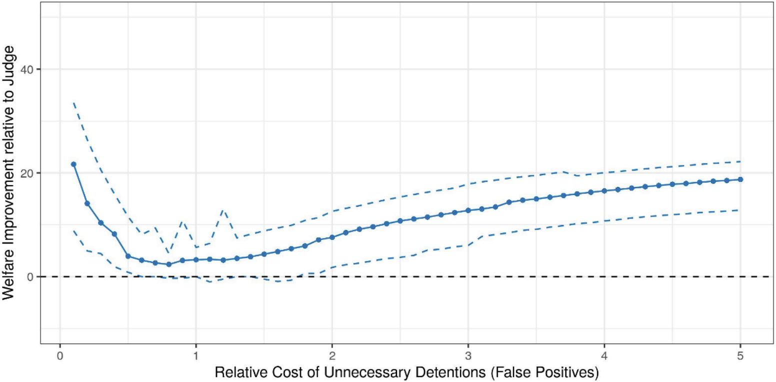
Key Results

Utility Functions allowed to vary by:	Rejection Rate
No Characteristics	24%
Race	24%
Race + Age	**20%**
Race + Charge	32%

At least 20% of judges make systematic prediction mistakes

Policy Implications





Two key findings:

1. Full automation has ambiguous effects
2. Targeted automation of tail decisions dominates status quo

Conclusions

Key Contributions:

1. Framework for testing prediction mistakes under weak assumptions
2. Shows substantial mistakes in important setting
3. Informs algorithm adoption decisions

Conclusions

Key Contributions:

1. Framework for testing prediction mistakes under weak assumptions
2. Shows substantial mistakes in important setting
3. Informs algorithm adoption decisions

Open Question:

- Learning dynamics