Examiner Inconsistency: Evidence from Refugee Appeals

Samuel Norris[1]

[1]Harris School of Public Policy, University of Chicago

Introduction

Uncertainty is costly

- Firms: distorts investment decisions
- Individuals: skill decay while waiting for SSDI decision

One major source of uncertainty:

- Different decision-makers make different decisions
- Courts, SSDI, patents, health inspections, teachers

Introduction

Uncertainty is costly

- Firms: distorts investment decisions
- Individuals: skill decay while waiting for SSDI decision

One major source of uncertainty:

- Different decision-makers make different decisions
- Courts, SSDI, patents, health inspections, teachers

This paper: measure the level and causes of disagreement

- Goal: measure disagreement for all pairs of examiners
- Challenge: disagreement not directly observed

Why?

- Can use to improve institutional consistency
- Too much disagreement violates IV monotonicity
    - Introduce new, more powerful tests

Two approaches

Nonparametric bounds on disagreement

- Disagreement not directly observed, so can only be bounded
- Bounds can be sharpened using two-stage decisions
  - Appeals, second medical opinions, etc

Structural model generalizing index decision-making

- Judges observe case quality with error, approve if above standard
  - Judges differ in standards and ordering errors
- Parsimonious characterization of decision-making
- Allows study of judge-level determinants of inconsistency

Preview of results

Judge disagreement in Canadian refugee appeals

- Approval requires two consecutive judges
- Random assignment in both rounds

Disagreement rates are very high!

- Average pair of judges approves 14% of cases, disagrees on 17%
  - Info on 2nd round tightens disagreement bound from 7 to 17%

Preview of results

Judge disagreement in Canadian refugee appeals

- Approval requires two consecutive judges
- Random assignment in both rounds

Disagreement rates are very high!

- Average pair of judges approves 14% of cases, disagrees on 17%
  - Info on 2^{nd} round tightens disagreement bound from 7 to 17%

Ordering errors:

- decline with experience
- increase with workload
- are lower for judges appointed through non-partisan process

Optimal allocation

- Could lower costs by 18% by optimally assigning judges to rounds

Setup

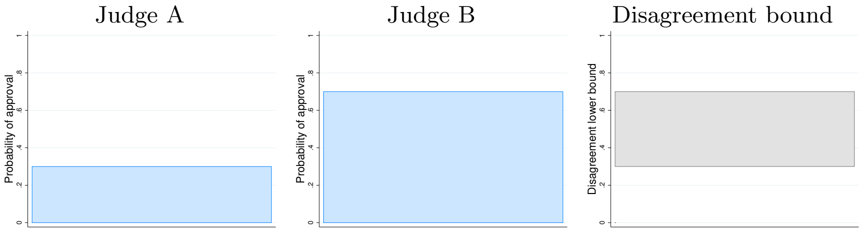
Disagreement between examiners A and B

$$\delta_{AB} \equiv P[y_{i1}(A_1) \neq y_{i1}(B_1)] \tag{1}$$

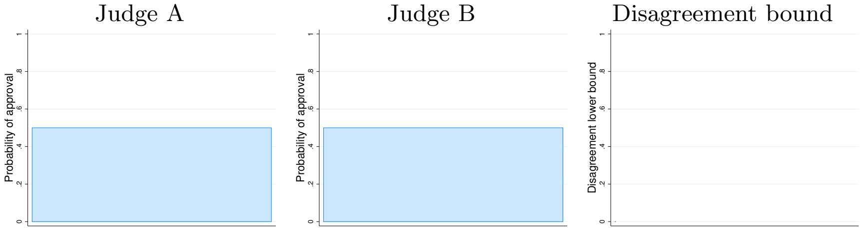
Empirical challenge: joint distribution of judge decisions not observed

- Disagreement not identified

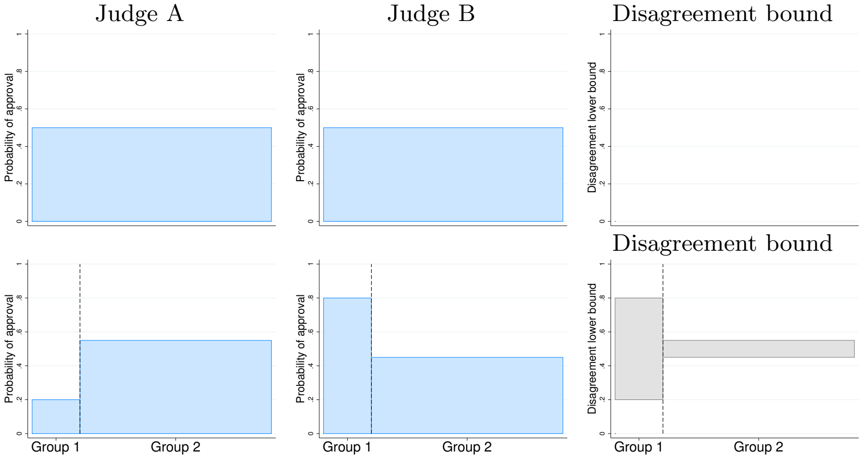
Second best: bounds on disagreement



Second best: bounds on disagreement



Second best: bounds on disagreement



Tricks with Bayes

Characteristic: second-round approval $y_{i2}(C)$

- By random assignment, same share to judges A and B
- Define judge assignment D_1^j

Rewrite bounds in terms of observables:

$$\delta_{AB}(C) \geq \overbrace{E[y_1|D_1^A] - E[y_1|D_1^B]}^{\text{Difference in approval rate}} + \underbrace{2\Big[E[y_1|D_1^B]E[y_2|D_1^B, D_2^C, y_1 = 1] - E[y_1|D_1^A]E[y_2|D_1^A, D_2^C, y_1 = 1]}_{\text{Extra information from second round}}$$

Tricks with Bayes

Characteristic: second-round approval $y_{i2}(C)$

- By random assignment, same share to judges A and B
- Define judge assignment D_1^j

Rewrite bounds in terms of observables:

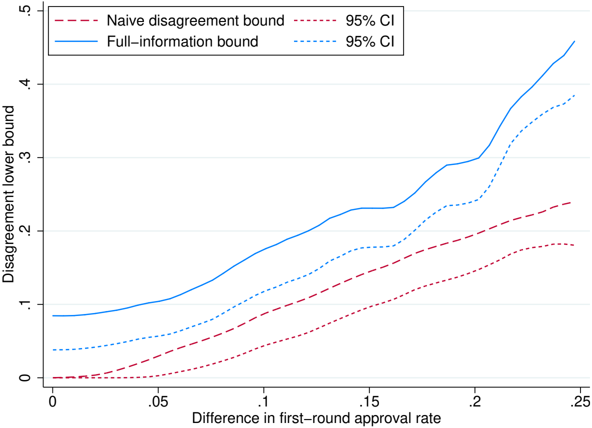
$$\delta_{AB}(C) \geq \overbrace{E[y_1|D_1^A] - E[y_1|D_1^B]}^{\text{Difference in approval rate}} + \underbrace{2\Big[E[y_1|D_1^B]E[y_2|D_1^B, D_2^C, y_1 = 1] - E[y_1|D_1^A]E[y_2|D_1^A, D_2^C, y_1 = 1]}_{\text{Extra information from second round}}$$

This is true for all different second-round judges j:

$$\delta_{AB} \geq \max_j(\delta_{AB}(j)) \quad (2)$$

- Report half-median unbiased bounds and CIs

Lower bounds on 1^{th} round judge-pair disagreement



Average pair disagrees on at least 17% of cases

- 7% bound using only first-round info

1st round information 2nd round information

Structural model

Judges observe common quality r_i with some error $\widetilde{\varepsilon}_{ijs}$

$$\text{Approve}_{ijs} = \mathbb{1}[r_i > \gamma_{js} + X_{ijs}\beta_s + \widetilde{\varepsilon}_{ijs}(W_{ijs})] \tag{3}$$

$$\widetilde{\varepsilon}_{ijs}(W_{ijs}) \sim \mathcal{N}(0, e^{\sigma_{js} + W_{ijs}\psi})$$

- Judges differ in standards γ_{js} and inaccuracy σ_{js}
- Cross-judge variation in γ_{js} increases disagreement
- Higher σ_{js} increases disagreement

Structural model

Judges observe common quality r_i with some error $\widetilde{\varepsilon}_{ijs}$

$$\text{Approve}_{ijs} = \mathbb{1}[r_i > \gamma_{js} + X_{ijs}\beta_s + \widetilde{\varepsilon}_{ijs}(W_{ijs})] \quad (3)$$

$$\widetilde{\varepsilon}_{ijs}(W_{ijs}) \sim \mathcal{N}(0, e^{\sigma_{js} + W_{ijs}\psi})$$

- Judges differ in standards γ_{js} and inaccuracy σ_{js}
- Cross-judge variation in γ_{js} increases disagreement
- Higher σ_{js} increases disagreement

Identification:

- Large-support X_{ijs} as special regressors
- Random assignment of judges
- Implemented with parametric restrictions on r_i and $\widetilde{\varepsilon}_{ijs}$

Coefficients

Table: First-round judge accuracy by experience and workload

	(1)	(2)	(3)	(4)
	Coefficients ψ_1 affecting log judge inaccuracy σ_1			
Experience > 1 year	-0.777^{***}	-1.154^{***}		
	(0.035)	(0.055)		
Experience > 5 years	-0.388^{***}	-0.524^{***}		
	(0.022)	(0.029)		
Experience > 10 years	-0.411^{***}	-0.747^{***}		
	(0.026)	(0.050)		
Log caseload				
Log caseload (≤ 5 yrs exp)				
Log caseload (> 5 yrs exp)				
Second-round experience control	Yes	Yes		
Career number of cases	No	Yes		

Standard errors clustered at judge level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table: First-round judge accuracy by experience and workload

	(1)	(2)	(3)	(4)
Coefficients ψ_1 affecting log judge inaccuracy σ_1				
Experience > 1 year	-0.777^{***}	-1.154^{***}	-0.736^{***}	-0.328^{***}
	(0.035)	(0.055)	(0.029)	(0.027)
Experience > 5 years	-0.388^{***}	-0.524^{***}	-0.414^{***}	0.049
	(0.022)	(0.029)	(0.144)	(0.048)
Experience > 10 years	-0.411^{***}	-0.747^{***}	-0.762^{***}	-0.936^{***}
	(0.026)	(0.050)	(0.073)	(0.032)
Log caseload			0.146^{***}	
			(0.022)	
Log caseload ($\leq$ 5 yrs exp)				0.305^{***}
				(0.005)
Log caseload (> 5 yrs exp)				0.034^{***}
				(0.007)
Second-round experience control	Yes	Yes	Yes	Yes
Career number of cases	No	Yes	No	No

Standard errors clustered at judge level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Judicial Advisory Councils (introduced 1988)

Gave independent committee veto over judicial candidates

- Designed to stop appointment of unqualified party supporters

$$\widehat{\sigma}_{j1}(W_{j1}) = \beta \text{Post-reform} + u_j \tag{4}$$

Judicial Advisory Councils (introduced 1988)

Gave independent committee veto over judicial candidates

- Designed to stop appointment of unqualified party supporters

$$\widehat{\sigma}_{j1}(W_{j1}) = \beta \text{Post-reform} + u_j \tag{4}$$

Table: Experience-adjusted inaccuracy for judges before and after reform

	(1)	(2)	(3)
Appointed after reform (=1)	-1.182***	-1.162***	-1.189***
	(0.180)	(0.259)	(0.271)
Liberal appointee (=1)			-0.0237
			(0.104)
Male judge (=1)			-0.132
			(0.129)
Year appointed	No	Yes	Yes
Pre-reform mean	1.72	1.72	1.72
Number of judges	53	53	53

Estimated with Hanushek (1974) correction for estimated dependent variable. Standard errors clustered at the judge level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Approval Promotion Approval control

Conclusion

Disagreement across decision-makers → costly uncertainty

- Never directly observed, so partially identified
- Multiple-round decision processes dramatically tighten bounds
- Canadian refugee judges disagree on 17% of cases
  - Halfway between perfect ordering agreement and flipping coins

Ordering accuracy improves:

- With experience
- When workloads lower
- When political appointments are limited

Conclusion

Disagreement across decision-makers → costly uncertainty

- Never directly observed, so partially identified
- Multiple-round decision processes dramatically tighten bounds
- Canadian refugee judges disagree on 17% of cases
    - Halfway between perfect ordering agreement and flipping coins

Ordering accuracy improves:

- With experience
- When workloads lower
- When political appointments are limited

Monotonicity in IV designs

- Level of disagreement inconsistent with monotonicity
- Appendix/website: new, powerful tests
    - In my setting, MTEs biased

Differences in approval rates

First part of bound:

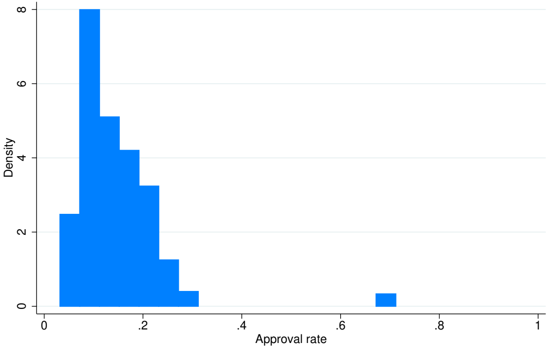
$$E[y_1|D_1^A] - E[y_1|D_1^B]$$

Differences in approval rates

First part of bound:

$$E[y_1|D_1^A] - E[y_1|D_1^B]$$

Figure: First round approval rates by judge



Back

Differences in subsequent success

Second part of bound (when judges equally lenient):

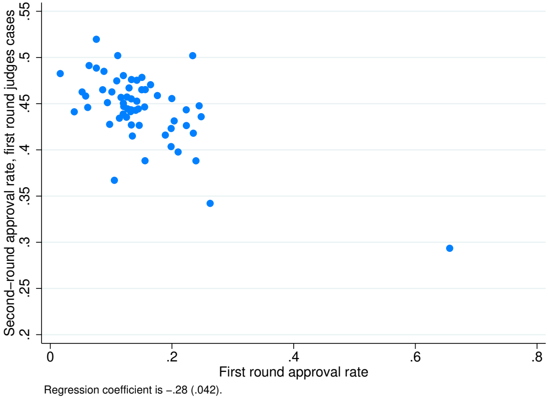
$$2E[y_1]\Big[E[y_2|D_1^B, D_2^C, y_1 = 1] - E[y_2|D_1^A, D_2^C, y_1 = 1]$$

Differences in subsequent success

Second part of bound (when judges equally lenient):

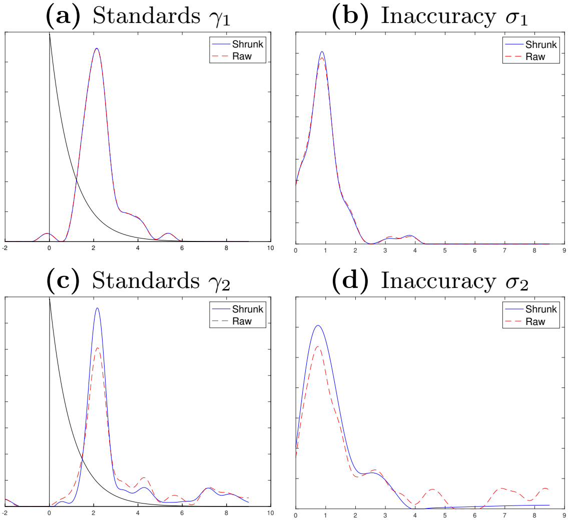
$$2E[y_1]\Big[E[y_2|D_1^B, D_2^C, y_1 = 1] - E[y_2|D_1^A, D_2^C, y_1 = 1]$$

Figure: Share of approved cases subsequently approved vs 1st round approval



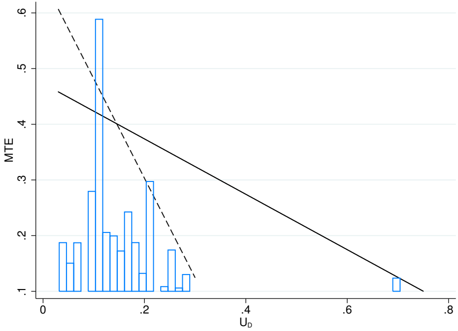
Back

$$\text{Approve}_{ijs} = \mathbb{1}[r_i > \gamma_{js} + X_{ijs}\beta_s + \widetilde{\varepsilon}_{ijs}], \quad \widetilde{\varepsilon}_{ijs} \sim \mathcal{N}(0, \sigma^2_{js})$$



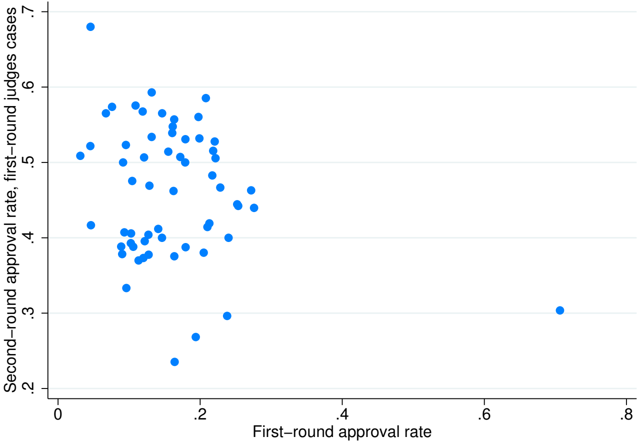
Back

Figure: MTE of second-round approval on first-round approval judge



The black line represents the MTE of first-round approval on second-round approval (implicitly, no one who is rejected in the first round is approved in the second round). First-round approval is instrumented by judge effects, the distribution of which is displayed as a histogram. The dashed line is the MTE estimated without the outlier point.

Figure: Second-round approval versus first-round approval



Regression coefficient is –.24 (.120).

Figure displays the second round approval rate for the claimants approved by each first round judge, plotted against judges first-round approval rates.

Back

Judge severity uncorrelated with characteristics

Table: Randomization

	Male	Africa	Asia	South America	IRB mean approval	Predicted approval	1st-round mean approval
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: First round judges							
First-round approval rate	0.011	-0.031	-0.089	-0.006	0.010	-0.002	
	(0.021)	(0.030)	(0.071)	(0.023)	(0.014)	(0.002)	
F-stat	0.88	2.82	9.54	2.47	4.23	2.99	
Prob	0.73	0.00	0.00	0.00	0.00	0.00	
Observations	50,435	50,435	50,435	50,435	50,435	50,435	
Panel B: Second round judges							
Second-round approval rate	-0.048	-0.012	0.000	0.032	-0.005	0.002	-0.024
	(0.030)	(0.037)	(0.042)	(0.042)	(0.019)	(0.002)	(0.018)
F-stat	1.01	1.53	1.71	1.22	1.54	2.11	3.07
Prob	0.45	0.01	0.00	0.12	0.00	0.00	0.00
Observations	7,143	7,143	7,143	7,143	7,143	7,143	7,143

IRB mean approval is the approval rate of the IRB Member who initially denied refugee status to the claimant. Predicted approval comes from a regression of approval on gender and continent of origin. F-stats come from separate regression of outcome on judge fixed effects. Standard errors clustered at the judge level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Table: Randomization using name-imputed continent of origin

	Male	Africa	Asia	South America	Predicted approval	1st-round mean approval
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First round judges						
First-round approval rate	-0.002	-0.068***	-0.011	0.098	-0.002	
	(0.020)	(0.023)	(0.040)	(0.071)	(0.002)	
F-stat	0.87	2.90	3.04	6.65	3.51	
Prob	0.75	0.00	0.00	0.00	0.00	
Observations	58,604	58,604	58,604	58,604	58,604	
Panel B: Second round judges						
Second-round approval rate	-0.042	0.032	-0.022	0.027	0.002	-0.027
	(0.033)	(0.039)	(0.042)	(0.041)	(0.003)	(0.018)
F-stat	1.07	1.83	1.19	1.61	1.54	4.02
Prob	0.33	0.00	0.15	0.00	0.01	0.00
Observations	8,446	8,446	8,446	8,446	8,446	8,446

Gender and continent of origin predicted from claimant name. IRB mean approval is the approval rate of the IRB Member who initially denied refugee status to the claimant. Predicted approval comes from a regression of approval on gender, continent of origin and IRB Member approval rate. F-stats come from separate regression of outcome on judge fixed effects. All regressions include office X pre-2002 fixed effects to account for cross-office differences in case strength and changes in government policy in 2002. Standard errors clustered at the judge level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Consistency definition

Want to compare first-round judges A and B with same approval rate

$$\mathbb{1}[r_i > \gamma_{j1} + \widetilde{\varepsilon}_{ij1}] \tag{5}$$

- $\widetilde{\varepsilon}_{ij1} \sim G_{j1} \implies \Pr[\text{Approval}|r] = G_{j1}(r - \gamma_{j1}) = \widetilde{G}_{j1}(r)$
- **Consistency:** there exists point of single-crossing v such that
    1. $\widetilde{G}_{A1}(v) = \widetilde{G}_{B1}(v)$
    2. $\forall\ w > v, \widetilde{G}_{A1}(w) \geq \widetilde{G}_{B1}(v)$, with a strict equality for some w
    3. $\forall\ w < v, \widetilde{G}_{A1}(w) \leq \widetilde{G}_{B1}(v)$, with a strict equality for some w

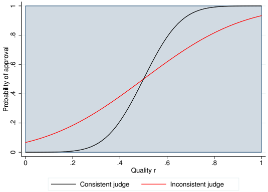
Consistency definition

Want to compare first-round judges A and B with same approval rate

$$\mathbb{1}[r_i > \gamma_{j1} + \widetilde{\varepsilon}_{ij1}] \tag{5}$$

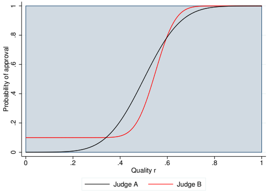
- $\widetilde{\varepsilon}_{ij1} \sim G_{j1} \implies \Pr[\text{Approval}|r] = G_{j1}(r - \gamma_{j1}) = \widetilde{G}_{j1}(r)$
- **Consistency:** there exists point of single-crossing v such that
  1. $\widetilde{G}_{A1}(v) = \widetilde{G}_{B1}(v)$
  2. $\forall\ w > v, \widetilde{G}_{A1}(w) \geq \widetilde{G}_{B1}(v)$, with a strict equality for some w
  3. $\forall\ w < v, \widetilde{G}_{A1}(w) \leq \widetilde{G}_{B1}(v)$, with a strict equality for some w

(d) Fits definition



(e) Fits definition

(f) Does not fit



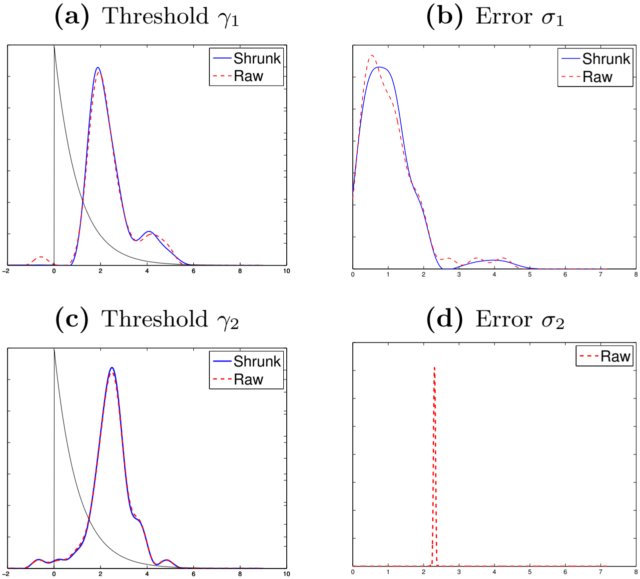
Back

Want to compare second-round judges A and B

- Known non-limiting judge C accepts nearly everyone
- Known comparison judge D
- Judge error distributions $\widetilde{\varepsilon} \sim \widetilde{G}_J(r) = G_J(r - \gamma_J)$
- Require: $\widetilde{G}_C / \widetilde{G}_D$ monotonically increasing whenever $\widetilde{G}_A \neq \widetilde{G}_B$
  - True whenever $G_D(r) = 1 \; \forall \; r$

Back

Figure: Distribution of judge coefficients



Back

Want to estimate

$$y_i = X\beta + e_i$$

but observe noisy estimate of y_i,

$$y_i^* = y_i + u_i$$

Efficient estimator:

$$y_i^* = X\beta + v_i \tag{6}$$

Estimate weights using knowledge of v_i

1. Note that $\sigma_v^2 = \sigma_u^2 + \sigma_i^2$
2. Estimate Equation 6
3. Combine $\hat{\sigma}_v^2$ and $\hat{\sigma}_i^2$ into $\hat{\sigma}_u^2$
4. Estimate Equation 6, weights $\frac{1}{\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_i^2}}$

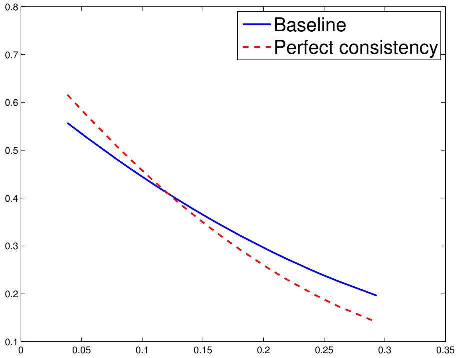
Back

MTE bias

Idea: calculate MTE of 1st-round approval on 2nd-round approval

- Baseline, perfect consistency ($\sigma_{j1} = 0$)

Figure: MTE and bias



Back

Table: Placebo tests and relevance for regressors, with judge fixed effects

	Predicted approval		Actual approval	
	(1)	(2)	(3)	(4)
Panel A: First round				
End of week	0.000	0.000	-0.008***	-0.007***
	(0.000)	(0.000)	(0.002)	(0.002)
Observations	58604	58604	58604	58604
Panel B: Second round				
End of week	0.001	0.001	-0.022*	-0.022*
	(0.001)	(0.001)	(0.012)	(0.012)
Noon hearing	-0.001	-0.001	-0.078***	-0.075***
	(0.001)	(0.001)	(0.022)	(0.023)
Controls	No	Yes	No	Yes
Observations	8446	8446	8446	8446

Controls include year filed and office. All specifications include judge fixed effects. Standard errors clustered at the judge level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Figure: Model estimates of first- versus second-round approval

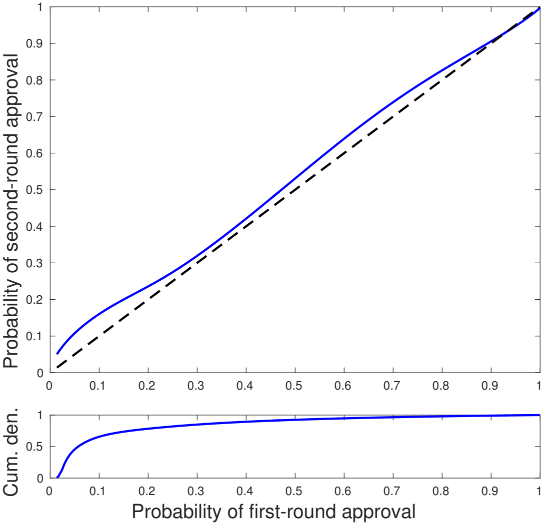
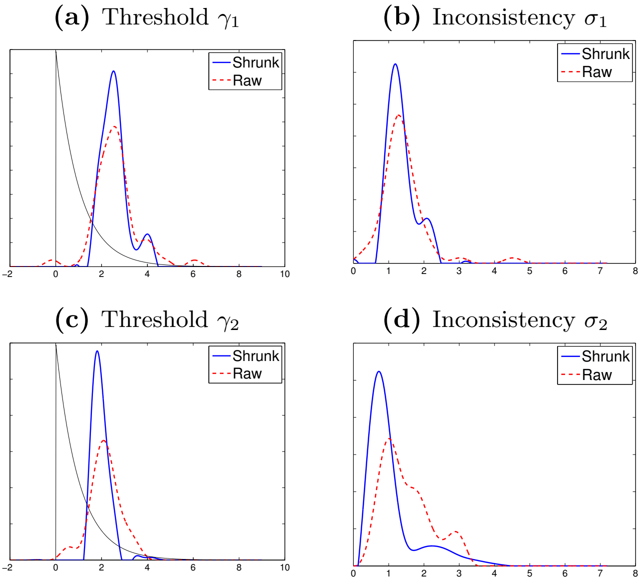


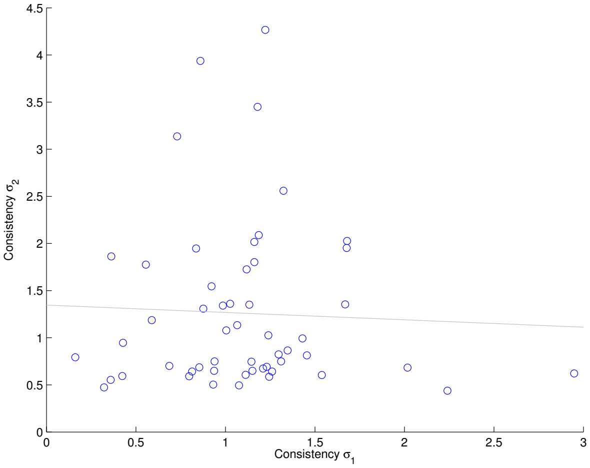
Figure plots first-round approval probability against second-round approval probability conditional on first-round approval for each case strength r_i. Secondary graph displays cumulative density of approval probability.

Figure: Distribution of judge coefficients, no regressors



Back

Figure: Scatter of σ_{j2} vs σ_{j1}



Back

Identification

Local identification (judges with similar approval rates)

- Judge assignment
- Identifies relative consistency

Identification

Local identification (judges with similar approval rates)

- Judge assignment
- Identifies relative consistency

Global identification (duration model, Chen et al 1999)

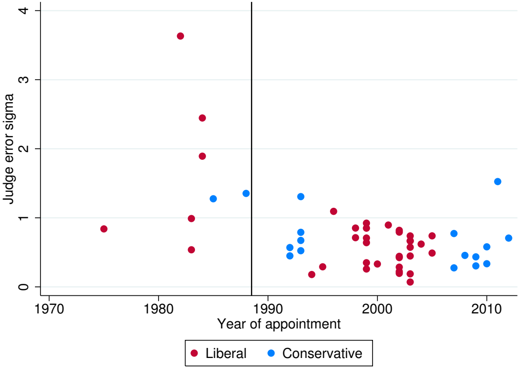
- Regressors X_{ijs} adjusts threshold γ_{js}

$$\text{Approve}_{ijs} = \mathbb{1}[r_i - \widetilde{\varepsilon}_{ijs} > X_{ijs}\beta + \gamma_{js}] \tag{7}$$

1. Use X_{ijs} to identify distribution of composite error
2. Identify relative size of r_i, $\widetilde{\varepsilon}_{ij1}$, $\widetilde{\varepsilon}_{ij2}$ from variance/covariance of round outcomes
    - Assumption: one component of β_s same across rounds

Back

Figure: Estimated σ_1 by year of appointment



Back

Approval rates before and after reform

Table: Approval rate for first-round judges before and after reform

	Approval rate			Approval, year residualized		
	(1)	(2)	(3)	(4)	(5)	(6)
Appointed after reform (=1)	0.00896	0.0764	0.0734	0.0103	0.0891	0.0831
	(0.0318)	(0.0716)	(0.0757)	(0.0309)	(0.0716)	(0.0751)
Liberal appointee (=1)			0.00317			0.0102
			(0.0192)			(0.0194)
Male judge (=1)			-0.00876			-0.0140
			(0.0257)			(0.0250)
Year appointed	No	Yes	Yes	No	Yes	Yes
Pre-reform mean	0.15	0.15	0.15	0.00	0.00	0.00
Number of judges	53	53	53	53	53	53

Regression and dependent variable mean estimated with Hanushek (1974) weights for estimated dependent variable. Robust standard errors in parentheses and clustered at the judge level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Potential promotion bias

Were accurate pre-reform lawyers promoted out of sample?

Table: Inaccuracy σ_{j1} for judges by whether ever promoted

	(1)	(2)	(3)
Ever promoted (=1)	-0.178	0.0202	0.0154
	(0.135)	(0.134)	(0.132)
Liberal appointee (=1)			-0.0996
			(0.114)
Male judge (=1)			-0.0435
			(0.175)
Year control	No	Yes	Yes
Number judges	53	53	53

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

Baseline model (no experience adjust)

Table: Decision error for judges before and after reform

	(1)	(2)	(3)
After 1988 reform (=1)	-0.226	-0.370	-0.382
	(0.154)	(0.236)	(0.241)
Liberal appointee (=1)			-0.0294
			(0.105)
Male judge (=1)			-0.137
			(0.105)
Year control	No	Yes	Yes
Dependent mean	1.07	1.07	1.07
N judges	53	53	53

Estimated with Hanushek (1974) correction for estimated dependent variable. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Control for judge approval rate

Table: Inaccuracy σ_{j1} before and after reform

	(1)	(2)	(3)
After 1988 reform (=1)	-1.177***	-1.118***	-1.144***
	(0.179)	(0.268)	(0.279)
Liberal appointee (=1)			-0.0219
			(0.104)
Male judge (=1)			-0.137
			(0.133)
Year control	No	Yes	Yes
Dependent mean	0.71	0.71	0.71
N judges	53	53	53

Estimated with Hanushek (1974) correction for estimated dependent variable. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Testing exclusion of X_{ijs}

Table: Testing effect of regressors on distribution of judge errors

	(1)	(2)	(3)
Coefficients β affecting judge threshold γ_1			
End-of-week decision	0.057***	0.087***	0.051**
	(0.004)	(0.029)	(0.021)
Hearing schedule over lunch	0.411***	0.381**	0.510***
	(0.077)	(0.193)	(0.165)
Coefficients ψ affecting judge inconsistency σ_1			
End-of-week decision		0.040	
		(0.058)	
Hearing schedule over lunch			0.371
			(1.287)
SD of γ_1	0.836	0.833	0.840
SD of σ_1	0.485	0.467	0.475

Reports coefficients for choice model $\mathbb{1}[r_i > X_{ijs}\beta_s + \gamma_{js} + \widetilde{\varepsilon}_{ijs}(X'_{ijs})]$, $\widetilde{\varepsilon}_{ijs}(X'_{ijs}) \sim \mathcal{N}(0, e^{\sigma_{js}+X'_{ijs}\psi})$. All models include controls for office of origination in X_{ijs}, and allow the parameters of the Pareto distribution of r_i to vary flexibly after 2002 relative to before 2002. Standard errors clustered at the level of the first stage judge. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Simplifying parametric assumptions

$$\text{Approve}_{ijs} = \mathbb{1}[r_i > X_{ijs}\beta + \gamma_{js} + \widetilde{\varepsilon}_{ijs}(X'_{ijs})] \tag{8}$$

- $r_i \sim$ exponential-Pareto(1,1) (mean = variance = 1)
- $\widetilde{\varepsilon}_{ijs}(X'_{ijs}) \sim \mathcal{N}(0, e^{\widetilde{\sigma}_{js}+X'_{ijs}\psi})$, $\widetilde{\sigma}_{j2} = \widetilde{\sigma}_2 \;\; \forall j$

Closed-form expressions for likelihood

- Generic (hard to calculate):

$$P[Approve_{ij1}] = \int G_{\widetilde{\varepsilon}_{ij1}(X'_{ij1})}(r_i - X_{ij1}\beta - \gamma_{j1}) f_r dr \tag{9}$$

- Parametric assumptions (easy):

$$P[Approve_{ij1}] = \Phi\left[-\frac{X_{ij1}\beta_1 + \gamma_{j1}}{\sigma_{j1}}\right] + \tag{10}$$

$$e^{-(X_{ij1}\beta_1+\gamma_{j1})+\frac{\sigma_{j1}^2}{2}}\left[1 - \Phi\left(-\frac{X_{ij1}\beta_1 + \gamma_{j1}}{\sigma_{j1}} + \sigma_{j1}\right)\right] \tag{11}$$

Back

Table: Model coefficients on survey responses

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Threshold γ_1 (SD=.92)			
Favorability, SD	-0.144	-0.237***			-0.055	-0.186*
	(0.104)	(0.086)			(0.114)	(0.102)
Inconsistency, SD			0.316***	0.212***	0.304***	0.134*
			(0.046)	(0.055)	(0.052)	(0.071)
			Panel B: Threshold γ_2 (SD=.99)			
Favorability, SD	-0.280**	-0.402***			-0.264***	-0.420***
	(0.107)	(0.080)			(0.097)	(0.090)
Inconsistency, SD			0.121	0.104	0.058	-0.053
			(0.078)	(0.093)	(0.069)	(0.079)
			Panel C: Inconsistency σ_1 (SD=2.17)			
Favorability, SD	0.082	0.020			0.152***	0.092
	(0.060)	(0.099)			(0.044)	(0.094)
Inconsistency, SD			0.184***	0.163***	0.224***	0.194***
			(0.047)	(0.056)	(0.049)	(0.061)
Respondent FE	No	Yes	No	Yes	No	Yes
Observations	182	182	182	182	182	182

Estimated with Hanushek (1974) correction for estimated dependent variable. Standard errors clustered at the judge level in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Table: Inconsistency σ_1 on survey responses with approval controls

	(1)	(2)	(3)	(4)	(5)	(6)
Favorability, SD	0.0408	-0.108			0.125**	-0.0430
	(0.0694)	(0.105)			(0.0539)	(0.0981)
Inconsistency, SD			0.166***	0.151**	0.210***	0.131***
			(0.0563)	(0.0583)	(0.0505)	(0.0437)
Approval rate control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	136	136	136	136	136	136

Estimated with Hanushek (1974) correction for estimated dependent variable. All models include respondent fixed effects. Standard errors clustered at the judge level in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Back

Lawyer survey predicts model estimates

Survey of lawyers with judge experience

- Favorability to claimants
- Consistency: predictability of decision, same standards as other judges

Lawyer survey predicts model estimates

Survey of lawyers with judge experience

- Favorability to claimants
- Consistency: predictability of decision, same standards as other judges

Table: Model coefficients on survey responses

	Threshold γ_1	Threshold γ_2	Inconsistency σ_1
	(1)	(2)	(3)
Favorability, SD	-0.186*	-0.420***	0.092
	(0.102)	(0.090)	(0.094)
Inconsistency, SD	0.134*	-0.053	0.194***
	(0.071)	(0.079)	(0.061)
Lawyer FE	Yes	Yes	Yes
Observations	182	182	182

Estimated with Hanushek (1974) correction for estimated dependent variable. All models include respondent fixed effects. Standard errors clustered at the judge level in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Full No FEs Approval controls Accuracy Back

Table: Model coefficients on survey responses

	(1)	(2)	(3)
Panel A: Threshold γ_1 (SD=.92)			
Favorability, SD	-0.237***		-0.186*
	(0.086)		(0.102)
Inconsistency, SD		0.212***	0.134*
		(0.055)	(0.071)
Panel B: Threshold γ_2 (SD=.99)			
Favorability, SD	-0.402***		-0.420***
	(0.080)		(0.090)
Inconsistency, SD		0.104	-0.053
		(0.093)	(0.079)
Panel C: Inconsistency σ_1 (SD=2.17)			
Favorability, SD	0.020		0.092
	(0.099)		(0.094)
Inconsistency, SD		0.163***	0.194***
		(0.056)	(0.061)
Lawyer FE	Yes	Yes	Yes
Observations	182	182	182

Estimated with Hanushek (1974) correction for estimated dependent variable. All models include respondent fixed effects. Standard errors clustered at the judge level in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Back

Table: Model coefficients on survey responses

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Threshold γ_1 (mean=2.38, SD=.93)						
Favorability, SD	-0.144		0.034	-0.206**		-0.202
	(0.129)		(0.194)	(0.101)		(0.164)
Accuracy, SD		-0.252***	-0.269**		-0.152***	-0.006
		(0.064)	(0.121)		(0.050)	(0.099)
Observations	174	174	174	174	174	174
Panel B: Threshold γ_2 (mean=2.09, SD=1.07)						
Favorability, SD	-0.269**		-0.177*	-0.398***		-0.284***
	(0.110)		(0.091)	(0.082)		(0.072)
Accuracy, SD		-0.260**	-0.122		-0.357***	-0.150
		(0.108)	(0.090)		(0.088)	(0.091)
Observations	174	174	174	174	174	174
Panel C: Inconsistency σ_1 (mean=1.31, SD=2.17)						
Favorability, SD	0.069		0.176*	0.047		0.043
	(0.069)		(0.099)	(0.108)		(0.152)
Accuracy, SD		-0.025	-0.151*		0.039	0.006
		(0.057)	(0.076)		(0.070)	(0.089)
Lawyer FE	No	No	No	Yes	Yes	Yes
Observations	174	174	174	174	174	174

Estimated with Hanushek (1974) correction for estimated dependent variable. Standard errors clustered at the judge level in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Taste-based vs observational errors

Is judge identity predictive beyond index model?

- Do pairs of judges disproportionately agree on cases?

Taste-based vs observational errors

Is judge identity predictive beyond index model?

- Do pairs of judges disproportionately agree on cases?
- $P_{jk} = P[\text{Approval by } j_2 | \text{Approval by } k_1]$

$$\mathbb{1}[\text{Approval by } j_2 | \text{Approval by } k_1] = \beta P_{jk} + \nu_{jk} + u_{ijk} \tag{12}$$

Taste-based vs observational errors

Is judge identity predictive beyond index model?

- Do pairs of judges disproportionately agree on cases?
- $P_{jk} = P[\text{Approval by } j_2 | \text{Approval by } k_1]$

$$\mathbb{1}[\text{Approval by } j_2 | \text{Approval by } k_1] = \beta P_{jk} + \nu_{jk} + u_{ijk} \qquad (12)$$

	Judge-pair round FEs		Judge-pair FEs	
	(1)	(2)	(3)	(4)
Model approval probability	0.945***	0.938***	0.967***	0.962***
	(0.146)	(0.169)	(0.0463)	(0.0470)
Model controls	No	Yes	No	Yes
Mean approval	0.44	0.44	0.44	0.44
F-stat for judge pairs	1.01	1.02	0.99	0.99
P–value	0.694	0.693	0.810	0.809
SD of judge-pair EB means	0.004	0.004	0.004	0.004
Observations	8,196	8,196	8,196	8,196

Regresses second-round approval on model-predicted likelihood of approval and judge-pair fixed effects. Standard errors clustered at the judge level in parentheses.

Dist p | No regressors | Characteristics | Back

Ideology vs observational errors, no regressors

Table: Second-round outcome on predicted approval and judge-pair FEs

	Judge-pair round FEs		Judge-pair FEs	
	(1)	(2)	(3)	(4)
Model approval probability	0.949***	0.947***	0.976***	0.975***
	(0.166)	(0.166)	(0.0478)	(0.0478)
Model controls	No	Yes	No	Yes
Mean approval	0.44	0.44	0.44	0.44
F-stat for judge pairs	1.02	1.02	0.98	0.99
P-value	0.318	0.314	0.641	0.610
Bootstrap p-value	0.674	0.657	0.834	0.804
SD of judge-pair EB means	0.006	0.006	0.004	0.003
Observations	8,196	8,196	8,196	8,196

Regresses second-round approval on model-predicted likelihood of approval and judge-pair fixed effects. Model estimated without using regressors for identification. Standard errors clustered at the judge level in parentheses.

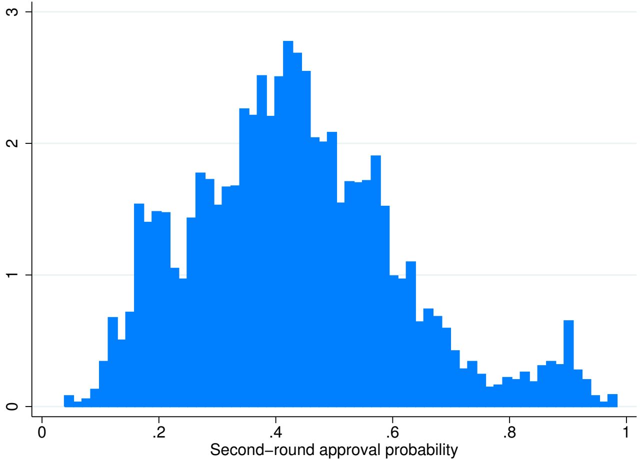
Back

Table: Second-round outcome on predicted approval and judge characteristics

	Outcome = approved in second round					
Characteristic =	French	Liberal	Male	Pre-reform	Exp>1	Exp>5
	(1)	(2)	(3)	(4)	(5)	(6)
Model approval probability	0.993***	1.004***	1.009***	1.001***	1.000***	1.001***
	(0.0321)	(0.0297)	(0.0294)	(0.0300)	(0.0290)	(0.0291)
Characteristic, first-round judge (=1)	-0.0128	-0.0172	-0.000628	0.00361	-0.0143	-0.0110
	(0.0146)	(0.0217)	(0.0272)	(0.0186)	(0.0411)	(0.0142)
Characteristic, second-round judge (=1)	-0.0115	-0.0303	-0.0216	-0.00727	-0.0962**	-0.0131
	(0.0157)	(0.0217)	(0.0266)	(0.0166)	(0.0433)	(0.0148)
Characteristic, both judges (=1)	0.0122	0.0443*	0.0264	0.0251	0.0918**	0.0258
	(0.0215)	(0.0258)	(0.0302)	(0.0378)	(0.0463)	(0.0205)
Model controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean approval	0.44	0.44	0.44	0.44	0.44	0.44
F-stat for characteristic pairs	0.39	1.48	1.28	0.27	4.69	0.53
Prob	0.759	0.217	0.281	0.847	0.003	0.661
SD of judge-pair EB means	0.077	0.036	0.015	0.001	0.015	0.018
Observations	8,196	8,196	8,196	8,196	8,196	8,196

Regresses second-round approval on model-predicted likelihood of approval and characteristics of the judges first and second round judges. Model controls include office of origination, pre-post 2002, and an end-of-week and noon hearing dummy for the second-round hearing. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Distribution of approval probability



Back

Optimal judge allocation

[label=optimal]
Problem: minimize judge effort such that

1. Posterior r_i of approved claimants first-order dominates baseline
2. Approve at least as many claimants
3. No judge works more than in baseline

Efficiency gains by:

- Moving judges to round where they are relatively precise
- Minimizing expensive second-round decisions

Optimal judge allocation

[label=optimal]
Problem: minimize judge effort such that

1. Posterior r_i of approved claimants first-order dominates baseline
2. Approve at least as many claimants
3. No judge works more than in baseline

Efficiency gains by:

- Moving judges to round where they are relatively precise
- Minimizing expensive second-round decisions

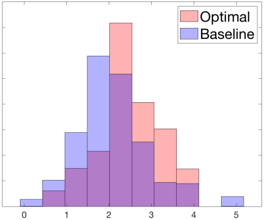
Results:

- 18% fewer judge hours $\rightarrow$ savings of \$4.9 million in salaries
- Higher posterior quality

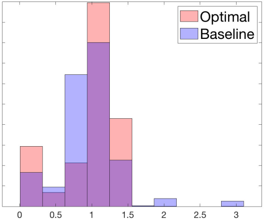
Back

Baseline vs optimal judge allocation

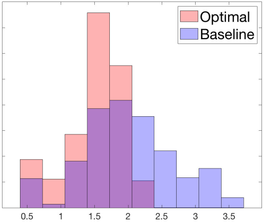
(a) Threshold γ_1



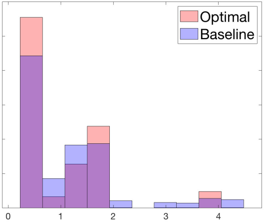
(b) Inconsistency σ_1



(c) Threshold γ_2



(d) Inconsistency σ_2

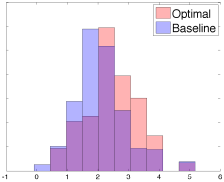


Back Second round informative

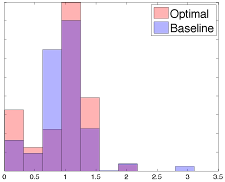
Second round informative

Maintain posterior distribution of $r_i + \widetilde{\varepsilon}_{ij2}$ (16% gains)

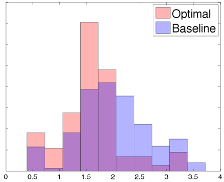
(a) Threshold γ_1



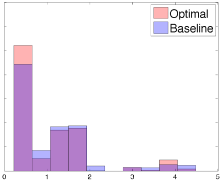
(b) Error σ_1



(c) Threshold γ_2

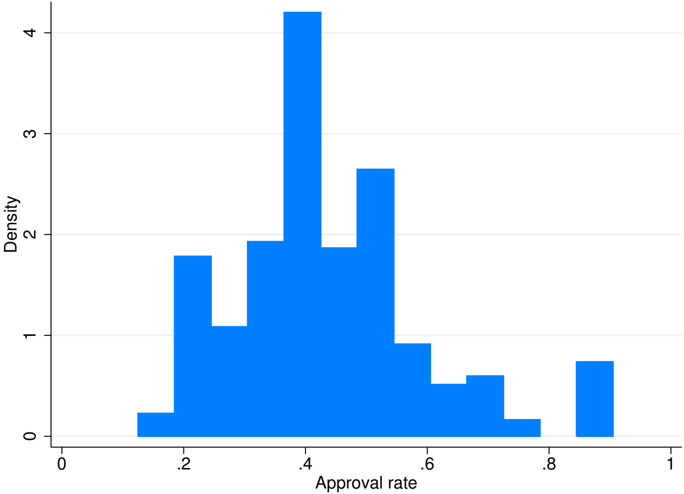


(d) Error σ_2



Back

Figure: Second round approval rates by judge



Back

Table: Second-round approval on mean approval rate of first-round judge

	(1)	(2)	(3)
Mean first round approval, exclusive	-0.264***	-0.312***	-0.324***
	(0.0521)	(0.0423)	(0.0437)
Mean second round approval, exclusive		0.958***	
		(0.0239)	
Second-round judge FE	No	No	Yes
Mean approval rate	0.44	0.44	0.44
Observations	8,446	8,446	8,446

Standard errors clustered by second-round judge. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure: Second-round approval versus first-round approval

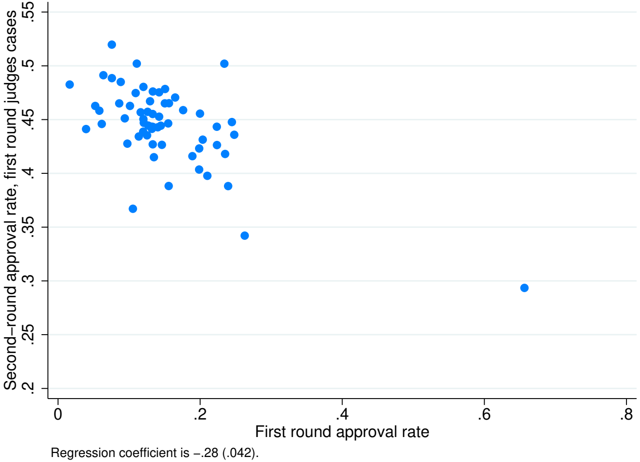


Figure displays the second round approval rate for the claimants approved by each first round judge, plotted against judges first-round approval rates. Empirical Bayes mean with second-round judge approval rate residualized out.

Non-residualized MTE