

IV in Practice: Judges and Examiners

PP5001 · Week 4 · Martinmas 2026

Professor David A. Jaeger

From Oregon to Judges and Examiners

Last week: a lottery changed the probability of obtaining insurance.

This week: assignment to an official changes the probability of receiving treatment.

- **Pre-trial detention:** assignment to bail judges.
- **Disability insurance:** assignment to disability examiners.

What makes these administrative assignments useful for causal inference?

Two Sources of Variation in Treatment

Officials apply common rules but may assess evidence or borderline cases differently.

Treatment can vary because:

- **Cases differ:** the same characteristics can influence treatment and the outcome.
- **Officials differ:** some grant treatment more readily than others.

A leniency design uses differences in officials' decisions together with an assignment process that makes their cases comparable.

Where Does Quasi-Random Variation Come From?

Imagine two officials receiving comparable applications from the same queue.

- One approves **40%** and the other approves **60%**.
- Assignment is unrelated to applicants' characteristics or potential outcomes.
- Encountering the more lenient official raises the probability of treatment by **20 percentage points**.

Assignment supplies independence. Differences in leniency supply relevance.

These are hypothetical rates illustrating the argument.

Understanding the Assignment Process

A lottery, rotation, or queue may make assignment as good as random within a group of cases.

Establish:

- Which officials could have received each case?
- Can applicants or staff influence the assignment?
- Do offices, shifts, or specialist teams receive different kinds of cases?

The comparison and controls must reflect these institutional details.

Comparable groups assigned to different officials can still contain systematically different **treated and untreated** people.

The IV Assumptions in Practice

- **Relevance:** does the official assigned affect treatment?
- **Independence:** why should assignment be unrelated to potential outcomes, conditional on the assignment controls?
- **Exclusion:** through which channels can the official affect the outcome?
- **Monotonicity:** do officials order treatment decisions in a common direction?

For each paper, connect the assumptions to the institution and the evidence.

Measuring Leniency

Let $R_i = 1$ denote a favourable initial decision for person i .

Measure official j 's leniency using decisions on **other cases**:

$$Z_i = \frac{\sum_{k:j(k)=j(i), k \neq i} R_k}{n_{j(i)} - 1}.$$

Here $j(i)$ is the official assigned to person i , and n_j is the official's number of cases.

Leaving out the person's own decision removes its direct contribution to the instrument.

A Leave-One-Out Example

An official approves **three of five** applications.

Applicant's decision	Approvals among the other four	Leave-one-out rate
Approved	2	$2/4 = 0.50$
Rejected	3	$3/4 = 0.75$

The useful variation comes from differences in underlying leniency **across officials**.

In the detention study, the authors also account for court and time, and exclude all cases belonging to the same defendant.

Pre-trial Detention

Dobbie, Goldin, and Yang (2018)

The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges

Philadelphia and Miami-Dade: court records linked to tax records.

The following exhibits are from the published article.

The Policy Question and Data

For discussion

- Read Sections I and II of Dobbie, Goldin, and Yang (2018).
- Explain the policy problem, the treatment, and the court and tax data.
- Why might released and detained defendants have different outcomes even without a causal effect of release?
- Identify the outcomes relevant to a policy recommendation.

Defining the Treatment

In these regressions,

$D_i = 1$ if released within three days of the bail hearing.

Z_i measures the assigned bail judge's tendency to release other defendants, after accounting for court and time.

Outcomes include guilty pleas, conviction, rearrest, and employment.

Read the signs as effects of release.

Judge Assignment and Balance: Table 3

For discussion

TABLE 3—TEST OF RANDOMIZATION		
	Pretrial release (1)	Judge leniency (2)
Male	−0.11781 (0.00716)	0.00007 (0.00015)
Black	−0.03941 (0.00362)	0.00003 (0.00017)
Age at bail decision	−0.01287 (0.00236)	−0.00005 (0.00006)
Prior offense in past year	−0.15492 (0.00739)	0.00019 (0.00012)
Number of offenses	−0.02409 (0.00120)	0.00000 (0.00002)
Felony offense	−0.25575 (0.01821)	0.00005 (0.00010)
Any drug offense	0.12528 (0.00909)	0.00013 (0.00019)
Any DUI offense	0.10966 (0.01679)	0.00019 (0.00024)
Any violent offense	−0.01740 (0.01838)	0.00003 (0.00017)

Dobbie, Goldin, and Yang (2018). Upper portion. Continued on the next slide.

Table 3: Balance (continued)

For discussion

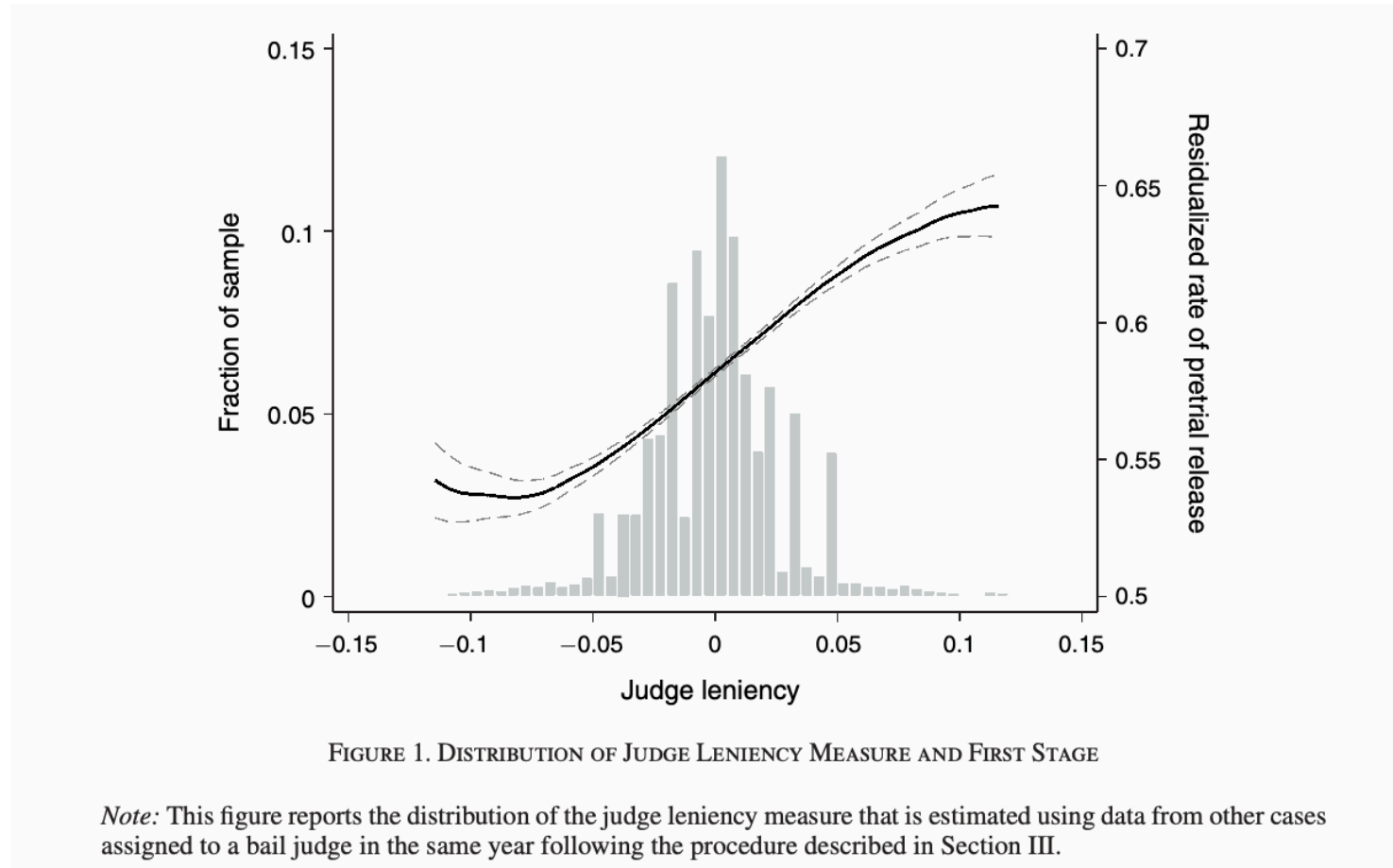
TABLE 3—TEST OF RANDOMIZATION		
	Pretrial release (1)	Judge leniency (2)
Any property offense	0.01097 (0.01688)	−0.00011 (0.00016)
Matched to IRS data	0.00868 (0.00194)	−0.00002 (0.00012)
Baseline earnings	0.00113 (0.00009)	−0.00001 (0.00000)
Baseline UI	0.00279 (0.00048)	−0.00001 (0.00002)
Baseline EITC	0.01233 (0.00087)	0.00002 (0.00008)
Baseline filed return	0.05136 (0.00387)	−0.00018 (0.00017)
Baseline employed	0.02523 (0.00272)	0.00019 (0.00015)
Baseline any EITC	−0.01856 (0.00418)	−0.00003 (0.00021)
Baseline any income	0.00000 (0.00000)	0.00000 (0.00000)
Baseline any UI	0.02431 (0.00363)	0.00026 (0.00029)
Joint <i>F</i> -test	[0.00000]	[0.78320]
Observations	421,065	421,065

Notes: This table reports reduced form results testing the random assignment of cases to bail judges. Judge leniency is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section III. Column 1 reports estimates from an OLS regression of pretrial release on the variables listed and court-by-time fixed effects. Column 2 reports estimates from an OLS regression of judge leniency on the variables listed and court-by-time fixed effects. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. The *p*-value reported at the bottom of columns 1 and 2 is for an *F*-test of the joint significance of the variables listed in the rows with the standard errors two-way clustered at the individual and judge-by-year level. See the online data Appendix for additional details on the sample and variable construction.

Dobbie, Goldin, and Yang (2018).

The First Stage: Figure 1

For discussion



Dobbie, Goldin, and Yang (2018).

Table 2

For discussion

TABLE 2—JUDGE LENIENCY AND PRETRIAL RELEASE			
	Sample mean	Judge leniency	
	(1)	(2)	(3)
<i>Panel A. Initial release</i>			
Released in 3 days	0.556 (0.497)	0.639 (0.063)	0.641 (0.062)
<i>Panel B. Subsequent bail outcomes</i>			
Bail modification petition	0.208 (0.406)	−0.407 (0.058)	−0.407 (0.052)
Released in 14 days	0.600 (0.490)	0.629 (0.053)	0.632 (0.052)
Released before trial	0.738 (0.440)	0.496 (0.032)	0.496 (0.029)
Court × time fixed effects	—	Yes	Yes
Baseline controls	—	No	Yes
Observations	421,065	421,065	421,065
<i>Notes:</i> This table reports first-stage results. The regressions are estimated on the sample as described in the notes to Table 1. Judge leniency is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section III. Column 1 reports the mean and standard deviation of the dependent variable. Column 2 reports results controlling for our full set of court-by-time fixed effects. Column 3 adds baseline controls: defendant race, defendant gender, defendant age, whether the defendant had a prior offense within the past year, number of offenses, indicators for whether the defendant is arrested for a drug, DUI, violent, or property offense, whether the most serious offense is a felony, whether the defendant was matched to the IRS data, baseline individual wages, baseline household wages, baseline UI, baseline EITC, baseline tax filing status, baseline employment, baseline any UI, baseline any EITC, baseline any income, and indicators for missing characteristics. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses in columns 2 and 3.			

Dobbie, Goldin, and Yang (2018).

Exclusion and Monotonicity

For discussion

- Explain why the separate assignment of bail and trial judges matters.
- Consider other ways bail conditions could affect later outcomes.
- Give an example in which two judges reverse their relative leniency across types of defendants.
- Connect each argument to the relevant IV assumption.

Interpreting the Results: Table 4

For discussion

TABLE 4—PRETRIAL RELEASE AND CRIMINAL OUTCOMES						
	Detained mean (1)	OLS results			2SLS results	
		(2)	(3)	(4)	(5)	(6)
<i>Panel A. Case outcomes</i>						
Any guilty offense	0.578 (0.494)	−0.072 (0.014)	−0.057 (0.009)	−0.046 (0.007)	−0.123 (0.047)	−0.140 (0.042)
Guilty plea	0.441 (0.497)	−0.188 (0.008)	−0.099 (0.010)	−0.082 (0.007)	−0.095 (0.056)	−0.108 (0.052)
Any incarceration	0.300 (0.458)	−0.161 (0.012)	−0.104 (0.006)	−0.110 (0.007)	0.006 (0.029)	−0.012 (0.030)
<i>Panel B. Court process outcomes</i>						
Failure to appear in court	0.121 (0.326)	0.063 (0.004)	0.010 (0.008)	0.021 (0.007)	0.158 (0.046)	0.156 (0.046)
Absconded	0.002 (0.045)	0.005 (0.000)	0.002 (0.000)	0.002 (0.000)	0.005 (0.004)	0.005 (0.004)
<i>Panel C. Future crime</i>						
Rearrest in 0–2 years	0.462 (0.499)	−0.050 (0.011)	−0.015 (0.006)	0.016 (0.005)	0.024 (0.061)	0.015 (0.063)
Rearrest prior to disposition	0.155 (0.362)	0.051 (0.008)	0.066 (0.007)	0.100 (0.007)	0.192 (0.038)	0.189 (0.042)
Rearrest after disposition	0.343 (0.475)	−0.075 (0.006)	−0.049 (0.002)	−0.041 (0.003)	−0.114 (0.057)	−0.121 (0.055)
Court × time fixed effects	—	Yes	Yes	Yes	Yes	Yes
Baseline controls	—	No	Yes	Yes	No	Yes
Complier weights	—	No	No	Yes	No	No
Observations	186,938	421,065	421,065	421,065	421,065	421,065

Dobbie, Goldin, and Yang (2018). Original table excerpt. Standard errors in parentheses, clustered by defendant and judge.

Table 5

For discussion

TABLE 5—PRETRIAL RELEASE AND LABOR MARKET OUTCOMES						
	Detained mean	OLS results			2SLS results	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Years 1–2</i>						
Earnings (\$ thousands)	5.224 (15.196)	2.689 (0.073)	0.389 (0.033)	0.162 (0.056)	0.030 (1.404)	−0.524 (0.966)
Household income (\$ thousands)	10.179 (22.844)	2.703 (0.162)	0.232 (0.090)	−0.042 (0.083)	1.809 (1.939)	−0.015 (1.439)
Employed	0.378 (0.485)	0.134 (0.003)	0.050 (0.002)	0.040 (0.003)	0.065 (0.049)	0.036 (0.042)
Any income	0.458 (0.498)	0.104 (0.003)	0.036 (0.002)	0.020 (0.003)	0.135 (0.073)	0.113 (0.064)
<i>Panel B. Years 3–4</i>						
Earnings (\$ thousands)	5.887 (15.897)	2.426 (0.093)	0.199 (0.055)	−0.039 (0.087)	−0.005 (1.441)	0.948 (1.128)
Household income (\$ thousands)	10.922 (23.974)	2.456 (0.171)	0.020 (0.111)	−0.107 (0.104)	1.090 (2.109)	0.181 (1.883)
Employed	0.378 (0.485)	0.104 (0.003)	0.031 (0.002)	0.021 (0.003)	0.099 (0.053)	0.094 (0.057)
Any income	0.461 (0.498)	0.090 (0.003)	0.030 (0.002)	0.023 (0.003)	0.125 (0.055)	0.107 (0.056)
Court × time fixed effects	—	Yes	Yes	Yes	Yes	Yes
Baseline controls	—	No	Yes	Yes	No	Yes
Complier weights	—	No	No	Yes	No	No
Observations	144,290	334,943	334,943	334,943	334,943	334,943

Dobbie, Goldin, and Yang (2018). Original table excerpt. Standard errors in parentheses, clustered by defendant and judge.

Disability Insurance

Maestas, Mullen, and Strand (2013)

Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt

Applications and examiner assignments linked to benefit decisions and subsequent earnings.

The Policy Question and Comparison

For discussion

- Read Sections I and II of Maestas, Mullen, and Strand (2013).
- Explain the policy question, data, and why comparing approved and rejected applicants may be misleading.
- Distinguish an initial allowance decision from eventual benefit receipt and explain the role of appeals.

Initial Decisions and Benefit Receipt

- **Treatment**, D_i : eventual receipt of SSDI benefits.
- **Instrument**, Z_i : the assigned examiner's initial allowance rate among other applicants.
- **Outcomes**, Y_i : employment and earnings.

An initial rejection may be followed by an appeal and eventual receipt of benefits.

The initial allowance decision and the treatment are therefore distinct.

Examiner Assignment: Table 2

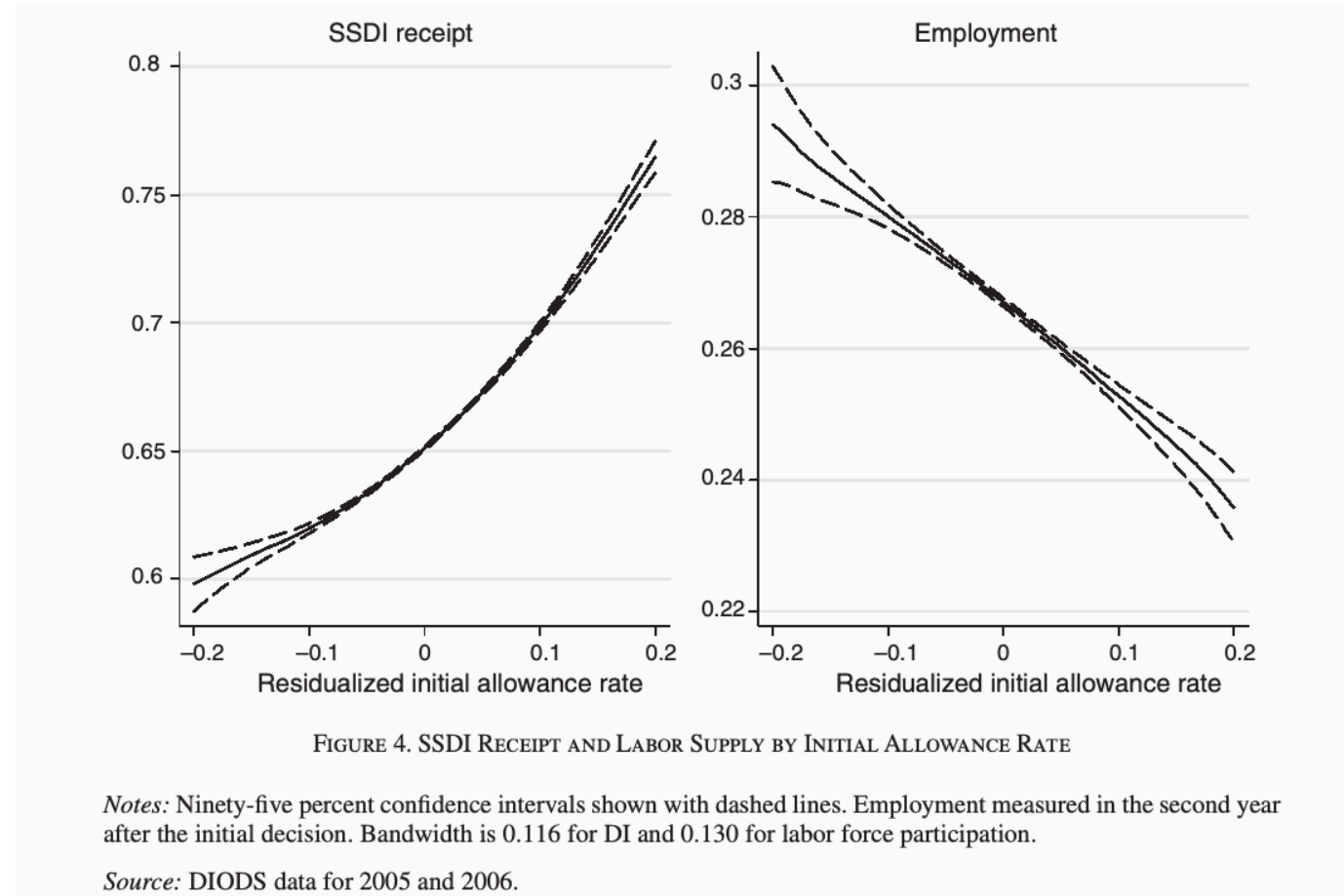
For discussion

TABLE 2— FIRST-STAGE REGRESSIONS: EFFECT OF EXALLOW ON SSDI RECEIPT							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>2005</i>							
Coefficient on EXALLOW	0.278***	0.271***	0.220***	0.215***	0.209***	0.204***	0.204***
<i>t</i> -stat	31.01	30.45	26.56	25.97	26.82	26.15	26.1
Fraction of always takers (P_A)							0.59***
<i>t</i> -stat							3.53
R^2	0.012	0.017	0.042	0.043	0.118	0.128	0.128
<i>2006</i>							
Coefficient on EXALLOW	0.309***	0.306***	0.258***	0.254***	0.245***	0.242***	0.243***
<i>t</i> -stat	31.73	31.66	30.88	30.7	32.48	32.16	32.28
Fraction of always takers (P_A)							0.56***
<i>t</i> -stat							3.32
R^2	0.013	0.018	0.044	0.045	0.117	0.127	0.128
<i>2005 and 2006</i>							
Coefficient on EXALLOW	0.294***	0.289***	0.240***	0.235***	0.227***	0.224***	0.226***
<i>t</i> -stat	41.87	40.73	38.81	38.09	40.2	39.33	39.9
Fraction of always takers (P_A)							0.57***
<i>t</i> -stat							3.42
R^2	0.012	0.017	0.042	0.043	0.117	0.127	0.127
Control variables included							
Three-digit zip code		X	X	X	X	X	X
Body system codes			X	X	X	X	X
Terminal illness indicator				X	X	X	X
Age group dummies					X	X	X
Average previous earnings						X	X
Month dummies							X
Control variables	112	1,015	1,031	1,031	1,040	1,041	1,052
<i>Notes:</i> Standard errors clustered at examiner level; <i>t</i> -statistics in parentheses. Mean of dependent (independent) variable = 0.65 (0.37). All regressions include DDS dummies. Pooled specification includes year interacted with month dummies. Number of observations is 1,090,345 for 2005 and 1,123,721 for 2006.							
***Significant at the 1 percent level.							
**Significant at the 5 percent level.							

Maestas, Mullen, and Strand (2013). p. 1814.

First Stage and Reduced Form: Figure 4

For discussion



Maestas, Mullen, and Strand (2013). p. 1813.

Table 5

For discussion

TABLE 5—REDUCED FORM REGRESSIONS: EFFECT OF EXALLOW ON EMPLOYMENT AND EARNINGS

Outcome	2005 decisions			2006 decisions	
	Two years	Three years	Four years	Two years	Three years
1) Earn $\geq$ \$1,000/year					
Coefficient on EXALLOW	−0.057***	−0.047***	−0.033***	−0.069***	−0.050***
	(−8.05)	(−6.60)	(−4.55)	(−10.51)	(−8.00)
R^2	0.082	0.086	0.077	0.077	0.070
2) Earn $\geq$ SGA					
Coefficient on EXALLOW	−0.039***	−0.034***	−0.023***	−0.042***	−0.037***
	(−7.21)	(−6.23)	(−4.32)	(−8.34)	(−7.79)
R^2	0.049	0.054	0.051	0.046	0.044
3) Earnings					
Coefficient on EXALLOW	−772***	−619***	−355	−1,114***	−804***
	(−2.97)	(−2.83)	(−1.57)	(−5.77)	(−4.49)
R^2	0.066	0.062	0.054	0.071	0.058
Observations	1,090,345	1,069,494	1,042,666	1,123,721	1,094,347

Notes: Standard errors clustered at examiner level; *t*-statistics in parentheses. Control variables include: DDS dummies, three-digit zip codes, body system codes, terminal illness indicator, age group dummies, average previous earnings, and month dummies.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Maestas, Mullen, and Strand (2013). p. 1820.

Connecting the Two Stages

With one excluded instrument and one endogenous explanatory variable,

$$\hat{\beta}_1^{IV} = \frac{\hat{\delta}_1}{\hat{\pi}_1}.$$

Use the **same sample and controls** in the reduced form and first stage.

For employment two years after a 2005 decision:

$$\frac{-0.057}{0.204} \approx -0.279.$$

What does this estimate mean, and for whom?

Interpreting Table 4: Panel A

For discussion

TABLE 4—EFFECTS OF SSDI RECEIPT ON EMPLOYMENT AND EARNINGS

Outcome	Two years after decision		Three years after decision		Four years after decision	
	OLS	IV	OLS	IV	OLS	IV
<i>Panel A. 2005 decisions</i>						
1) Earn $\geq$ \$1,000/year						
Mean dependent variable allowed	0.148		0.128		0.106	
Mean dependent variable denied	0.522		0.515		0.471	
Coefficient on ALLOW	−0.347*** (−322.48)	−0.279*** (−8.64)	−0.361*** (−336.60)	−0.227*** (−6.99)	−0.345*** (−321.81)	−0.158*** (−4.83)
R^2	0.200	0.195	0.218	0.200	0.209	0.171
2) Earn $\geq$ SGA						
Mean dependent variable allowed	0.050		0.043		0.033	
Mean dependent variable denied	0.293		0.302		0.270	
Coefficient on ALLOW	−0.242*** (−256.29)	−0.192*** (−7.62)	−0.255*** (−264.05)	−0.166*** (−6.70)	−0.233*** (−252.71)	−0.113*** (−4.59)
R^2	0.149	0.144	0.166	0.152	0.156	0.128
3) Earnings						
Mean dependent variable allowed	1,951		1,737		1,494	
Mean dependent variable denied	8,928		9,191		8,496	
Coefficient on ALLOW	−7,435*** (−126.50)	−3,781*** (−3.05)	−7,715*** (−182.51)	−3,007*** (−2.92)	−7,221*** (−176.24)	−1,716 (−1.60)
R^2	0.133	0.117	0.145	0.114	0.125	0.084
Observations	1,090,345		1,069,494		1,042,666	

Maestas, Mullen, and Strand (2013). p. 1819. Original excerpt. Parentheses contain t-statistics. For a test of zero, $SE(\hat{\beta}) = |\hat{\beta}| / |t|$. Standard errors are clustered by examiner. All specifications include the controls listed in the paper.

Processing Times and Exclusion: Figure 6

For discussion

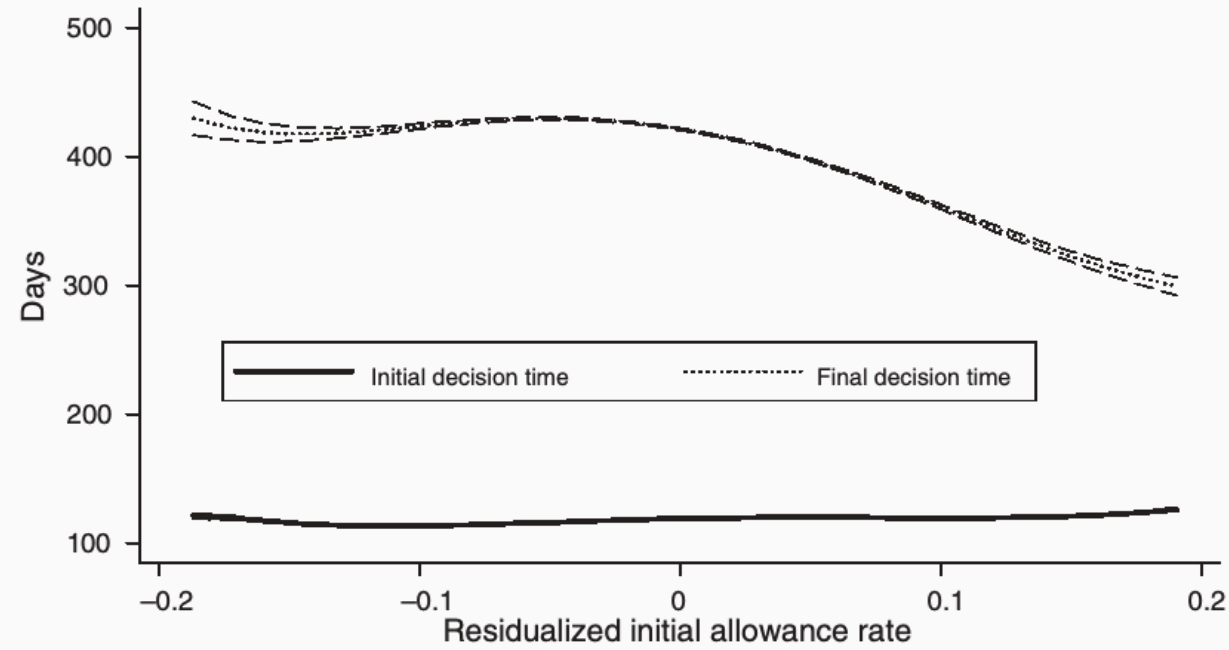


FIGURE 6. INITIAL AND FINAL PROCESSING TIMES BY INITIAL ALLOWANCE RATE

Notes: Ninety-five percent confidence intervals shown with dashed lines. Final processing times are measured through the appeals stage. Bandwidth is 0.067 for initial and 0.082 for final processing time.

Source: DIODS data for 2005 and 2006.

Maestas, Mullen, and Strand (2013). p. 1821.

Whose Treatment Effect?

With two officials, a complier receives treatment under the more lenient official but not under the stricter official:

$$D_i^1 = 1, \quad D_i^0 = 0.$$

The binary-instrument LATE is

$$E [Y_i^1 - Y_i^0 \mid D_i^1 > D_i^0] .$$

With many officials, a common monotone ordering supports a weighted average of effects across the margins changed by leniency.

Comparing the Designs and Policy Implications

For discussion

- For each paper, give the strongest institutional argument for independence, a plausible threat to exclusion, and a description of the people at the treatment margin.
- Propose a policy change for which the estimate is informative.
- What further evidence would a broader reform require?

References

- Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. “The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges.” *American Economic Review* 108(2): 201–240.
- Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. “Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt.” *American Economic Review* 103(5): 1797–1829.
- Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2026. “Leniency Designs: An Operator’s Manual.” *Journal of Economic Perspectives* 40(3): 213–240.