

Week 4: IV in Practice — Judges and Examiners

Pre-trial detention and disability insurance

[Class slides](#) · [Download slides as PDF](#) · [Download notes as PDF](#) · [Download Quarto source](#) · [After-class exercise](#)

Learning Objectives

By the end of this topic, you should be able to:

- Explain how the assignment of cases to judges or examiners can generate quasi-random variation in treatment.
- Construct a leave-one-out leniency instrument and explain why the assignment process determines which comparisons are credible.
- Assess relevance, independence, exclusion, and monotonicity using the institutional details and empirical evidence in the papers.
- Interpret the estimated effects for marginal defendants or applicants and assess what they imply for changes in policy.

Before class

Read these notes and both empirical papers below in full. Both papers are available on Moodle. Our discussion will focus on how the institutions generate an instrument, the evidence supporting its assumptions, and the policy interpretation of the estimates. The “Your turn” boxes below contain the discussion questions. Prepare your assigned items individually, using the discussant list on Moodle.

- **Dobbie, Goldin, and Yang (2018)**, “The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges.” Our class discussion will draw particularly on **Figure 1 and Tables 2–5**. [Published paper](#).
- **Maestas, Mullen, and Strand (2013)**, “Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt.” Our class discussion will draw particularly on the assignment of examiners, **Figure 4, Table 2, Tables 4–5, and Figure 6**. [Published paper](#).

Textbook references

Revisit the IV assumptions and LATE from Week 3:

- **The Effect:** [Chapter 19, Instrumental Variables](#).
- **Causal Inference: The Remix:** [Chapter 7, Instrumental Variables](#).
- **Mastering 'Metrics:** [Chapter 3, Instrumental Variables](#). [Companion resources](#) · [Mastering Econometrics videos at Marginal Revolution University](#).
- **Mostly Harmless Econometrics — advanced reading:** [Chapter 4](#), particularly the discussion of heterogeneous effects and LATE.

Additional reading: Goldsmith-Pinkham, Hull, and Kolesár (2026), “Leniency Designs: An Operator’s Manual”.

1. How leniency designs generate a comparison

Decisions made by different officials

Many policies give an official discretion over whether someone receives a treatment. The official reviews a case and decides whether to approve an application, grant access to a service, or impose a restriction. Even when officials apply the same formal rules, they can differ in how they assess evidence or interpret a borderline case. An official who grants the treatment more readily is described as **more lenient**. Here leniency refers to the tendency to grant the particular treatment being studied.

There are therefore two reasons why treatment decisions can differ across people: their cases differ, and the officials deciding their cases differ. The first source raises the selection problem we have discussed throughout the module. Characteristics that influence approval can also influence the outcome. The second source can supply an instrument when the assignment process gives otherwise comparable people different chances of encountering a lenient official.

Where does the quasi-random variation come from?

Imagine an office with two officials who receive applications from the same queue. A central allocation system sends each new application to one of them. Suppose this allocation is unrelated to the applicant's characteristics or potential outcomes. One official approves 40% of the office's applicants, while the other approves 60%. These are hypothetical population rates, illustrating a persistent difference in their decisions.

An applicant's chance of approval then depends partly on which official happens to receive the file. Across many assignments, the two officials receive comparable applicants, including comparable distributions of characteristics the researcher cannot observe. Assignment to the more lenient official raises the probability of treatment by 20 percentage points.

The assignment process supplies the quasi-random comparison. Differences in leniency make that comparison predict treatment. Together they provide the independence and relevance arguments familiar from IV.

An institution might allocate cases by an explicit lottery. In other settings, a rotation, a queue, or the official available when a case arrives may produce an assignment that is plausibly as good as random within a suitable group of cases. The researcher establishes why the actual procedure has this property. For example, a rotation is persuasive if applicants cannot time their arrival to obtain a particular official and case characteristics do not systematically change with the rotation. Preferential allocation of difficult cases to experienced officials would require a different argument and an appropriate comparison group.

The term **quasi-random** captures this institutional claim: within the comparison being used, who receives the more lenient official is effectively a matter of chance with respect to the determinants of the outcome.

Assignment and treatment selection

For the two-official example, let $Z_i = 1$ mean assignment to the more lenient official and $Z_i = 0$ assignment to the stricter official. Let $D_i = 1$ mean that the applicant receives treatment.

The official still assesses the case. Applicants with stronger claims may be approved by either official, while those with weaker claims may be rejected by either. Thus the treated and untreated groups can differ systematically even when the **assigned-official groups** are comparable. This is the same distinction we made in Oregon between lottery assignment and obtaining insurance.

We can first compare outcomes by assigned official. With quasi-random assignment, this comparison identifies the effect of assignment. To interpret the difference as operating through treatment, we need the exclusion restriction: the official affects the outcome through the treatment decision. The Wald estimator then scales the outcome difference by the treatment-probability difference. Monotonicity supplies the local treatment-effect interpretation when effects differ across individuals.

Which applicants are comparable?

An office may allocate different categories of cases to different teams. Officials may also serve different locations or shifts. Suppose applicants with more serious conditions go to a specialist team whose members approve applications more often. Comparing that team's applicants with the rest of the office would combine differences in official leniency with differences in the applicants' conditions.

The useful comparison would instead be **within the teams and periods in which applicants could have been assigned to the same officials**. Controls for those assignment groups preserve that comparison in the regression. Their choice follows from how the institution allocates cases.

To assess the argument, establish who can be assigned to whom, whether applicants or staff can influence the assignment, and whether particular officials receive systematically different cases. Balance checks on characteristics measured before assignment provide empirical evidence alongside that institutional description. The argument about unobserved characteristics rests on understanding the assignment process.

This week's applications

The papers below use these ideas in two settings. As you read, identify the assignment process first, then consider how differences in officials' decisions change treatment probabilities.

	Pre-trial release	Disability insurance
Individual	Defendant	Applicant
Treatment, $D_i = 1$	Released within three days of the bail hearing	Receives SSDI benefits
Instrument, Z_i	Assigned bail judge's release tendency, estimated using other defendants	Assigned examiner's initial allowance rate, estimated using other applicants
Examples of Y_i	Conviction, guilty plea, employment	Employment, earnings

Treatment coding matters: the detention paper's principal regressions estimate the effect of **release**. A negative coefficient on release corresponds to a positive effect of detention for the same comparison.

The assumptions in this setting

- **Relevance:** assignment to a more lenient official changes the probability of treatment.
- **Independence, conditional on assignment controls:** within the relevant assignment groups, the official assigned is unrelated to potential outcomes and potential treatment decisions.
- **Exclusion:** changing the assigned official affects the outcome through treatment.
- **Monotonicity:** moving from a stricter to a more lenient official moves treatment decisions in a common direction for individuals.

We also retain the treatment definitions and no-interference conditions discussed under SUTVA. For example, if one official's decisions change the resources available for other cases, we should consider whether those spillovers matter for the comparison.

In the linear model, the assignment and exclusion arguments support the orthogonality condition for IV. With heterogeneous effects, the additional monotonicity argument gives the estimates a local treatment-effect interpretation.

2. Measuring leniency

Suppose official j handles n_j cases. Write $R_i = 1$ when the official makes a favourable initial decision for person i . The official's average decision rate includes R_i itself. Thus a person's own circumstances help determine both their decision and this average.

We instead calculate the official's favourable-decision rate among **other cases**:

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

Here $j(i)$ denotes the official assigned to person i . This is a **leave-one-out** measure: we exclude person i when measuring the leniency of their assigned official.

For example, an official approves three of five cases. For an approved applicant, the leave-one-out rate is $2/4 = 0.50$. For a rejected applicant, it is $3/4 = 0.75$. Each applicant's instrument uses the decisions on the other four cases. Within this official's five cases, approved applicants therefore have a lower leave-one-out value (0.50) than rejected applicants (0.75). This happens because the calculation removes an approval for the former and a rejection for the latter. The useful first-stage variation comes from differences in underlying leniency **across officials**.

The administrative details determine how this calculation is implemented. Dobbie, Goldin, and Yang first remove court-and-time effects from release decisions, then calculate judge-year averages excluding **all cases belonging to the same defendant**. Maestas, Mullen, and Strand use other applicants' **initial** allowance decisions to instrument **eventual** benefit receipt, allowing for appeals and subsequent decisions.

Your turn

Why do we need to leave out individual i when constructing the instrument for the i th observation?

Accounting for the assignment process

If different kinds of cases are assigned to different offices, shifts, or specialists, compare cases within the groups for which assignment is plausibly as good as random. The corresponding controls belong in the first stage, reduced form, and outcome equation.

Using our notation, the first stage is

$$D_i = \pi_0 + \pi_1 Z_i + \sum_{k=1}^K \pi_{k+1} X_{ki} + v_i,$$

and the reduced form is

$$Y_i = \delta_0 + \delta_1 Z_i + \sum_{k=1}^K \delta_{k+1} X_{ki} + e_i.$$

The X_{ki} include the assignment controls. With one excluded instrument and one endogenous explanatory variable, the model is exactly identified. Using the same observations and controls in both regressions,

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

The first stage regresses the endogenous explanatory variable on all the exogenous variables. The reduced form regresses the outcome on all the exogenous variables.

3. Pre-trial detention

Policy question and data

Dobbie, Goldin, and Yang study bail hearings in Philadelphia and Miami-Dade, linking court records to administrative tax records. Detention can affect several policy outcomes at once: appearance in court, subsequent arrests, the resolution of the current case, and employment.

Your turn: Explain the policy question

What decision is being evaluated? Explain the treatment and outcome variables. Why might defendants released before trial have different outcomes from detained defendants even if release had no causal effect? Which outcomes would you want a policy evaluation to consider?

Assignment and balance

Bail judges work rotating schedules. The comparison accounts for court and time of the hearing. Philadelphia's shift system requires additional attention to time of day. The paper explains why defendants generally have limited ability to choose their bail judge, while also discussing possible selection in who reaches a hearing.

The balance exercise compares two relationships: the association of observed characteristics with **actual release**, and their association with the **assigned judge's leniency**. These answer distinct questions about treatment selection and instrument assignment.

Your turn: Assess assignment and balance

Use the institutional description and Table 3. How is a judge assigned? Which comparisons require controls for place or time? Explain what the two columns tell us. What concerns would remain even if the observed characteristics were balanced across the instrument?

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)
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.

Table 3. Dobbie, Goldin, and Yang (2018).

Balance provides evidence about the assignment process. Independence also requires an argument about characteristics we cannot observe. Relate the empirical check to the way cases arrive and judges rotate.

Relevance

Your turn: Explain the first stage

Use Figure 1 and Panel A of Table 2. Identify the axes in the figure and the dependent variable in the table. What does a 0.10 increase in judge leniency predict for the probability of release? Explain why the first-stage coefficient need not equal one.

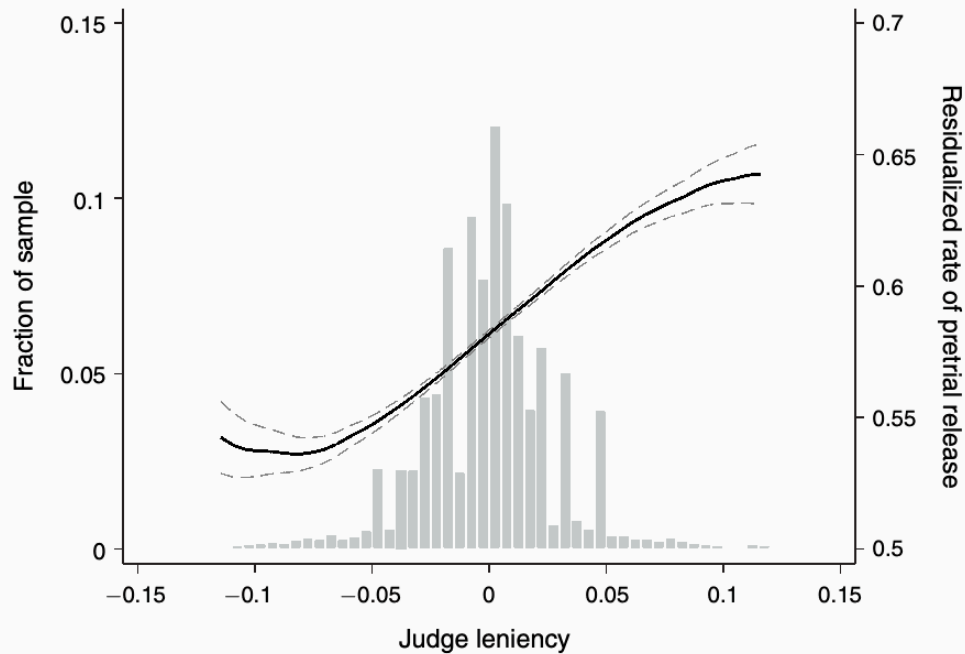


FIGURE 1. DISTRIBUTION OF JUDGE LENIENCY MEASURE AND FIRST STAGE

Note: This figure reports the distribution of the judge leniency measure that is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section III.

Figure 1. Dobbie, Goldin, and Yang (2018).

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

Notes: 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.

Table 2. Dobbie, Goldin, and Yang (2018).

Exclusion and monotonicity

Your turn: Assess the causal interpretation

Why does it matter that bail judges and trial judges are assigned through different processes? Could bail conditions affect later outcomes through a channel besides whether the defendant is released? Describe a pair of defendants for whom judges might reverse their ranking in leniency. Which IV assumption would that challenge?

A useful way to assess exclusion is to list everything the assigned official can change. If a judge changes financial conditions, supervision, or another aspect of the defendant's experience that independently affects the outcome, a binary release indicator may not capture the full effect of assignment. The separation of bail and trial responsibilities helps address a different concern: the possibility that the instrument also assigns a more lenient decision-maker at conviction or sentencing.

Monotonicity concerns individuals' potential decisions. A judge with a higher average release rate could still be stricter for a particular type of case. Evidence that first stages are positive across observed groups is useful when assessing this assumption. The underlying requirement concerns how the same defendant would be treated by different judges.

Results and policy interpretation

Your turn: Interpret the results

Use Tables 4 and 5, focusing on the OLS estimate with baseline controls and the corresponding 2SLS estimate (columns 3 and 6). Interpret the effects on guilty pleas, rearrest, and employment. State the units and describe uncertainty. How might release influence a defendant's willingness to plead guilty? Which policy trade-offs emerge across the outcomes?

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

Notes This table reports OLS and two-stage least squares results of the impact of pre-trial release. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Two-stage least squares models instrument for pretrial detention using a judge leniency measure that is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section III. All specifications control for court-by-time fixed effects. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses in columns 2–6.

Table 4. Dobbie, Goldin, and Yang (2018).

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

Notes: This table reports OLS and two-stage least squares results of the impact of pretrial release. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is listed in each row. Two-stage least squares models instrument for pretrial detention using a judge leniency measure that is estimated using data from other cases assigned to a bail judge in the same year following the procedure described in Section III. All specifications control for court-by-time fixed effects. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses in columns 2–6.

Table 5. Dobbie, Goldin, and Yang (2018).

The IV comparison concerns defendants whose release depends on the assigned judge. A proposal to release all defendants would extend beyond that margin. OLS and IV can also differ because the treatment effects vary across defendants and the estimators put weight on different comparisons.

4. Disability insurance

Policy question and assignment

Maestas, Mullen, and Strand study the US Social Security Disability Insurance programme. Benefit receipt can reduce the incentive to work, while applicants' health limitations affect both their chance of approval and their ability to work. Their administrative data connect applications, assigned examiners, decisions, and subsequent earnings.

Your turn: Explain the policy question and comparison

Why is employment among rejected applicants informative about the policy question? Why might a comparison of approved and rejected applicants misstate the effect of benefits? Explain the distinction between an initial allowance decision and eventual benefit receipt.

Applicants are assigned within state Disability Determination Services (DDS) offices. Assignment can reflect impairment categories or terminal illness. The authors therefore include office indicators and observed assignment characteristics. Table 2 adds further controls, allowing readers to see how the first stage changes.

i Your turn: Assess examiner assignment

Use the institutional description and Table 2. Why are DDS office controls necessary? Which applicant characteristics can affect assignment? What do changes across columns suggest about the relevant comparisons? Explain why unconditional comparisons across all examiners would be problematic.

TABLE 2— FIRST-STAGE REGRESSIONS: EFFECT OF EXALLOW ON SSDI RECEIPT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2005							
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
2006							
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
2005 and 2006							
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

Notes: Standard errors clustered at examiner level; *t*-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.

Table 2. Maestas, Mullen, and Strand (2013), p. 1814.

First stage, reduced form, and IV

i Your turn: Connect the two relationships

Explain both panels of Figure 4. Identify which panel shows the first stage and which shows the reduced form. What sign would you expect for the IV estimate? Use Table 2, column 7, and Table 5 to calculate the estimate for employment two years after a 2005 decision.

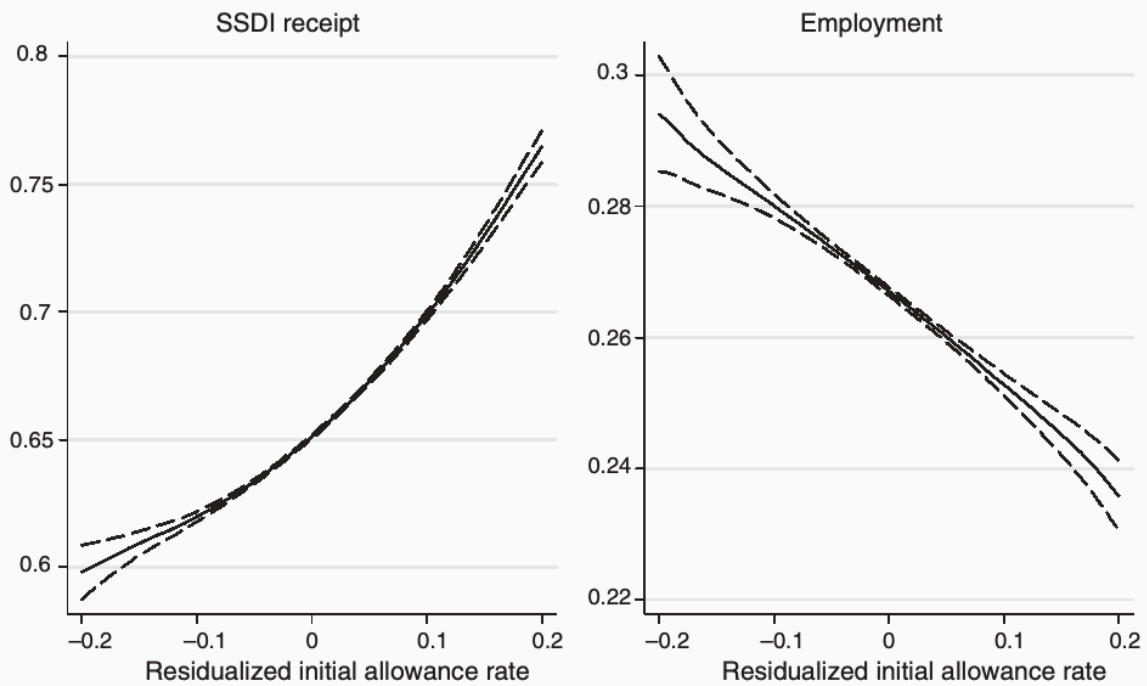


FIGURE 4. SSDI RECEIPT AND LABOR SUPPLY BY INITIAL ALLOWANCE RATE

Notes: Ninety-five percent confidence intervals shown with dashed lines. Employment measured in the second year after the initial decision. Bandwidth is 0.116 for DI and 0.130 for labor force participation.

Source: DIODS data for 2005 and 2006.

Figure 4. Maestas, Mullen, and Strand (2013), p. 1813.

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 > = \$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 > = 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.

Table 5. Maestas, Mullen, and Strand (2013), p. 1820.

The first-stage coefficient for the 2005 cohort is 0.204, and the reduced-form coefficient for employment two years later is −0.057. Their ratio is approximately −0.279, or a **27.9 percentage-point reduction** in employment for the applicants whose benefit receipt is affected by examiner assignment. Small rounding differences arise because the published coefficients are rounded.

Interpreting the effect

For a test of a zero coefficient, $t = \hat{\beta} / \text{SE}(\hat{\beta})$. Rearranging gives $\text{SE}(\hat{\beta}) = |\hat{\beta}| / |t|$. Published rounding makes the recovered standard error approximate.

i Your turn: Interpret Table 4

Focus on Panel A, employment two years after the decision. Compare OLS and IV. Explain the employment definition and the magnitude of the estimated effect. The parentheses contain t -statistics: how would you recover an approximate standard error? Who are the applicants whose treatment status changes with examiner assignment?

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 > = \$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***	−0.279***	−0.361***	−0.227***	−0.345***	−0.158***
		(−322.48)	(−8.64)	(−336.60)	(−6.99)	(−321.81)	(−4.83)
R ²		0.200	0.195	0.218	0.200	0.209	0.171
2) Earn > = 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***	−0.192***	−0.255***	−0.166***	−0.233***	−0.113***
		(−256.29)	(−7.62)	(−264.05)	(−6.70)	(−252.71)	(−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***	−3,781***	−7,715***	−3,007***	−7,221***	−1,716
		(−126.50)	(−3.05)	(−182.51)	(−2.92)	(−176.24)	(−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	
<i>Panel B. 2006 decisions</i>							
1) Earn > = \$1,000/year							
Mean dependent variable	allowed	0.133		0.103			
Mean dependent variable	denied	0.489		0.448			
Coefficient on ALLOW		−0.335***	−0.283***	−0.330***	−0.207***		
		(−297.39)	(−11.25)	(−285.84)	(−8.51)		
R ²		0.194	0.192	0.195	0.177		
2) Earn > = SGA							
Mean dependent variable	allowed	0.042		0.030			
Mean dependent variable	denied	0.267		0.242			
Coefficient on ALLOW		−0.226***	−0.175***	−0.213***	−0.152***		
		(−218.54)	(−8.82)	(−208.98)	(−8.16)		
R ²		0.141	0.136	0.140	0.132		
3) Earnings							
Mean dependent variable	allowed	1,688		1,365			
Mean dependent variable	denied	8,044		7,562			
Coefficient on ALLOW		−6,915***	−4,581***	−6,580***	−3,308***		
		(−84.14)	(−5.93)	(−117.85)	(−4.60)		
R ²		0.139	0.131	0.129	0.112		
Observations		1,123,721		1,094,347			

Notes: Standard errors clustered at examiner level; t -statistics in parentheses; ALLOW denotes actual award decision (rather than initial decision). 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.

Table 4. Maestas, Mullen, and Strand (2013), p. 1819.

Employment here means annual earnings of at least \$1,000. The local effect refers to applicants at the margin of receiving benefits under the observed examiner assignments. A reduction in employment is one consequence to consider alongside the insurance and income protection the programme provides. An overall recommendation also requires evidence about those benefits and the consequences of changing eligibility.

Processing times and exclusion

Your turn: Examine another channel

Use Figure 6. Distinguish initial and final processing times. How might time spent waiting affect later employment? Explain how this bears on interpreting the IV coefficient as the effect of benefit receipt. Read the authors' discussion of the direction of the resulting bias and their sensitivity calculation.

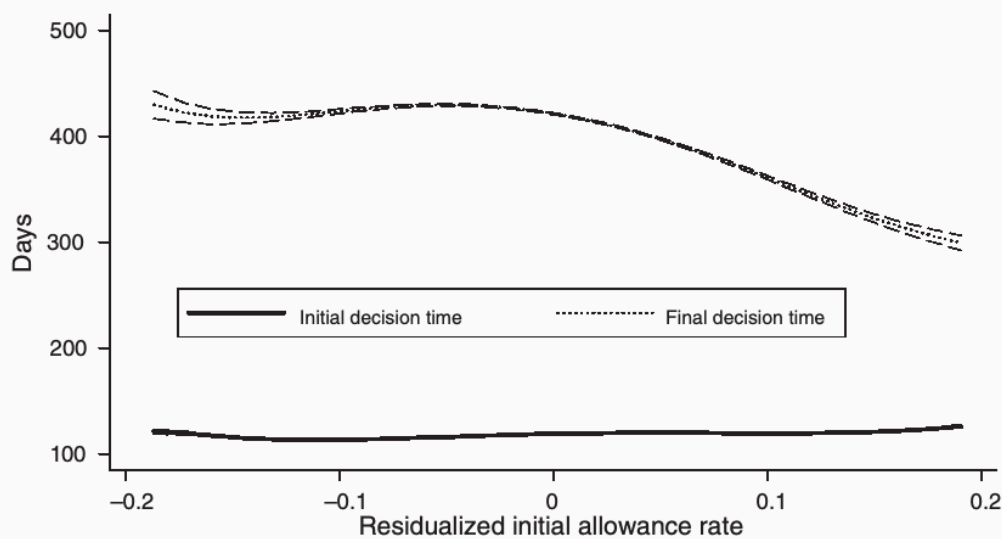


FIGURE 6. INITIAL AND FINAL PROCESSING TIMES BY INITIAL ALLOWANCE RATE

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.

Figure 6. Maestas, Mullen, and Strand (2013), p. 1821.

Appeals can lengthen the application process. If assignment affects time out of work as well as benefit receipt, and that waiting time affects subsequent employment, the exclusion argument must address this pathway. The paper discusses it explicitly. Under the proposed mechanism, stricter examiners can lengthen waiting and reduce subsequent employment through skill depreciation. This offsets some of the negative reduced-form effect of greater leniency operating through increased benefit receipt, making the estimated employment reduction from benefits smaller in magnitude than it would otherwise be.

5. Whose effect, and which policy?

To connect this week's designs to Week 3, temporarily imagine just two officials: a stricter official ($Z_i = 0$) and a more lenient official ($Z_i = 1$). A complier receives treatment under the more lenient official but not under the stricter official:

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

Under the IV assumptions, the binary-instrument comparison identifies

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

With many officials and a leniency measure, the estimates combine comparisons across levels of leniency. Under a common monotone ordering, they can be interpreted as weighted averages of effects for individuals whose treatment changes along those margins. The identity of these individuals depends on the officials and assignment process in the study.

A higher average approval rate alone does not establish a common ordering. For example, examiner A could be more willing than examiner B to approve musculoskeletal claims but less willing to approve mental-health claims. This is why the institutional case for monotonicity deserves discussion.

Your turn: Compare the designs

For each paper, give the strongest institutional argument for independence, the most plausible alternative channel threatening exclusion, and a concrete description of the people at the treatment margin. Then propose a policy change to which the estimate is informative. Explain what further evidence would be needed for a broader reform.

A note on standard errors

Individuals assigned to the same official share a source of variation. The papers allow for dependence within relevant groups when calculating uncertainty: Maestas, Mullen, and Strand cluster by examiner. Dobbie, Goldin, and Yang use two-way clustering by defendant and judge. Heteroskedasticity-consistent standard errors address differences in error variances. Clustered standard errors additionally allow errors to be related within the specified groups. We will return to the implications for inference as we use designs with grouped assignment.

6. After-class exercise

[Exercise instructions](#) · [Download the Quarto notebook](#)

The notebook supplies Python code that generates applicants randomly assigned to caseworkers with different approval thresholds. Each student generates a fresh sample, with no fixed random seed. Construct the leave-one-out instrument, examine the first stage and reduced form, and estimate OLS and IV. Compare your estimates with the known causal effect and explain any differences. This exercise is ungraded.

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.