Prediction Policy

Susan Athey-Machine Learning and Causal Inference

Thanks to Sendhil Mullainathan for sharing his slides

Three Direct Uses of Prediction

1. Policy
2. Testing Whether Theories are Right
3. Testing Theory Completeness

When is Prediction Primary Focus?

- Economics: “allocation of scarce resources”
- An allocation is a decision.
  - Generally, optimizing decisions requires knowing the counterfactual payoffs from alternative decisions.
- Hence: intense focus on causal inference in applied economics

- Examples where prediction plays the dominant role in a decision
  - Decision is obvious given an unknown state
  - Many decisions hinge on a prediction of a future state

Prediction and Decision-Making: Predicting a State Variable

Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015)

- Motivating examples:
  - Will it rain? (Should I take an umbrella?)
  - Which teacher is best? (Hiring, promotion)
  - Unemployment spell length? (Savings)
  - Risk of violation of regulation (Health inspections)
  - Riskiest youth (Targeting interventions)
  - Creditworthiness (Granting loans)
- Empirical applications:
  - Will defendant show up for court? (Should we grant bail?)
  - Will patient die within the year? (Should we replace joints?)

Allocation of Inspections

- Examples:
  - Auditors
  - Health inspectors
  - Fire code inspectors
  - Equipment
- Efficient use of resources:
  - Inspect highest-risk units
  - (Assuming you can remedy problem at equal cost for all…)

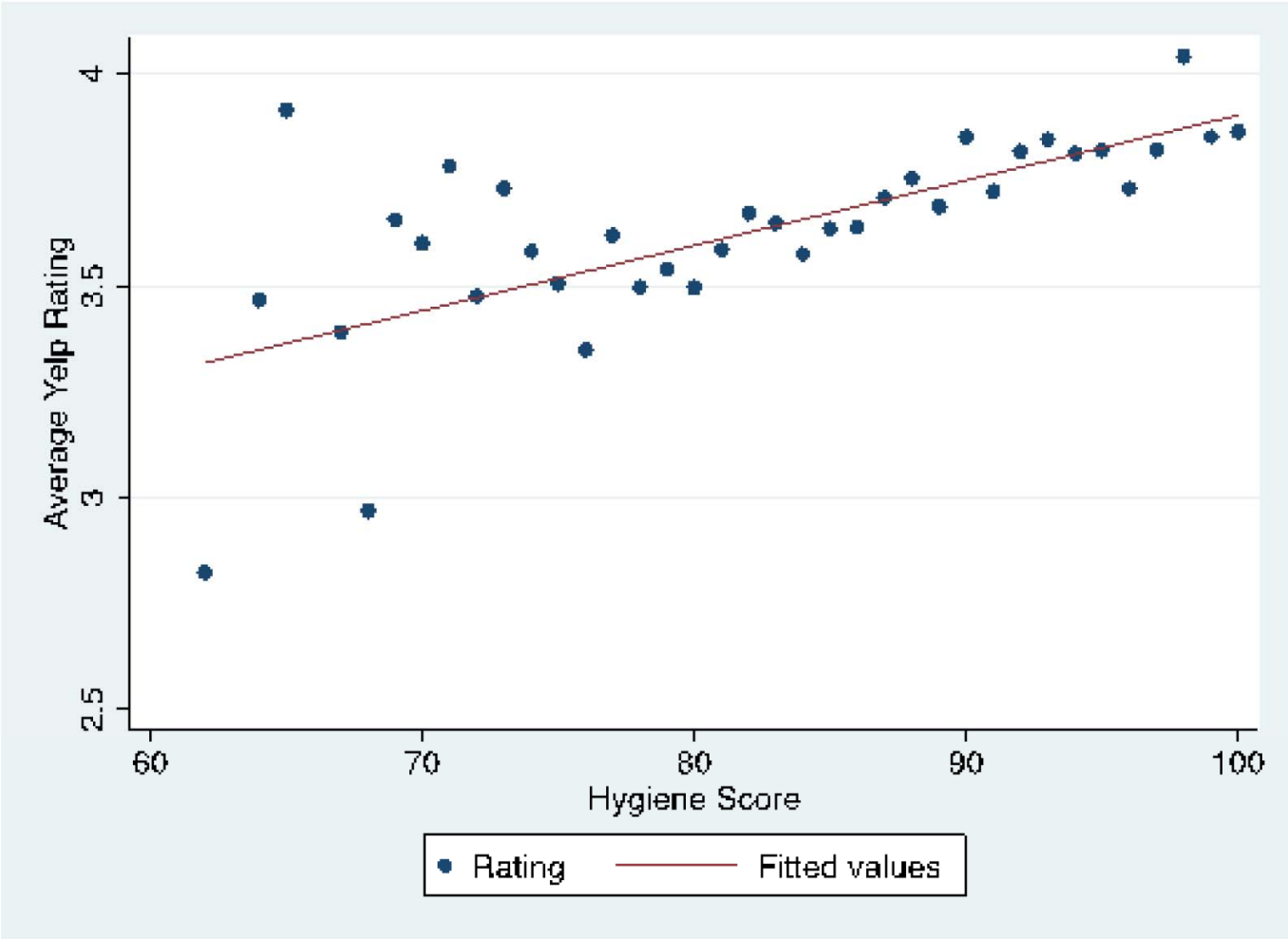
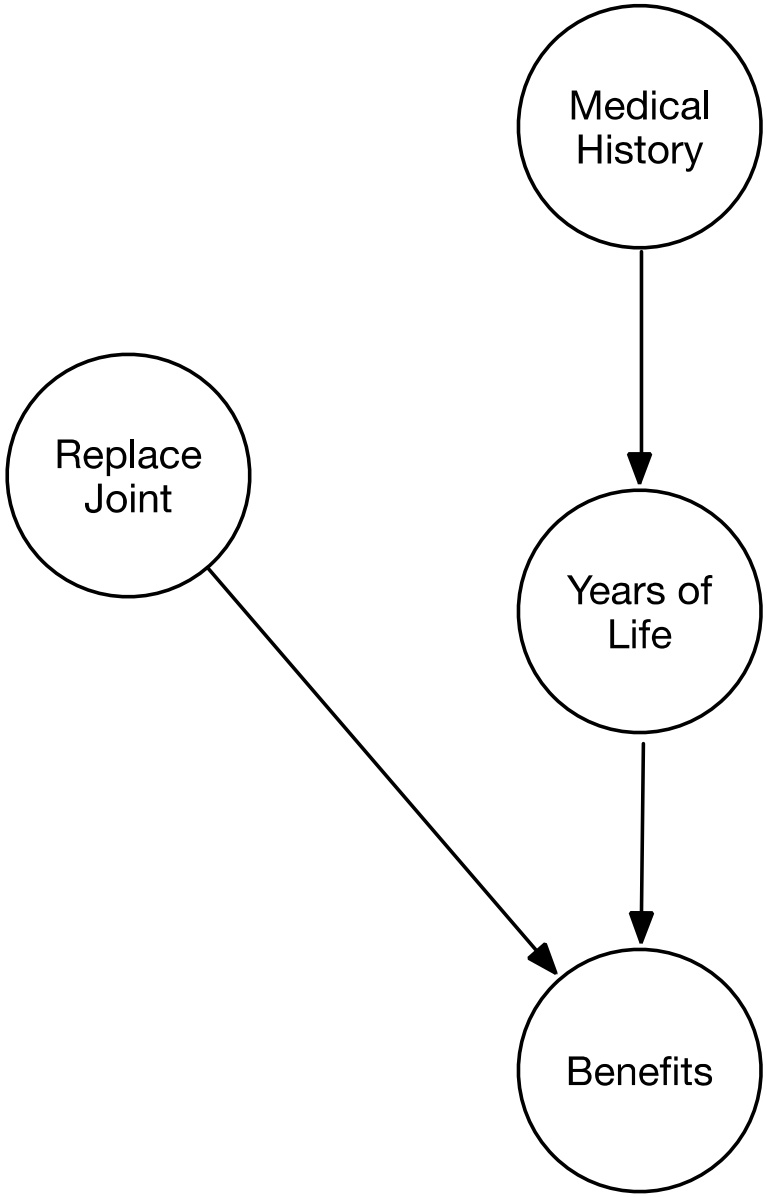


Figure 3: *Correlation between Yelp Ratings and Hygiene Inspection Scores*

Review data consists of reviews on Yelp.com for restaurants in San Francisco, CA from September, 2010 through September, 2013. Hygiene scores consist are from the San Francisco Department of Public Health, for the same timeframe.

Prediction Problem

- Over 750,000 joint replacements every year
- Benefits
  - Improved mobility and reduced pain
- Costs
  - Monetary: $15,000 (roughly)
  - Non-monetary: short-run utility costs as people recover from surgery



Look at death rate in a year
How well are we doing avoiding unnecessary surgery?

Medicare claims data 2010 surgeries for joint replacement

Average death rate is 5%

But is that the right metric for excess joint replacements?

Don’t want average patient

Want marginal patient

- Predictably highest risk patients

Table 2: The Riskiest People Receiving Joint Replacements

Predicted Mortality Percentile	Observed Mortality Rate	Total Number Annually
1	0.562	4905
	(.027)	
2	0.530	9810
	(.02)	
5	0.456	24525
	(.012)	
10	0.345	49045
	(.008)	
20	0.228	98090
	(.005)	
30	0.165	147135
	(.004)	
100	0.057	490450
	(.001)	

Approach: use ML methods to predict mortality as a function of covariates

- e.g. regularized regression, random forest
- Put individuals into percentiles of mortality risk

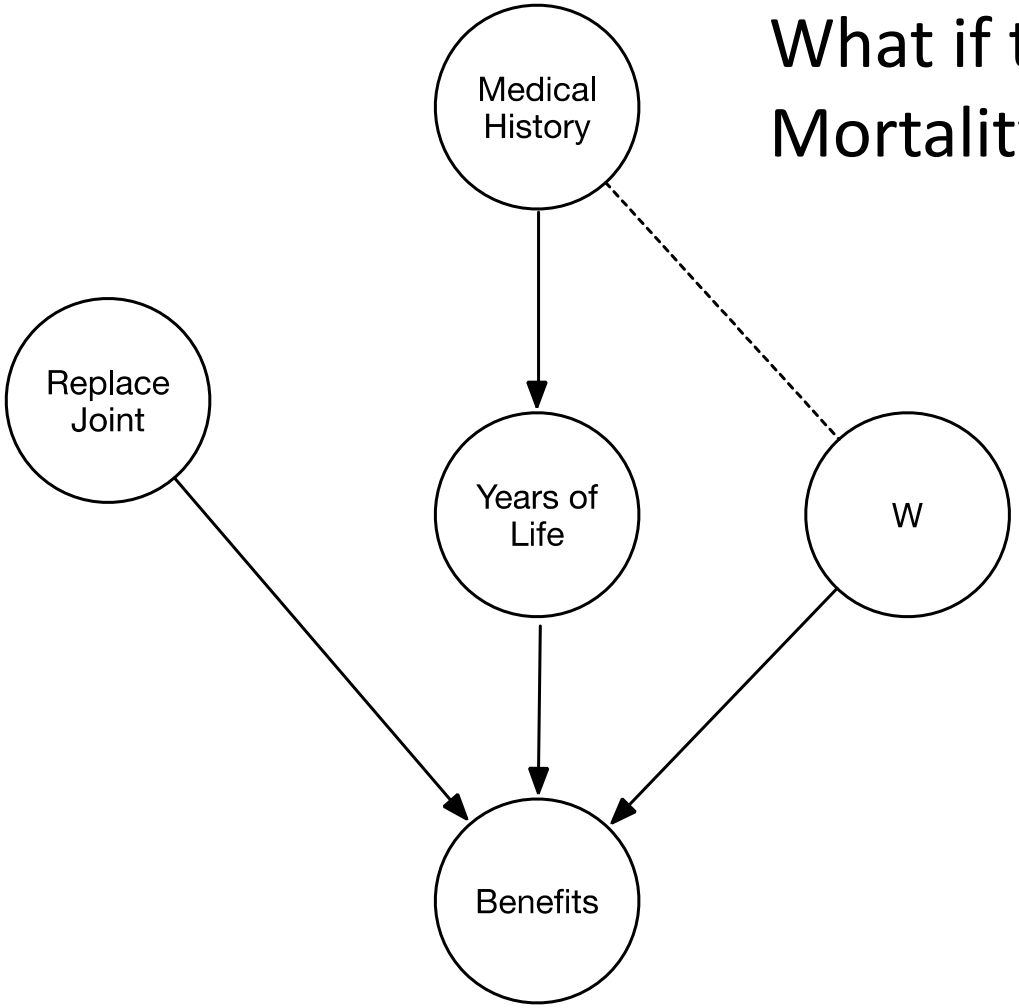
A large number of joint replacements going to people who die within the year

Could we just eliminate the ones above a certain risk?

Econometrics of Prediction Policy Problems

1. Problem: Omitted Payoff Bias

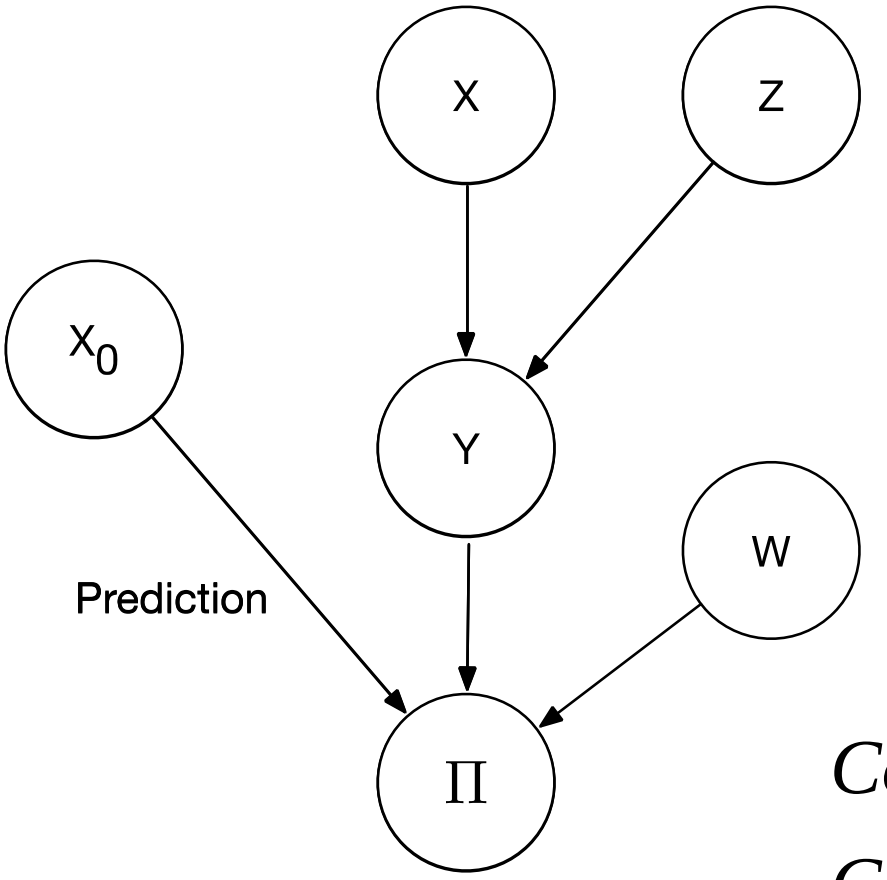
This Unobservable is a Problem



What if those with high Mortality also benefit most?

Pain

Omitted Payoff Bias



$$Y = f(X, Z)$$

$$\Pi = g(\boldsymbol{X_0}, W)$$

$Cov(X, Z)$ is not a problem

$Cov(X, W)$ is a problem

Econometrics of Prediction Policy Problems

1. Omitted Payoff Bias
   - Like omitted variable bias but not in *y*
   - Can partially assess on the basis of observables

Table 2: The Riskiest People Receiving Joint Replacements

Predicted Mortality Percentile	Observed Mortality Rate	Total Number Annually	PT + Joint Injections	Physician Visits for Osteo.
1	0.562	4905	4.4	1.4
	(.027)		(.356)	(.173)
2	0.530	9810	4.0	1.8
	(.02)		(.316)	(.13)
5	0.456	24525	3.9	2.0
	(.012)		(.208)	(.092)
10	0.345	49045	3.8	2.1
	(.008)		(.143)	(.066)
20	0.228	98090	3.9	1.8
	(.005)		(.091)	(.042)
30	0.165	147135	3.8	1.9
	(.004)		(.076)	(.035)
100	0.057	490450	3.9	2.1
	(.001)		(.046)	(.023)

No sign of bias:
Highest risk show no signs of greater benefit

Quantifying gain of predicting better

- Allocation problem:
  - Reallocate joints to other eligible patients
- How to estimate the risk of those who didn't get surgery?
  - Look at those who could get surgery but didn't
  - Doctors should choose the least risky first
  - So those who don't receive should be particularly risky.
- Take a conservative approach
  - Compare to median risk in this pool

Table 2: The Riskiest People Receiving Joint Replacements

Predicted Mortality Percentile	Observed Mortality Rate	Total Number Annually	Substitute with 50th percentile Eligibles	
			Futile Procedures Averted	Annual Savings (in millions)
1	0.562 (.027)	4905	2403	36
2	0.530 (.02)	9810	4485	67
5	0.456 (.012)	24525	9398	141
10	0.345 (.008)	49045	13350	200
20	0.228 (.005)	98090	15219	228
30	0.165 (.004)	147135	13548	203
100	0.057 (.001)	490450		

Assessing the Research Agenda

- Follows economic tradition of using data to improve policy
- In an area of economic interest
  - Similar to a lot of health econ work
- Of course this does not answer all questions of interest
  - Why not?

Another Prediction Policy Problem

- Each year police make over 12 million arrests
- Many detained in jail before trial
- Release vs. detain high stakes
  - Pre-trial detention spells avg. 2-3 months (can be up to 9-12 months)
  - Nearly 750,000 people in jails in US
  - Consequential for jobs, families as well as crime

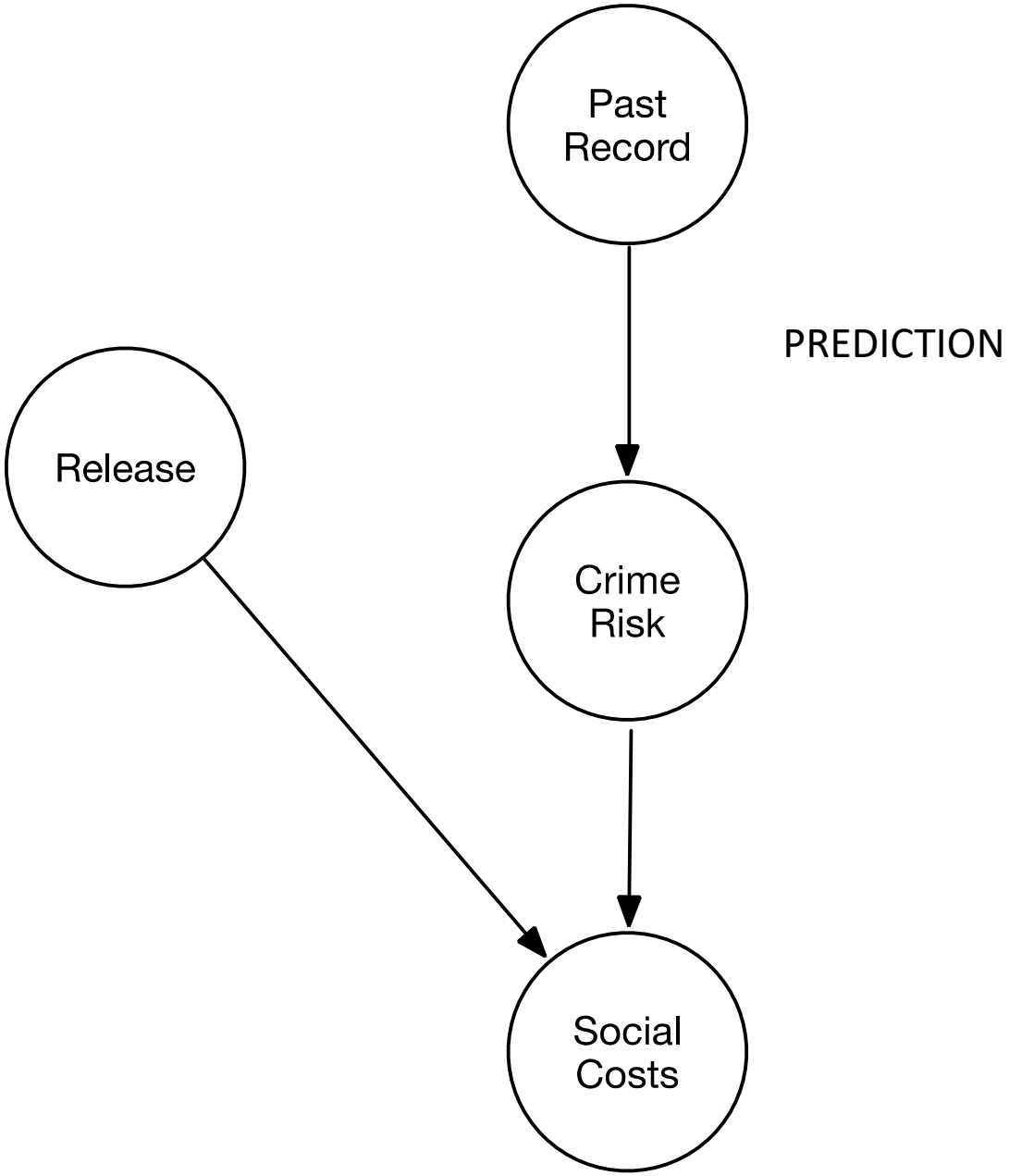
Kleinberg Lakkaraju Leskovec Ludwig and Mullainathan

Judge's Problem

- Judge must decide whether to release or not (bail)

- Defendant when out on bail can behave badly:
  - Fail to appear at case
  - Commit a crime

- The judge is making a *prediction*


Past
Record
PREDICTION
Release
Crime
Risk
Social
Costs

Omitted Payoff Bias?

- Bail carefully chosen
  - Unlike other sentencing no other concerns:
    - Retributive justice
  - Family & other considerations low

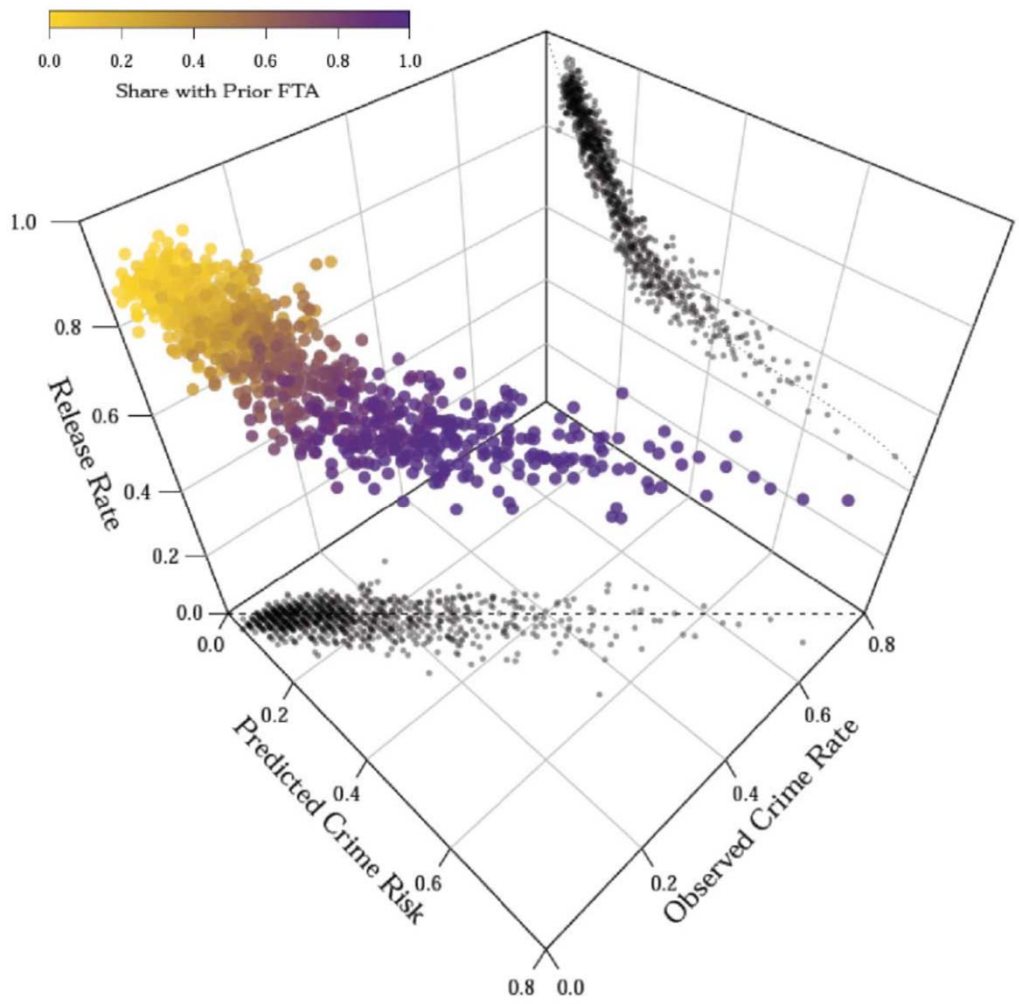
- Bad use case: Parole Decision

Evaluating the Prediction

- NOT just AUC or Loss
- Use predictions to create a release rule
- What is the release – crime rate tradeoff?
- Note: There's a problem

Econometrics of Prediction Policy Problems

1. Omitted Payoff Bias
2. "Selective Labels"
   - What do we do with people algorithm releases that judge jails?
   - (Like people who get surgery and didn't before)



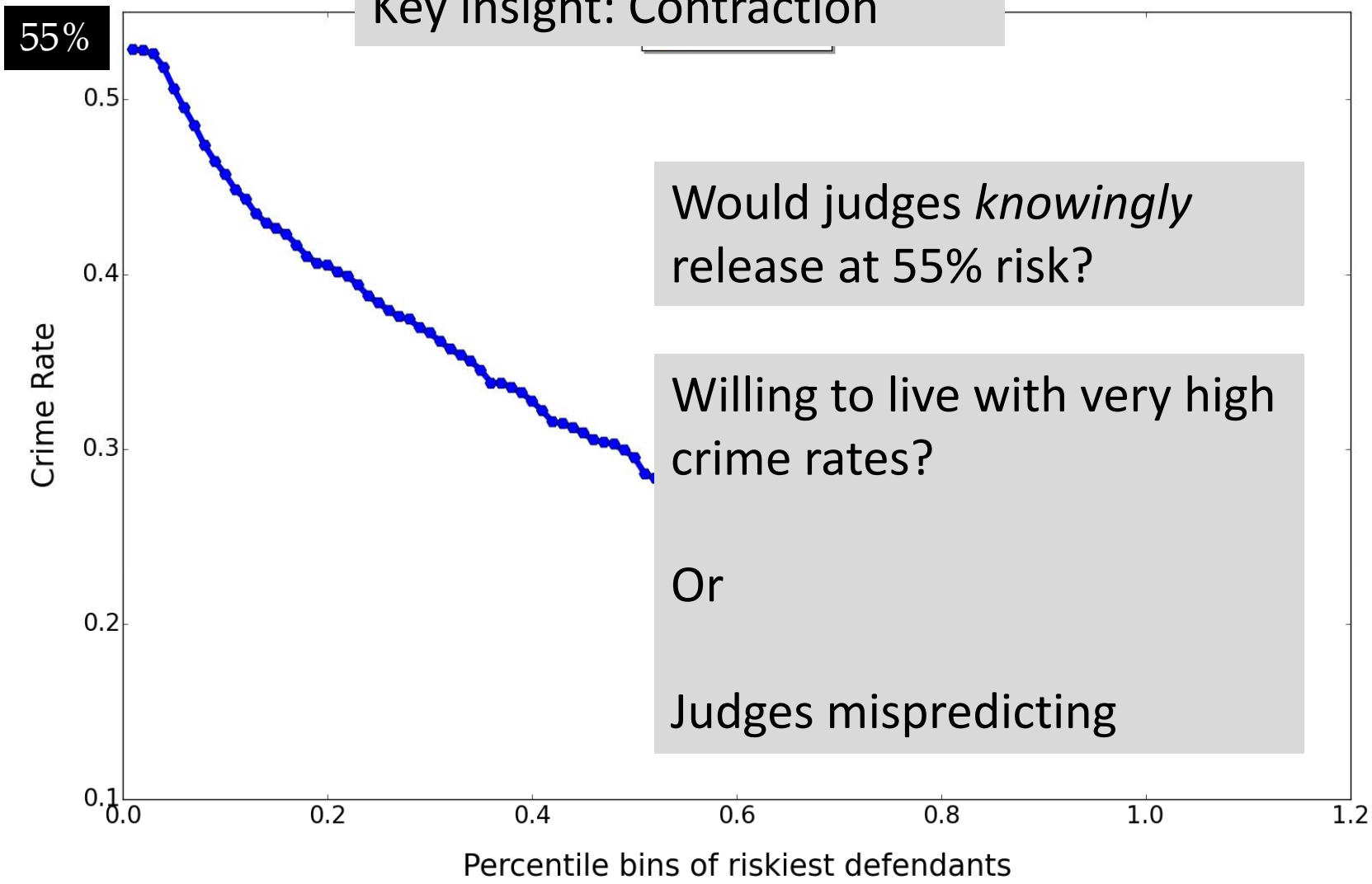
Notes: The figure above shows the results of an algorithm built using 221,876 observations in our NYC training set, applied to the 110,938 observations in our test set (see Figure 1). We report the observed judge's release rate (y-axis) against both the algorithm's predicted crime risk for each observation and the observed crime rate (observed only for those defendants in the test set released by the judges) for 1,000 bins sorted by predicted risk. The coloring shows share observations in each bin with a prior failure to appear. The bottom and back panels show the projection of the figure onto the two dimensional {predicted crime risk, observed crime rate} space and the {predicted crime risk,judge release rate} space.

Selective Labels Revisted

- What is the crime rate we must use?
  - For released defendants, empirical crime rate
  - For jailed ones, imputed crime rate
- But imputation may be biased…
  - Judge sees factors we don't
  - Suppose young people have dots on their foreheads
    - Perfectly predictive: judge releases only if no dot
  - In released sample: young people have no crime
    - We would falsely conclude young people have no risk.
    - But this is because the young people with dots are in jail.
  - We would then falsely presume when we release all young people we will do better than judge
- Key problem: unobserved factors seen by judge affect crime rate **(& judge uses these wisely)**
- How to fix?

Is not problem when we look just at released

Key insight: Contraction



Would judges *knowingly* release at 55% risk?

Willing to live with very high crime rates?

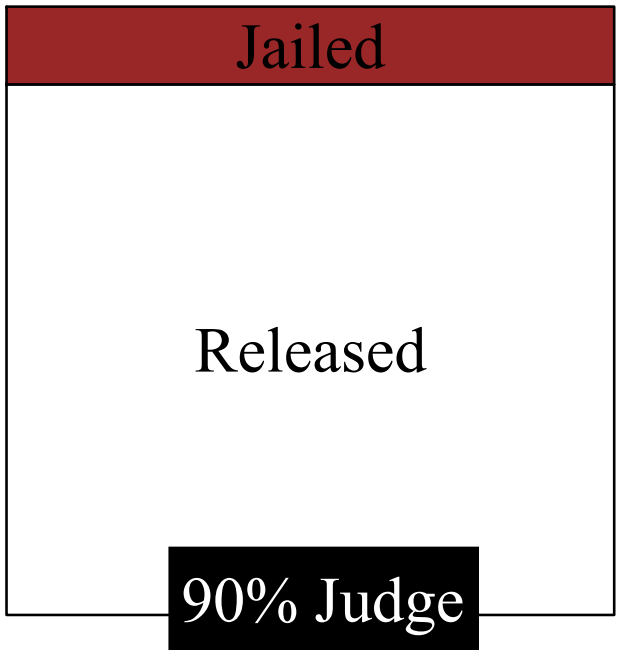
Or

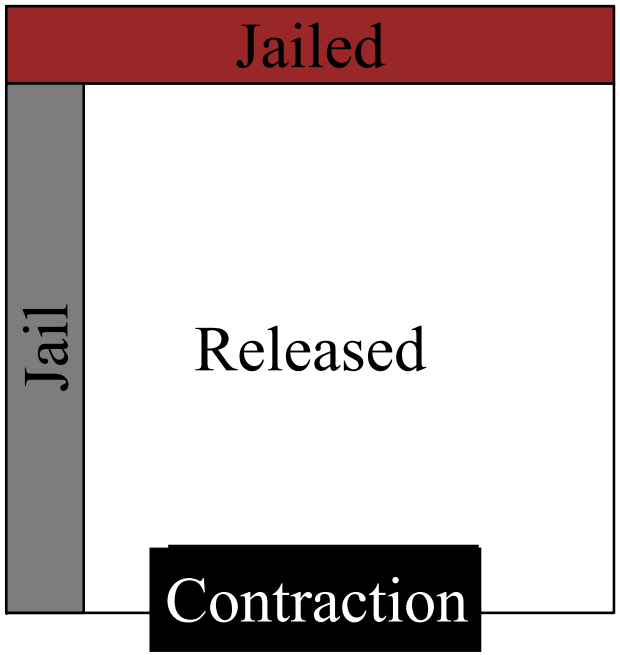
Judges mispredicting

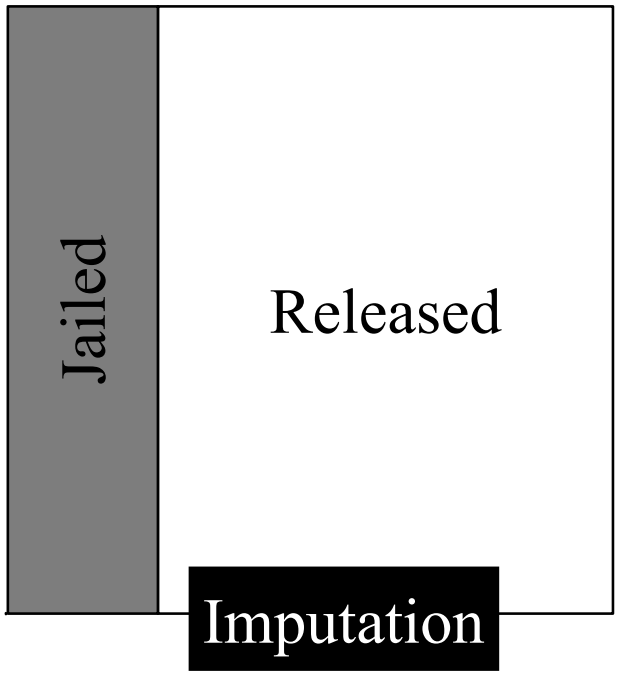
Contraction

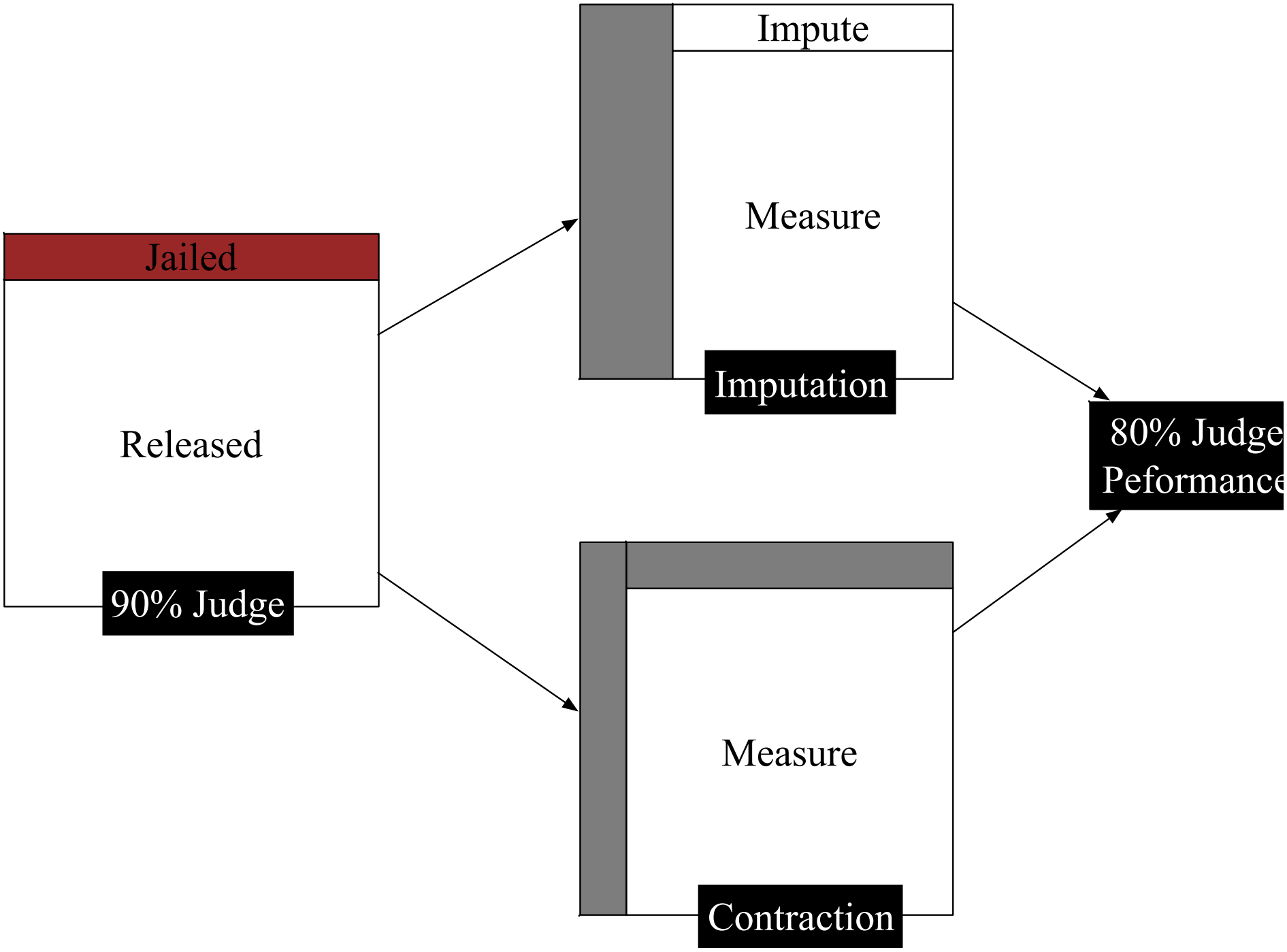
- Multiple judges with similar caseloads and different lenience

- Strategy: use most lenient judges.
  - Take their released population and ask which ***of those*** would you incarcerate to become less lenient
  - Compare to less lenient judges


Jailed
Released
90% Judge


Jailed
Jail
Released
Contraction


Jailed
Released
Imputation

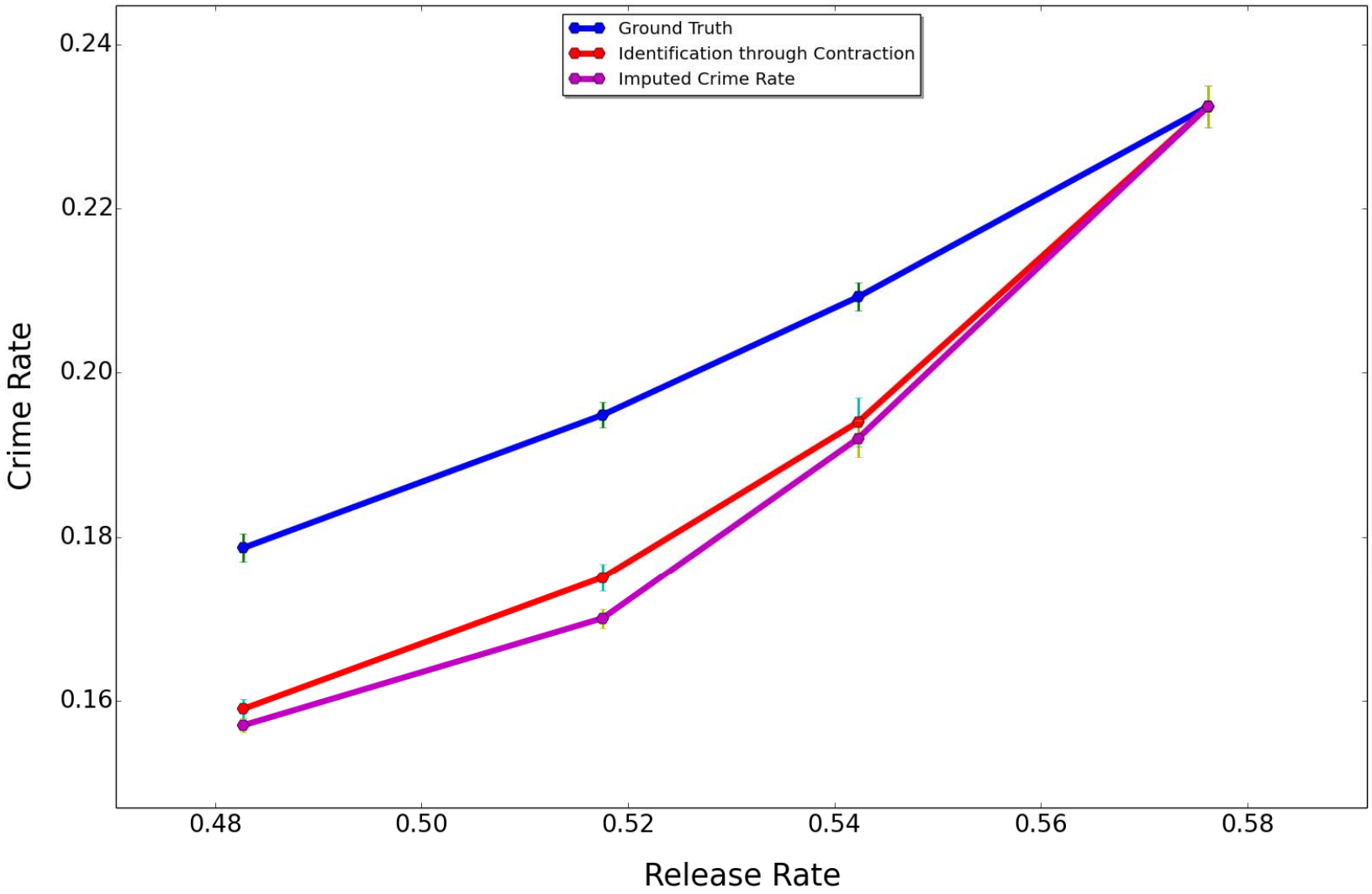

Impute
Measure
Jailed
Released
Imputation
80% Judge
Peformance
90% Judge
Measure
Contraction

Contraction

- Requires
  - Judges have similar cases (random assignment)
  - Does not require judges having “similar” rankings

- But does give performance of a different rule
  - “Human constrained” release rule

Contraction and Imputation Compared



Selective Labels

- In this case does not appear to be a problem

- But generically a problem
  - Extremely common problem – occurs whenever prediction -> decision -> treatment
  - Data generated by previous decisions

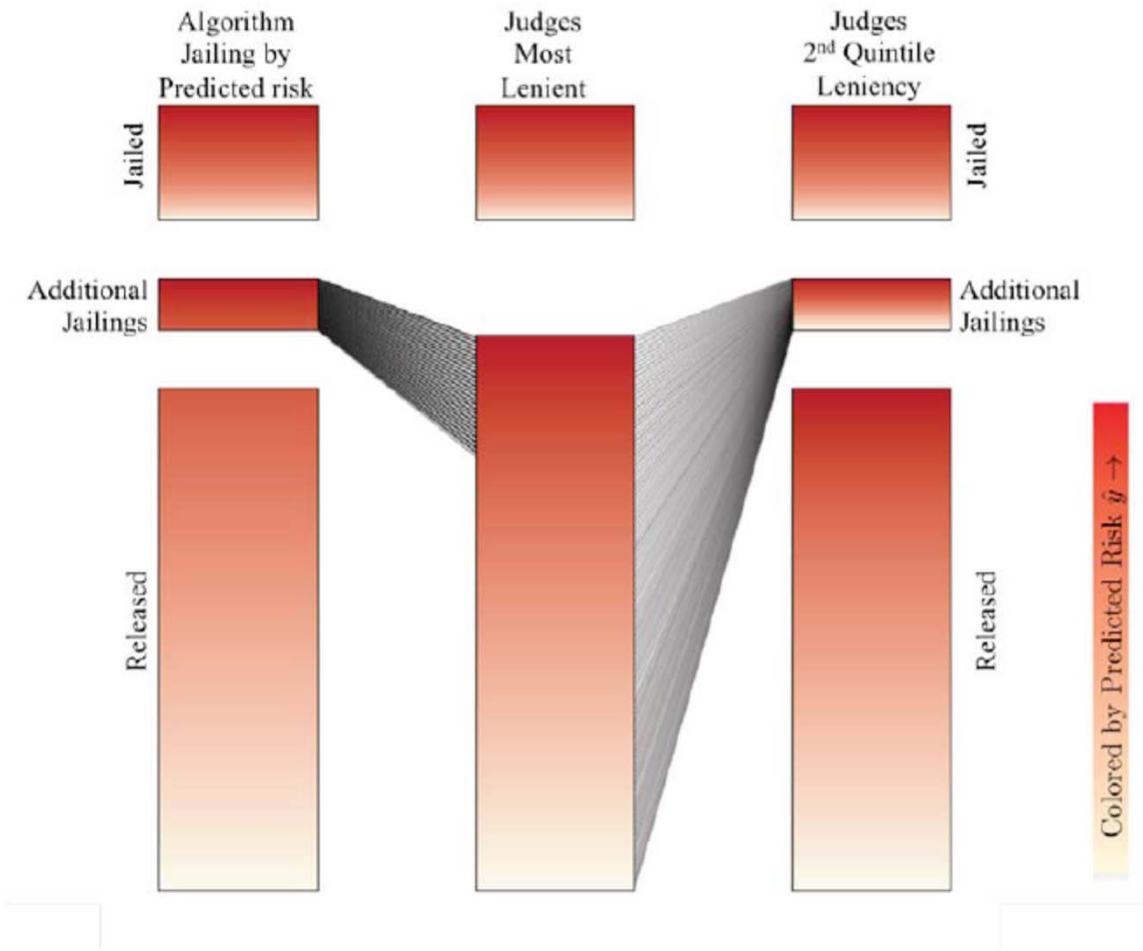


Figure 3: Who is Jailed as Judges Become More Stringent?

Notes: The bar in the middle of the figure shows who the most lenient quintile judges jail (top) and release (bottom) in our NYC dataset. The color shading shows the algorithm's predicted crime risk. The bar at the right shows how the 2nd most lenient quintile judges implicitly select their marginal detainees to get from the most lenient quintile's release rate down to their own release rate. The arrows show where within the risk distribution of the lenient quintile's released set the stricter judges are selecting the marginal detainees. The bar at the left shows how the algorithm would select marginal detainees to achieve the same reduction in release rate.

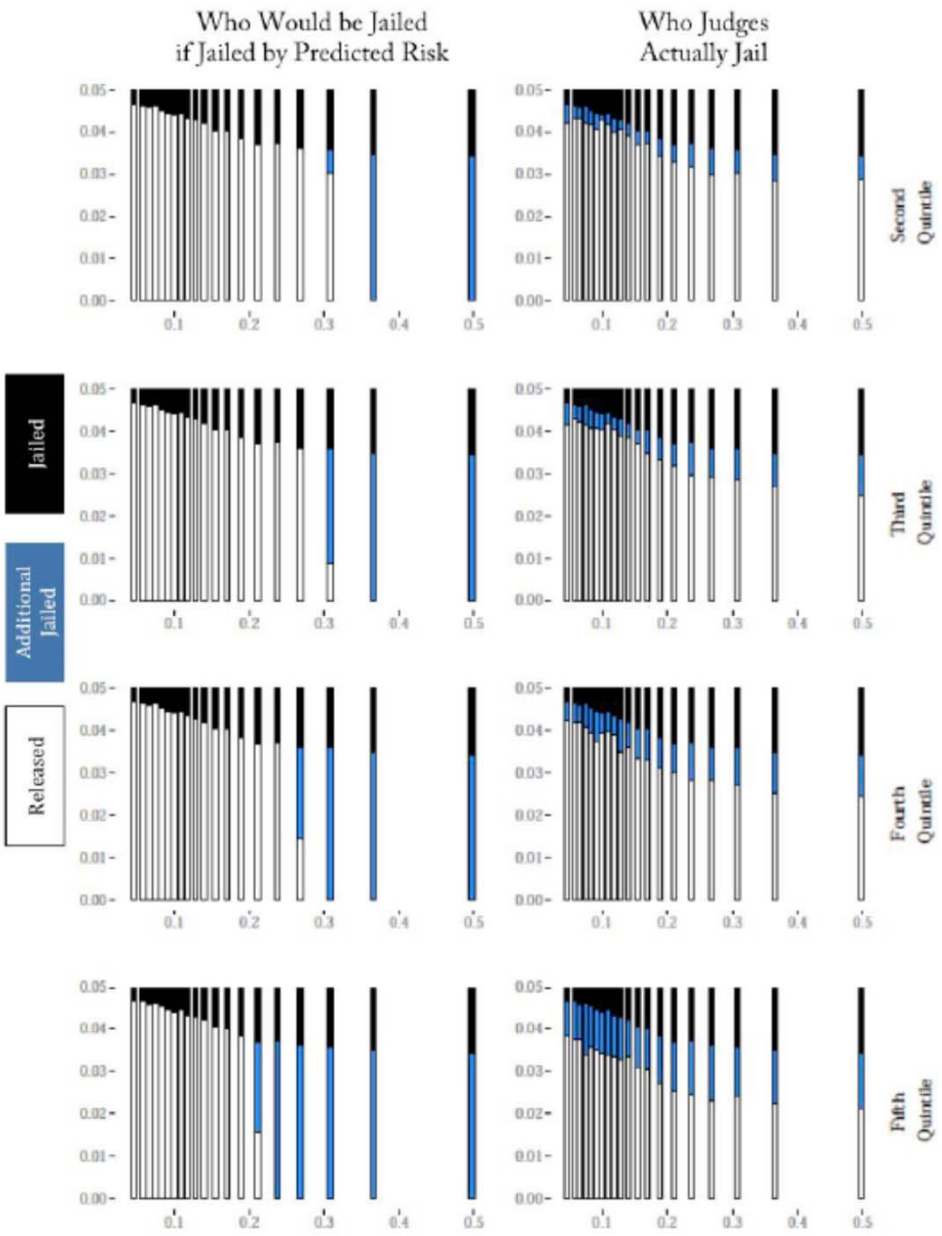


Figure 4: Who do Stricter Judges Jail? Predicted Risk of Marginal Defendants

Notes: This figure shows where each of the quintiles of stricter judges in NYC select their marginal defendants (relative to the most lenient quintile), compared to how the algorithm would select marginal detainees. Within each panel, we divide the sample up into 20 bins by predicted crime risk (shown on the x-axis). The black segment at the top of each bar shows the share of each bin the most lenient quintile judges jail. In the top right-hand panel, we show which defendants the second-most-lenient quintile judges implicitly select to jail to get from the most lenient judge's release rate down to their own lower release rate (blue), and who they continue to release (white). The left-hand top panel shows whom the algorithm would select instead. Each of the remaining rows shows the same comparison between the judge and algorithm decisions for the other less-lenient judge quintiles.

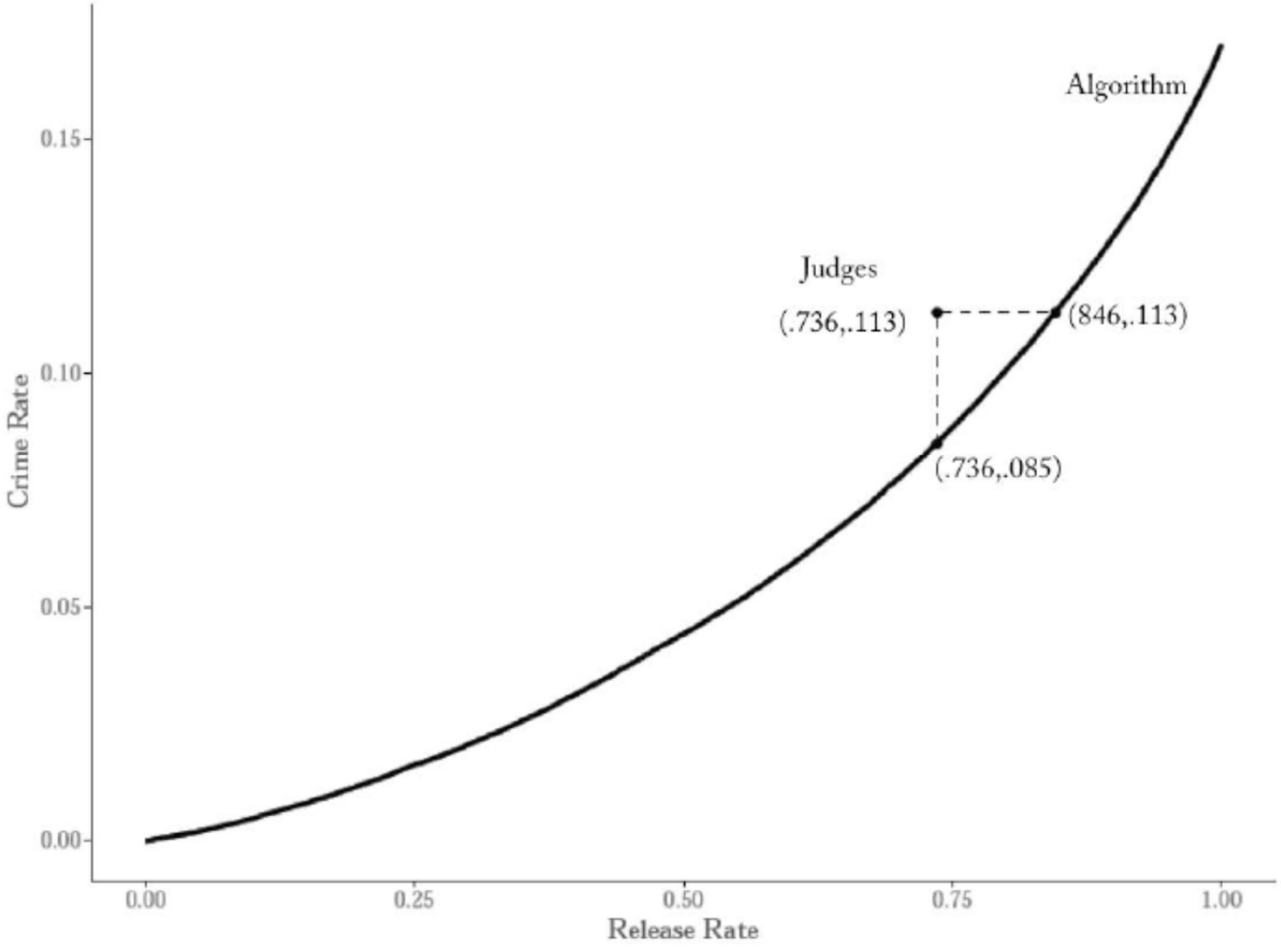


Figure 8: Simulation of Crime Rate - Release Tradeoff Algorithm Allows

Notes: The curve in the top panel shows the crime rate and release rate combinations that would be possible in NYC if judges were given a risk tool that could re-rank all defendants by their predicted crime risk and recommend them for detention in order of risk. Since we would like a crime rate that can be meaningfully compared across release rates, the y-axis shows the ratio of crimes committed by released defendants to the *total* number of defendants, not just the number released.The curve shows what gains would be possible relative to actual current judge decisions, assuming perfect compliance with the new tool. The curve in the bottom panel shows the risk level of the marginal person detained at each possible release rate under the algorithmic release rule.

Econometrics of Prediction Policy Problems

1. Omitted Payoff Bias
2. Selective Labels
3. Restricted Inputs

Restricted Inputs

- Race and gender are not legal to use
  - We do not use them
- But is that enough?
  - Reconstruction problem
  - Optimizing in presence of this additional reconstruction constraint
- Rethinking disparate impact and disparate treatment

Racist Algorithms?

Table 7: Racial Fairness

Release Rule	Crime Rate	Drop Relative to Judge	Percentage of Jail Population Black	Hispanic	Minority
Distribution of Defendants (Base Rate)			.4877	.3318	.8195
Judge	.1134 (.0010)	0%	.573 (.0029)	.3162 (.0027)	.8892 (.0018)
Algorithm					
Usual Ranking	.0854 (.0008)	-24.68%	.5984 (.0029)	.3023 (.0027)	.9007 (.0017)
Match Judge on Race	.0855 (.0008)	-24.64%	.573 (.0029)	.3162 (.0027)	.8892 (.0018)
Equal Release Rates for all Races	.0873 (.0008)	-23.02%	.4877 (.0029)	.3318 (.0028)	.8195 (.0023)
Match Lower of Base Rate or Judge	.0876 (.0008)	-22.74%	.4877 (.0029)	.3162 (.0027)	.8039 (.0023)

Notes: Table reports the potential gains of the algorithmic release rule relative to the judge at the judge s release rate with respect to crime reductions and share of the jail population that is black, Hispanic or either black or Hispanic. The first row shows the share of the defendant population overall that is black or Hispanic. The second row shows the results of the observed judge decisions. The third row shows the results of the usual algorithmic re-ranking release rule, which does not use race in predicting defendant risk and makes no post-prediction adjustments to account for race. In the fourth row we adjust the algorithm s ranking of defendants for detention to ensure that the share of the jail population that is black and Hispanic under the algorithmic release rule are no higher than those under current judge decisions. The next row constraints the algorithmic release rule s jail population to have no higher share black or Hispanic than that of the general defendant pool, while the final row constrains the algorithm s jail population to have no higher share black or Hispanic than either the judge decisions or the overall defendant pool.

Econometrics of Prediction Policy Problems

1. Omitted Payoff Bias
2. Selective Labels
3. Restricted Inputs
4. Response to Decision Rule

Comparing Judges to Themselves

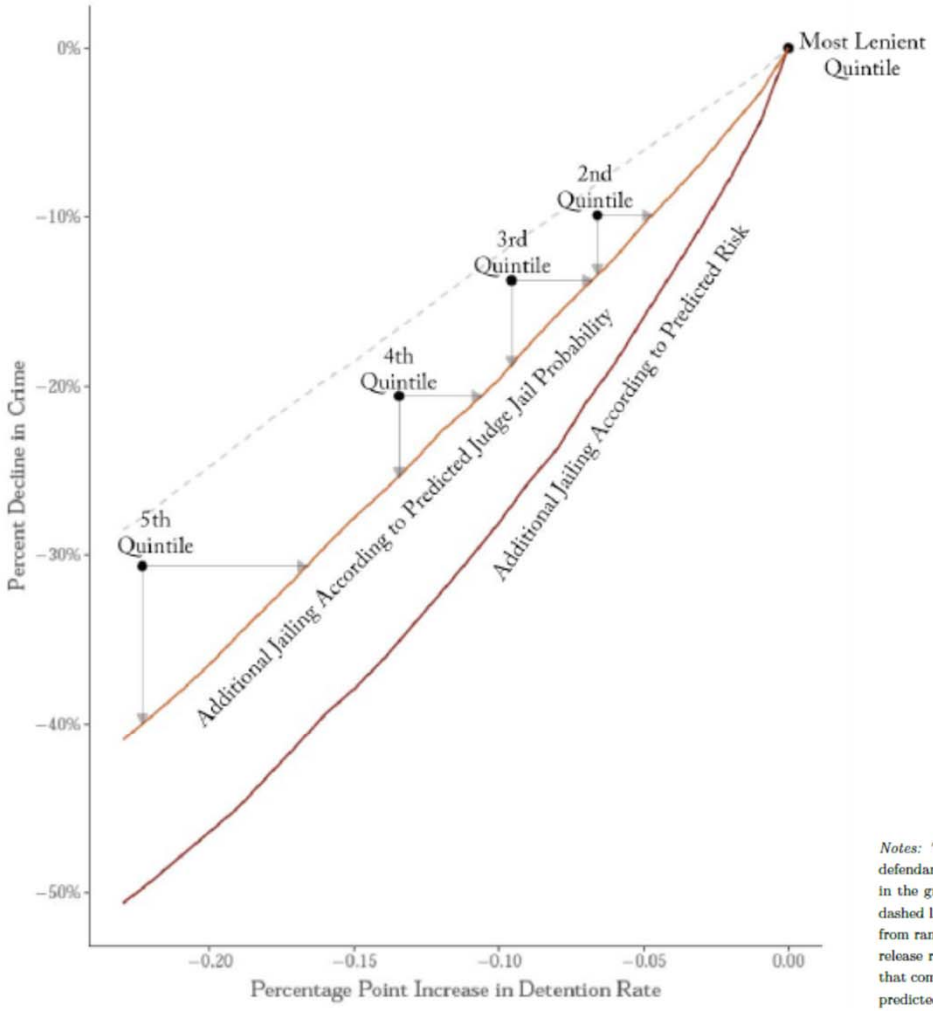


Figure 13: Effect of Detaining Defendants Judges Usually Detain

Notes: This figure compares the change in crime rates and release rates that could be achieved by jailing additional defendants using the algorithm's predicted crime risk compared to the decisions of stricter judges. The right-most point in the graph represents the release rate of the most lenient quintile of judges, with the crime rate that results. The light dashed line shows the decline in crime (as a percentage of the lenient quintile's crime rate, shown on the y-axis) that results from randomly selecting additional defendants to detain from within the lenient quintile's released cases, with the change in release rate relative to the lenient quintile shown on the x-axis. The red curve shows the crime rate / release rate tradeoff that comes from jailing additional defendants within the lenient quintile's released set in descending order of the algorithm's predicted crime risk. The additional curve on the graph shows the crime rate / release rate outcomes we would get from jailing additional defendants within the lenient quintile judges' caseloads in descending order of an algorithm's predicted probability that the judges jail a given defendant. The four points on the graph show the crime rate / release rate outcomes that are observed for the actual decisions made by the second through fifth most lenient quintile judges, who see similar caseloads on average to those of the most lenient quintile judges.

Why do we beat judges?

- Judges see more than we do
- Perhaps that is the problem

	Correctly used	Misused
Measured	Offence history	Time since break
Unmeasured	Private info	Demeanor

- Suggests behavioral economics of salience important here
  - In general, any kind of "noise"

General points here

- Need more ways of comparing human and machine predictions
- Notion of private information called into question

Summary

- Many prediction policy problems
- Raise their own econometric challenges
- Can also provide conceptual insights