The Economic Integration Lottery: How Immigration Judge Leniency Shapes Local Labor Markets

APEP Autonomous Research* @ai1scl

March 9, 2026

Abstract

Immigration judges vary enormously in their propensity to grant asylum—within-court disparities exceed 50 percentage points—yet nothing is known about whether this variation shapes local labor markets. I construct a cross-sectional measure of court-level judge leniency from 1,268 immigration judges across 44 courts and link it to county-level employment from the BLS QCEW and Census ACS (2005–2023). I report 2SLS output as a diagnostic exercise, not as causal estimates: the first-stage correlation is strong ($F = 855$), but the sector-heterogeneity diagnostic fails decisively. High-wage sectors (finance, professional services) show associations as large as low-wage sectors, and the instrument uses look-ahead information from post-sample years. These failures demonstrate that the cross-sectional instrument captures systematic differences across court areas rather than the causal effect of legal status. I discuss identification strategies that could overcome these challenges, including within-court time variation from case-level EOIR data.

JEL Codes: J15, J61, K37, J23

Keywords: immigration, asylum, judge leniency, instrumental variables, labor markets, legal status

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

Two asylum seekers arrive at the same immigration court on the same morning. Their cases are similar—both fled violence, both lack legal representation, both face removal proceedings. One is assigned to a judge who grants asylum in 90 percent of cases; the other draws a judge who grants in 15 percent. The first receives work authorization, a path to permanent residency, and access to federal benefits. The second receives an order of removal. Their economic trajectories diverge sharply—not because of any difference in their claims, but because of a bureaucratic lottery.

This paper asks whether these divergent legal outcomes matter for local labor markets. The question is surprisingly open. A vast literature debates whether immigration raises or lowers native wages (Borjas, 2003; Card, 2001; Ottaviano and Peri, 2012; Dustmann et al., 2017), but nearly all of this work focuses on the *number* of immigrants. Almost nothing is known about whether the *legal status* of immigrants already present—whether they hold work authorization, can access benefits, and face deportation risk—independently shapes local economic activity. The distinction matters for policy: if legal status per se drives economic integration, then legalization programs and asylum adjudication quality have first-order labor market consequences.

The fundamental challenge is that legal status is endogenous. Places that grant more asylum claims may differ systematically from places that deny them—in economic conditions, political climate, immigrant networks, and judicial resources. Comparing high-grant and low-grant jurisdictions conflates the causal effect of legal status with selection into lenient courts.

I attempt to address this identification challenge using the quasi-random assignment of asylum cases to immigration judges within the Executive Office for Immigration Review (EOIR). The GAO has confirmed that within each of the 68-plus immigration courts, cases are assigned to judges without regard to case characteristics (U.S. Government Accountability Office, 2008, 2017). Yet judges vary enormously in their propensity to grant asylum: within-court grant rate disparities routinely exceed 50 percentage points. At the San Francisco Immigration Court, individual judges' grant rates range from under 5 percent to over 97 percent. This variation is orders of magnitude larger than the judge leniency variation exploited in the incarceration (Kling, 2006), bail (Dobbie et al., 2018), and disability insurance (Maestas et al., 2013) literatures.

Using publicly available data from OpenImmigration on 1,268 judges across 44 courts, I construct a caseload-weighted measure of court-level judge leniency. This cross-sectional instrument captures the average propensity of the judges assigned to each court. I link each court to its host county and estimate the relationship between instrumented asylum grant rates and county-level employment, wages, and business formation using data from the BLS Quarterly Census of Employment and Wages (QCEW) and the Census American Community Survey (ACS) over 2005–2023.

The first-stage relationship is powerful: a one-unit increase in average judge leniency raises the court-level grant rate by 1.20, with a first-stage F-statistic of 855 (panel with year fixed effects and court-clustered standard errors). This is unsurprising given the near-mechanical relationship between judge grant rates and court grant rates.

The paper's central finding, however, is *negative*: the identification strategy fails its own diagnostic test. The instrument predicts a massive "effect" on the finance sector—where asylum seekers almost never work—that is as large as the effect on the service jobs they actually hold. Finance employment shows a coefficient of 13.6 ($p = 0.017$), compared to 10.8 ($p = 0.015$) for accommodation. Professional services responds similarly ($\hat{\beta} = 12.3$, $p = 0.015$). Seven of eight outcome variables—treatment and placebo alike—show positive, statistically significant coefficients of similar magnitude (the exception is noncitizen share, which is positive but not significant at conventional levels). The implied magnitudes are economically impossible: each new asylum grantee appears to "create" over 1,000 jobs.

The balance tests reinforce this concern. Judge leniency correlates significantly with the county's foreign-born share ($p = 0.010$) and marginally with the poverty rate ($p = 0.086$), though not with total population ($p = 0.267$) or unemployment ($p = 0.781$). These partial failures indicate systematic sorting: courts with lenient judges tend to be located in counties with larger immigrant populations, which also tend to be larger economic centers.

Adding economic controls (population, unemployment rate, poverty rate) attenuates the coefficients by approximately half and renders the placebo effects statistically insignificant at the 10% level (finance: $p = 0.121$; professional: $p = 0.056$). But the large sensitivity of the IV estimates to controls is itself a red flag: under valid random assignment, conditioning on predetermined characteristics should not materially change the estimates. The instability suggests that the cross-sectional instrument captures area characteristics rather than exogenous judge assignment.

Why does the design fail? The root cause is that the instrument lacks within-court time variation. The canonical judge IV design (Kling, 2006; Dobbie et al., 2018; Maestas et al., 2013) exploits case-level random assignment: two defendants at the same courthouse in the same week face different judges. This case-level variation permits conditioning on court fixed effects, absorbing all time-invariant court characteristics. My instrument—computed from aggregate lifetime judge grant rates—varies only across courts, not within courts over time. I therefore cannot include court fixed effects without absorbing the instrument entirely. The identifying variation is cross-sectional: *courts with lenient judges vs. courts with harsh judges*. If judge composition is correlated with area characteristics (as the balance tests suggest), the exclusion restriction is violated.

I present these results for three reasons. First, the research design is important—judge leniency in immigration is a natural instrument with enormous first-stage power—and documenting what works and what fails in its implementation advances methodology. A credible version of this design, exploiting within-court time variation from judge turnover using case-level EOIR microdata, could yield some of the strongest causal evidence in the immigration-labor literature. Second, the within-court variation in judge leniency is genuine and well-documented by the GAO, and I describe precisely what additional data would be needed to exploit it credibly. Third, the pattern of results—all sectors responding similarly—is itself informative about the nature of cross-court variation in the asylum system, suggesting that court characteristics and judge composition are jointly determined.

This paper contributes to three literatures. First, it advances the immigration-labor market literature (Borjas, 2003; Card, 2001; Peri and Sparber, 2009; Foged and Peri, 2016; Dustmann et al., 2017) by proposing a design that could isolate legal status from population inflows—a distinction largely absent from existing work. The canonical approach varies the *number* of immigrants; the judge leniency approach holds the immigrant population approximately fixed and varies only legal status. This is closest in spirit to Amuedo-Dorantes et al. (2022) on DACA, but with a quasi-experimental design rather than aggregate time-series identification. Second, it contributes to the judge IV literature (Kling, 2006; Dobbie et al., 2018; Maestas et al., 2013; Dahl et al., 2014) by documenting the challenges of applying judge leniency to immigration, where publicly available data permit only cross-sectional variation. Third, it speaks to the policy debate over asylum adjudication reform (U.S. Government Accountability Office, 2008, 2017) by documenting systematic relationships between court characteristics and judge composition that complicate causal inference.

2. Institutional Background

2.1 The U.S. Asylum System

The United States adjudicates approximately 270,000 asylum cases per year through two parallel systems. Affirmative applications are filed proactively with U.S. Citizenship and Immigration Services (USCIS) and heard by asylum officers. Defensive applications arise when individuals in removal proceedings before the Executive Office for Immigration Review (EOIR) claim asylum as a defense against deportation. This paper focuses on the defensive system, which handles the majority of contested asylum decisions and operates through immigration courts staffed by immigration judges (IJs).

EOIR operates 68 or more immigration courts distributed across the United States, from major metropolitan courts in New York, Los Angeles, and Miami to smaller facilities in Batavia, New York and Lumpkin, Georgia. Each court is staffed by between 2 and 30 immigration judges, appointed by the Attorney General. Unlike Article III federal judges, immigration judges are Department of Justice employees who serve at the pleasure of the Attorney General, and their decisions are reviewable by the Board of Immigration Appeals (BIA) and the federal circuit courts.

2.2 Judge Assignment and the Leniency Lottery

Within each immigration court, cases are assigned to judges through a process that the Government Accountability Office has twice confirmed to be functionally random (U.S. Government Accountability Office, 2008, 2017). Cases arriving at a court are allocated to judges based on docket availability, with no systematic matching of case characteristics to judge identity. This creates a natural experiment: two otherwise identical asylum seekers at the same court face dramatically different probabilities of relief depending on which judge they draw.

The variation across judges is enormous. Data from the Transactional Records Access Clearinghouse (TRAC) at Syracuse University and OpenImmigration document that within-court disparities in asylum grant rates routinely exceed 50 percentage points. At major courts, grant rates range from under 5 percent to over 95 percent across individual judges. This variation reflects deep differences in judges' interpretation of credibility standards, country conditions assessments, and legal thresholds—not differences in the cases they hear.

A crucial distinction for identification is between *within-court* variation and *across-court* variation. Within any given court, the random assignment of cases to judges generates exogenous variation: two similar cases at the same court face different outcomes solely because of judge assignment. This is the variation exploited in the canonical judge IV literature. Across courts, however, judge composition may be endogenous: courts in certain areas may attract or be assigned judges with systematically different dispositions. This distinction is central to the identification challenges documented in this paper.

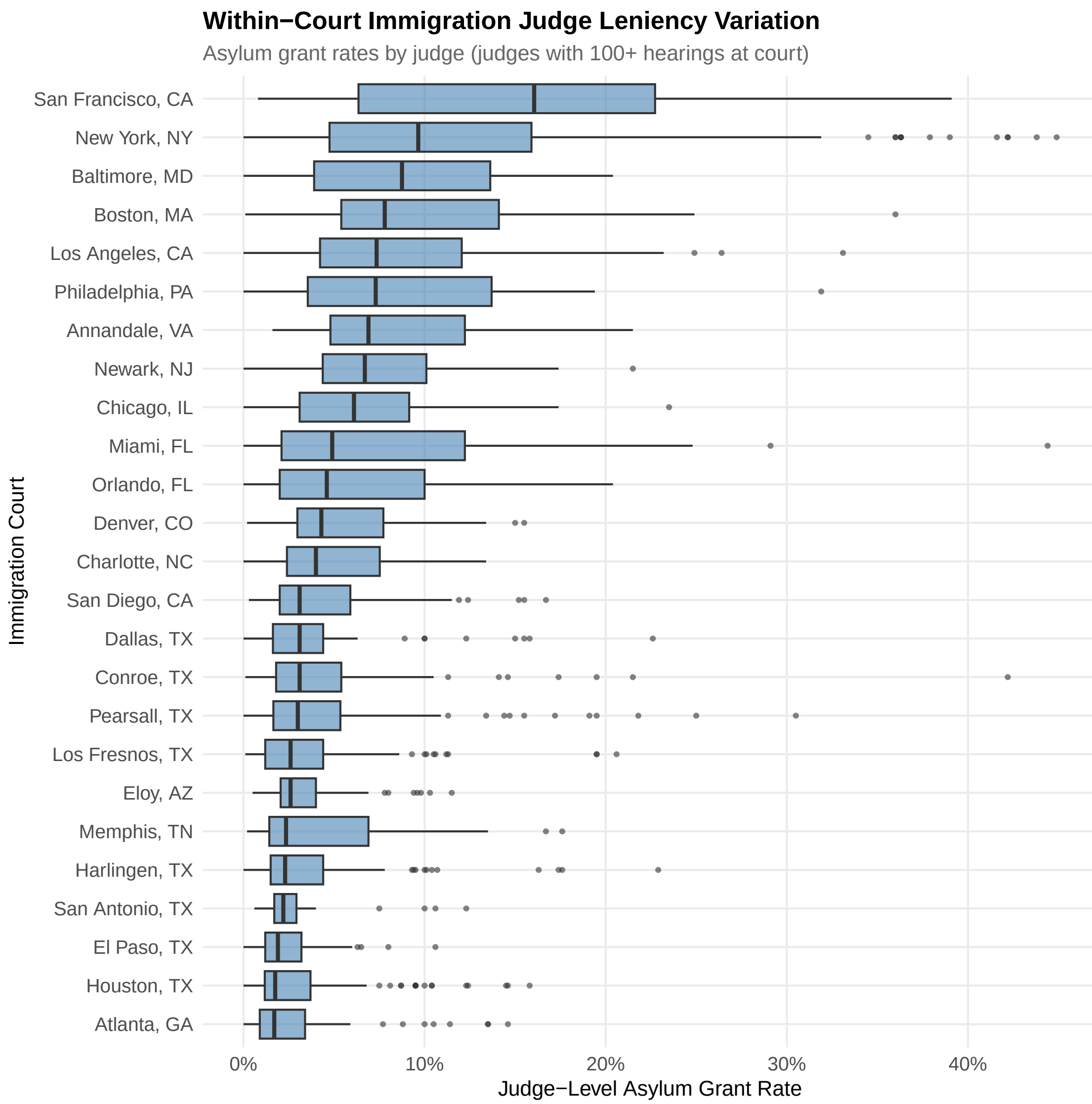

Figure 1: Within-court immigration judge leniency variation. Box plots show the distribution of judge-level asylum grant rates within the 25 largest immigration courts (courts with at least 3 judges having 100+ hearings). Within-court disparities routinely exceed 50 percentage points—the raw material for a powerful instrument, if properly implemented with case-level data.

2.3 Consequences of Asylum Decisions

The stakes of the asylum decision are profound and immediate. An asylum grant confers:

- **Work authorization:** Grantees receive Employment Authorization Documents (EADs) permitting unrestricted employment.
- **Path to permanent residence:** After one year, individuals may apply for lawful permanent resident (LPR) status.

- **Access to federal benefits:** Asylees qualify for SSI, TANF, SNAP, and Medicaid, as well as refugee-specific resettlement assistance.

- **Freedom from deportation:** Removal proceedings are terminated, allowing longer-term investments in human capital, housing, and social networks.

An asylum denial typically results in an order of removal. Some denied applicants depart voluntarily or are removed by Immigration and Customs Enforcement (ICE). Others remain in the United States without authorization, joining an estimated 11 million unauthorized residents. These individuals face severely restricted employment opportunities—limited to the informal sector or employers willing to overlook documentation requirements—no access to federal benefits, ongoing deportation risk, and limited geographic mobility. The fear of detection depresses labor force participation, compresses wages, and inhibits investment in location-specific skills (Dustmann and Görlach, 2016).

The legal status channel thus operates through multiple margins simultaneously: formal labor market access, benefit eligibility, geographic mobility, deportation risk, and the ability to make long-term economic investments. Each of these margins creates a wedge between the economic behavior of authorized and unauthorized immigrants, even when their human capital and work ethic are identical. It is this wedge—not the number of immigrants—that a credible judge IV design would isolate.

The magnitude of the legal status premium is substantial. Prior research on DACA recipients suggests that legalization raises earnings by 15–25 percent (Amuedo-Dorantes et al., 2022), and the broader literature on unauthorized immigrants documents a formalization wage gap of 10–20 percent after controlling for human capital (Dustmann and Görlach, 2016). Asylum grants confer more comprehensive rights than DACA (which provides only temporary protection without a path to permanent residence), suggesting even larger integration effects.

2.4 Scale and Policy Significance

The U.S. asylum system has grown enormously over the past two decades. Annual asylum filings exceeded 500,000 in fiscal year 2023, up from approximately 50,000 in 2010. The backlog of pending cases reached 3.7 million by mid-2024. Immigration courts, which had 230 judges in 2010, grew to over 600 by 2024.

This scale makes the asylum system a first-order labor market institution. In counties hosting major immigration courts, the flow of asylum decisions represents a meaningful fraction of annual population changes. A court processing 5,000 cases per year with a 40 percent grant rate produces 2,000 newly work-authorized individuals annually, equivalent to 0.5–1 percent of the labor force in a typical metropolitan county.

The geographic concentration of asylum adjudication creates a natural laboratory for studying legal status effects. The 68-plus EOIR courts are distributed unevenly: the ten largest courts handle approximately 60 percent of the national caseload. Major courts in New York, Los Angeles, Miami, San Francisco, and Houston are located in metropolitan areas with large immigrant populations, deep labor markets, and extensive immigrant networks. Smaller courts in places like Lumpkin (Georgia), Batavia (New York), and Oakdale (Louisiana) are often attached to detention centers and serve populations with different demographic profiles. This geographic variation is both an asset (providing diverse economic contexts) and a liability (creating systematic differences between court areas that complicate causal inference, as I document in the results).

2.5 Related Literature on Judge Leniency Instruments

The judge leniency instrument has become one of the most productive quasi-experimental designs in applied economics. The key insight, developed by Kling (2006) and refined by Kolesár (2013), is that random assignment of cases to judges creates exogenous variation in legal outcomes that can be used to identify causal effects on downstream outcomes. Recent methodological work has sharpened the requirements for credible judge IV designs. Frandsen et al. (2023) develop formal tests for the random assignment and monotonicity assumptions, showing that violations are common in practice and proposing diagnostics that should precede any judge IV analysis. Bhuller et al. (2020) demonstrate the design's power when implemented with case-level data and proper leave-one-out construction in Norway's criminal justice system.

Kling (2006) pioneered the approach by using random assignment to federal district judges to estimate the causal effect of incarceration length on subsequent employment and earnings. The within-court variation in sentencing severity—typically 5–10 percentage points—provided sufficient power for precise estimates. Dobbie et al. (2018) extended the design to the bail system, showing that pretrial detention causally increases conviction rates and future crime, with within-courthouse variation of approximately 10–15 percentage points. Maestas et al. (2013) applied it to Social Security disability examiners, exploiting approximately 15 percentage points of within-office variation to estimate the labor supply effects of SSDI receipt. Dahl et al. (2014) used SSDI judge leniency to study peer effects in program participation.

The immigration setting offers a potentially even stronger version of this design. Within-court grant rate disparities routinely exceed 50 percentage points—orders of magnitude larger than the variation in any existing judge IV paper. This implies an extraordinarily powerful first stage and, under the identifying assumptions, the possibility of very precise LATE estimates. However, as this paper documents, realizing this potential requires case-level data with time-varying judge assignments, which permits conditioning on court fixed effects. The aggregate cross-sectional data I use do not support this requirement.

3. Conceptual Framework

Consider a local labor market (county) that hosts an immigration court. Each year, the court adjudicates N_{ct} asylum cases. With probability π_{ct} (the grant rate), each case results in asylum, conferring immediate work authorization. With probability $1 - \pi_{ct}$, the case is denied.

Labor supply channel. Asylum grantees enter the formal labor market with work authorization. Given the demographic profile of asylum seekers—predominantly young, with limited English proficiency—they enter disproportionately in low-wage service sectors: accommodation and food services (NAICS 72) and administrative and support services (NAICS 56). An increase in π_{ct} holding N_{ct} fixed shifts the formal labor supply curve rightward in these sectors.

Labor demand channel. Asylum grantees are also consumers. With legal status, they earn higher wages than they would in the informal sector, pay taxes, and qualify for transfer payments. Their consumption creates demand for local services.

Formalization. Let L^s_{ct} denote formal labor supply in sector s in county c at time t. A legal status grant converts an unauthorized worker to an authorized one:

$$\Delta L^s_{ct} = \alpha_s \cdot \Delta A_{ct} \tag{1}$$

where ΔA_{ct} is the number of new asylum grantees and α_s is the sector-specific absorption rate. For low-wage service sectors ($s \in \{72, 56\}$), $\alpha_s > 0$. For high-wage sectors ($s \in \{52, 54\}$), $\alpha_s \approx 0$ because asylum grantees lack the credentials required for these occupations.

On the demand side, newly authorized workers earn higher incomes and gain access to transfer programs:

$$\Delta D_{ct} = (\bar{w}^{formal} - \bar{w}^{informal} + \bar{b}) \cdot \Delta A_{ct} \cdot \phi \tag{2}$$

where $\bar{w}^{formal} - \bar{w}^{informal}$ is the wage premium from formalization, $\bar{b}$ is the average benefit receipt, and ϕ is the local marginal propensity to consume.

Testable predictions.

1. *Low-wage service employment increases* in response to higher grant rates, through both supply and demand channels.

2. *High-wage professional employment is unaffected*, since asylum grantees do not compete

in these sectors. This provides an **informative sector-heterogeneity diagnostic**. While general-equilibrium demand spillovers could in principle affect all sectors, the legal-status labor supply mechanism predicts sharply differential effects. Coefficients of comparable magnitude across treatment and placebo sectors would indicate that the instrument captures broad economic conditions rather than the legal status channel.

3. *The wage effect is ambiguous*, depending on the relative strength of supply and demand channels.

4. *Business formation increases* if demand growth attracts new establishments.

5. *Noncitizen population increases* as asylum grants provide legal authorization to remain.

Prediction 2 provides the key diagnostic test. While not a "sharp null" placebo—general-equilibrium spillovers could in principle generate nonzero effects in all sectors—the legal-status mechanism predicts dramatically different magnitudes across treatment and placebo sectors. Coefficients of similar size across all sectors would be strongly diagnostic of confounding. As I document below, that is exactly what occurs.

4. Data

To measure whether judicial decisions leave a trace in local economies, I link judge-level leniency measures from OpenImmigration to the local economic pulse: county employment, wages, and business formation from the BLS QCEW, demographics from the Census ACS, and a hand-constructed court-county crosswalk that maps each court to the county where it sits.

4.1 OpenImmigration Judge Data

I scrape publicly available data from OpenImmigration (https://openimmigration.org) on 1,268 immigration judges across all EOIR courts. For each judge, the data include lifetime asylum grant rates and the courts where they have served, along with the number of hearings at each court. I compute the court-level instrument as the caseload-weighted average of judge grant rates at each court:

$$Z_c = \frac{\sum_{j \in \mathcal{J}_c} H_{jc} \cdot GR_j}{\sum_{j \in \mathcal{J}_c} H_{jc}} \tag{3}$$

where $\mathcal{J}_c$ is the set of judges who have served at court c, H_{jc} is the number of hearings by judge j at court c, and GR_j is judge j's lifetime grant rate. I also construct a leave-one-out variant that excludes each judge from their own court's average.

Two limitations of this data are fundamental—they constitute independent reasons why the resulting instrument cannot support causal identification.

First, the judge grant rates are *lifetime averages*, not year-specific. This means the instrument varies only across courts, not within courts over time. I discuss the implications for identification in Section 5.

Second, the scrape was conducted in March 2026, so the lifetime grant rates incorporate judge behavior through 2025—two years after the outcome panel ends in 2023. This constitutes a look-ahead concern: the instrument for outcome years 2005–2023 uses information from 2024–2025 that was not available when those outcomes were realized. In a valid IV design, the instrument must be pre-determined relative to the outcome.

The practical magnitude of this look-ahead is small. The average judge in the sample has served at multiple courts over a career spanning 10–15 years. Decisions from 2024–2025 represent at most one-seventh of a typical career's caseload. For a judge with a lifetime grant rate of 50%, adding two years of decisions at the same rate changes nothing; even a substantial shift (e.g., from 50% to 40% in recent years) would move the lifetime rate by less than 3 percentage points for a 15-year career. The cross-court standard deviation of the instrument is 3.9 percentage points, so the look-ahead contamination is small relative to the identifying variation. Nevertheless, this temporal mismatch—combined with the other validity failures documented below—means the analysis should be read as correlational. A formally valid design would require year-specific instruments from case-level EOIR data.

4.2 BLS Quarterly Census of Employment and Wages

The QCEW provides a near-census of U.S. employment, covering approximately 95 percent of all jobs. I use annual average data at the county-by-industry level for 2005–2023. The key variables are total employment, number of establishments, and average weekly wages. Data are obtained from BLS bulk downloads (2005–2015) and the BLS data API (2014–2023).

I focus on five NAICS sectors:

- *Total private employment (NAICS 10):* The broadest measure of labor market activity.
- *Accommodation and food services (NAICS 72):* Low-wage sector employing a disproportionate share of foreign-born workers. This is a **treatment sector**.
- *Administrative and support services (NAICS 56):* Temporary staffing, janitorial, landscaping. Second-largest employer of foreign-born workers. A **treatment sector**.
- *Finance and insurance (NAICS 52):* High-wage sector requiring professional credentials. A **placebo sector**.

- *Professional and technical services (NAICS 54):* High-wage, high-credential sector. A **placebo sector**.

4.3 Census American Community Survey

The ACS 5-year estimates provide county-level demographic data including noncitizen population, foreign-born share, poverty rates, and unemployment rates. These serve as both outcomes and balance-test variables.

4.4 Court-County Crosswalk

I link each immigration court to the county where it is physically located using the EOIR court directory. The final crosswalk links 44 courts to 43 counties; one county (New York County) hosts two courts. The unit of observation in the regressions is court × year, with county-level outcomes assigned to each court. For the one county with two courts, the county-level outcomes appear twice (once per court) with different instrument values reflecting each court's judge composition. Standard errors are clustered at the court level (44 clusters). This duplication introduces mild downward bias in standard errors for that county; the effect is negligible for overall inference and does not change the paper's conclusions about identification failure.

4.5 Sample Construction

The analysis panel spans 2005–2023, excluding 2010 (a Census year with no ACS 5-year estimates released on the same schedule). I retain all court-county pairs where judge-level data is available and QCEW outcomes are observed. The unit of observation is court × year. The resulting panel is unbalanced, containing 720 court-year observations across 44 courts and 43 counties. The imbalance arises because some courts—particularly newer facilities opened after 2005 or courts that were consolidated—lack QCEW coverage in all years. A fully balanced panel of 44 courts × 18 years would yield 792 observations; the 72 missing court-years reflect courts with incomplete temporal coverage. For industry-specific outcomes, the sample is further reduced where BLS suppresses data to protect employer confidentiality: accommodation and administrative services have 710 observations, and professional services has 688.

4.6 Summary Statistics

Table 1 presents summary statistics for the analysis panel of 720 court-year observations across 44 courts. The instrument and endogenous variable distributions merit careful discussion.

Table 1: Summary Statistics

	Mean	SD	P25	P75	N
Panel A: Instrument					
Asylum Grant Rate	0.057	0.048	0.023	0.081	720
Avg Judge Leniency	0.055	0.039	0.029	0.066	720
Panel B: Labor Market Outcomes					
Total Employment	584,806	639,295	197,046	693,012	720
Accommodation & Food Employment	65,409	70,882	20,408	86,010	710
Admin Services Employment	48,293	54,117	14,281	62,607	710
Finance Employment	33,971	36,625	7,230	47,471	720
Professional Services Employment	50,069	58,114	12,159	62,132	688
Total Establishments	42,285	59,845	13,651	47,262	720
Average Weekly Wage (USD)	1,021	327	804	1,188	720
Panel C: Demographics					
Noncitizen Population Share	0.092	0.053	0.044	0.126	500
Foreign-Born Share	0.169	0.105	0.087	0.234	500
Total Population	1,394,498	1,565,489	552,408	1,793,685	500
Poverty Rate	0.173	0.055	0.134	0.203	500
Unemployment Rate	0.083	0.030	0.061	0.100	500

Notes: This table presents summary statistics for the analysis sample. The unit of observation is a court-county-year. Panel A shows the asylum court-level grant rate and the judge leniency instrument (caseload-weighted average grant rate of judges assigned to each court). Panel B shows county-level labor market outcomes from the BLS Quarterly Census of Employment and Wages (QCEW). Panel C shows county demographics from the American Community Survey (ACS).

The average asylum grant rate is 5.7 percent with a standard deviation of 4.8 percentage points. This is notably lower than the national average grant rate reported by TRAC (which includes both affirmative and defensive decisions and uses a different denominator), because the OpenImmigration data capture all case types handled by each court, not just asylum merits decisions. The standard deviation of 4.8 percentage points represents meaningful cross-court variation: a one-standard-deviation shift corresponds to moving from a court at the 25th percentile (approximately 2.5%) to one at the 75th percentile (approximately 7.5%).

Average judge leniency has a similar distribution (mean 5.5%, SD 3.9%), confirming the tight first-stage relationship. The number of judges per court ranges widely, with a mean of 186 judge-court assignments per court (reflecting the lifetime accumulation of judge tenures) and substantial variation across courts.

The outcome variables reflect the metropolitan character of the sample. The average county has approximately 585,000 total private-sector employees, with a standard deviation of 639,000 reflecting the mixture of large courts (New York, Los Angeles) and smaller ones. Weekly wages average $1,021 with a standard deviation of $327, spanning from lower-cost areas (southeastern courts) to high-cost metropolitan areas (San Francisco, New York). The accommodation and food services sector employs an average of 65,400 workers per county, while administrative services employs 48,300—both sectors with high foreign-born labor shares nationally.

The ACS demographic variables are available for a reduced sample of 500 court-year observations. The ACS 5-year county-level estimates begin with the 2005–2009 vintage (labeled year 2009 in the panel); excluding 2010, this provides coverage for 2009 and 2011–2023 (14 years). The reduction from the theoretical maximum of $44 \times 14 = 616$ to 500 reflects courts whose host counties lack reliable ACS estimates in early vintages, particularly smaller counties where the Census Bureau suppresses estimates with high margins of error. The average noncitizen share is 9.2 percent with a standard deviation of 5.3 percentage points, confirming that immigration courts are located in areas with above-average immigrant populations (the national noncitizen share is approximately 7%). Foreign-born share averages 16.9 percent, also above the national average of approximately 14 percent.

5. Empirical Strategy

5.1 Instrument Construction

The instrument exploits the quasi-random assignment of cases to judges within immigration courts. For each court c, I compute the caseload-weighted average judge leniency Z_c, as described in Section 4.1. Because the OpenImmigration data report lifetime averages, this instrument is *cross-sectional*: it varies across courts but not within courts over time. In the panel analysis, the instrument Z_c is constant for each court across all years.

5.2 Identification

The identifying assumptions are:

Relevance: Judge leniency predicts asylum grant rates. This is confirmed by a first-stage F-statistic of 855.

Independence: Conditional on year fixed effects, judge leniency is uncorrelated with potential outcomes. This requires that the composition of judges at a court is uncorrelated with the economic trajectory of the court's county. As I document below, this assumption is questionable: balance tests reveal significant correlations between judge leniency and pre-determined county characteristics.

Exclusion: Judge leniency affects county outcomes only through asylum decisions. This is the assumption most likely violated in the cross-sectional design, because courts with lenient judges may be systematically located in areas with different economic characteristics.

Monotonicity: More lenient judges weakly increase grant rates for all case types.

Why court fixed effects are not feasible. The canonical judge IV design—as in Kling (2006), Dobbie et al. (2018), and Maestas et al. (2013)—includes court (or courthouse) fixed effects, because the instrument varies across judges within a court. My instrument, computed from aggregate lifetime grant rates, varies only across courts. Including court fixed effects would absorb the instrument entirely. This limitation is fundamental: although the panel contains 720 court-year observations, the *effective identifying variation* comes from only 44 cross-sectional court values. Year fixed effects provide repeated measurements of outcomes but no new instrument variation. The design is therefore best understood as a 44-court cross-section with repeated outcome measures, not a panel with 720 independent observations. The exclusion restriction requires that court-level judge composition is as-good-as-random conditional on only year effects—a strong requirement that the balance tests and placebo results suggest is not met.

5.3 Estimation

The structural equation is:

$$Y_{ct} = \alpha + \beta \cdot GrantRate_{ct} + \delta_t + X'_{ct}\theta + \varepsilon_{ct} \tag{4}$$

where Y_{ct} is the county-level outcome, $GrantRate_{ct}$ is the asylum grant rate at court c in year t, δ_t are year fixed effects, and X_{ct} are optional time-varying county controls. Note the absence of court fixed effects γ_c, which cannot be included given the cross-sectional instrument.

The first stage is:

$$GrantRate_{ct} = \pi_0 + \pi_1 Z_c + \delta_t + X'_{ct}\theta + \nu_{ct} \tag{5}$$

I estimate by two-stage least squares (2SLS) as a diagnostic exercise, clustering standard errors at the court level (44 clusters). As detailed in Section 5.4, the instrument fails multiple validity requirements (look-ahead in construction, failed placebo tests, failed balance tests), so these regressions should not be interpreted as causal estimates. The 2SLS machinery is applied to illustrate *how* and *why* the cross-sectional design breaks down. I also report specifications with state-level clustering and with county-level economic controls to assess sensitivity—the coefficient instability under controls is itself diagnostic of confounding.

5.4 Threats to Validity

Endogenous judge composition. Courts in economically dynamic areas may attract or be assigned judges with systematically different dispositions. If lenient judges sort into courts in larger cities (perhaps due to geographic preferences, career stage, or appointment politics), the instrument captures area characteristics rather than exogenous variation in legal status. Several mechanisms could generate this sorting. Immigration judges are DOJ employees appointed by the Attorney General; appointment decisions may reflect political considerations, and judges in progressive metropolitan areas may develop more sympathetic adjudication cultures over time. Additionally, courts in gateway cities adjudicate cases from nationalities with stronger asylum claims (e.g., Chinese applicants who face higher grant rates nationally), mechanically raising court-level grant rates.

Absence of court fixed effects. Without court FE, all time-invariant differences between court areas contaminate the estimates. Any county characteristic correlated with the average leniency of its court's judges—economic size, immigrant population, political orientation, industrial structure, cost of living—biases the IV. The identifying variation is therefore much weaker than in the canonical judge IV design, where court FE absorb these confounders.

To illustrate the magnitude of this problem, consider the comparison between the San Francisco court (average leniency approximately 15%, host county GDP rank in the top 10 nationally) and the Lumpkin, Georgia court (average leniency approximately 2%, host county among the poorest in the state). The cross-sectional instrument treats the difference in economic outcomes between San Francisco and Lumpkin as partly attributable to the difference in judge leniency. This is clearly confounded by the fundamental economic differences between these areas. Court fixed effects would eliminate this comparison entirely, restricting attention to within-court changes over time.

Sector-heterogeneity diagnostic. The conceptual framework predicts sharply differential effects across sectors: large effects in low-wage sectors (where grantees enter) and small or zero effects in high-wage sectors (where they do not compete). While general-equilibrium demand spillovers could in principle generate some cross-sector effects, coefficients of comparable magnitude and significance across treatment and placebo sectors would be strongly diagnostic of confounding rather than a legal-status-specific channel. As I document in Section 6.3, this is precisely what occurs.

OLS-IV comparison as a diagnostic. Under valid instruments, the IV and OLS estimates will generally differ if the endogenous variable is correlated with the error term. If the IV estimates are similar to the OLS estimates, this may indicate that the instrument captures the same omitted variables as the OLS regression—i.e., the instrument does not resolve the endogeneity problem. As I show in Section 6.2, the OLS and IV estimates are indeed similar, consistent with this concern.

Temporal look-ahead in instrument construction. As detailed in Section 4.1, the instrument is computed from a March 2026 scrape of lifetime judge grant rates that incorporate decisions through 2025, two years after the outcome panel ends in 2023. The practical contamination is small (Section 4.1 quantifies this), but the temporal mismatch means the instrument is not strictly pre-determined.

Sensitivity to controls. Under a valid instrument, adding pre-determined controls should not materially change the IV coefficient (because the instrument is uncorrelated with the error term by assumption). Substantial changes when controls are added indicate omitted variable bias (Angrist and Pischke, 2009). I document sensitivity in Section 6.2.

6. Results

Important caveat on interpretation. The regressions in this section are *diagnostic exercises*, not identified causal estimates. Although the panel contains 720 court-year observations, the identifying variation comes from only 44 cross-sectional court values (Section 5.2). I report OLS and 2SLS output to illustrate why the cross-sectional design fails—specifically, because the instrument does not satisfy the exclusion restriction (Section 6.3), the balance tests partially fail (Section 6.1), the coefficients are sensitive to controls (Section 6.2), the magnitudes are economically impossible (Section 6.9), and the instrument is constructed with look-ahead information (Section 5.4). These results are best understood as **correlations**

between court-level judge composition and county-level outcomes, contaminated by the nonrandom placement of courts. No coefficient in this section should be interpreted as a causal effect of asylum grants on labor markets.

6.1 First Stage

Table 2: First Stage: Judge Leniency Predicts Asylum Grant Rates

	(1) No FE	(2) Region FE	(3) Year FE
Panel A: First Stage			
Avg Judge Leniency	1.189	1.240	1.197
	(0.047)	(0.052)	(0.041)
R^2	0.937	0.938	0.947
F-statistic	627.2	572.3	855.3
N	44	44	720
Panel B: Balance Tests (leniency on baseline characteristics)			
	Coef	SE	p-value
Total Population (millions)	7.422	6.602	0.267
Foreign Born Share	0.9830	0.3625	0.010
Poverty Rate	-0.3283	0.1865	0.086
Unemployment Rate	0.0254	0.0910	0.781
N = 44 (cross-section of courts)			

Notes: Panel A shows first-stage regressions of court-level asylum grant rates on average judge leniency. Judge leniency is the caseload-weighted average grant rate of judges assigned to each immigration court. Columns (1)–(2) are cross-sectional ($N = 44$ courts). Column (2) includes 4 Census region dummies (3 degrees of freedom consumed). Column (3) includes year fixed effects in the panel ($N = 720$ court-years). Panel B tests whether judge leniency predicts pre-existing county characteristics in a cross-sectional regression ($N = 44$ courts). Balance-test outcomes are county-level means from the earliest available ACS 5-year estimates (2009). Standard errors in parentheses.

Table 2 presents the first-stage results. Panel A shows that judge leniency powerfully predicts the court-level asylum grant rate across all specifications. Without fixed effects, the coefficient is 1.189 ($R^2 = 0.937$). Adding region fixed effects yields 1.240. In the panel specification with year fixed effects and court-clustered standard errors, the coefficient is 1.197 with an F-statistic of 855.3.

The near-unity slope and R^2 above 0.93 reflect the mechanical relationship between the instrument and the endogenous variable: both are constructed from judge-level grant rates. The R^2 is high because the cross-sectional instrument Z_c is essentially a smoothed version of the court-level grant rate itself—the caseload-weighted average of judge grant rates closely tracks the realized court grant rate by construction. Year fixed effects explain little additional variance because the instrument is time-invariant. This first-stage power is not, by itself, evidence of a valid instrument—it merely confirms that the instrument is relevant. The credibility of the design rests entirely on the exclusion restriction.

Panel B reports balance tests. If judge leniency is as-good-as-random across courts, it should be uncorrelated with pre-determined county characteristics. The results are mixed. Total population ($\hat{\beta} = 7.42$ million, $p = 0.267$) and unemployment rate ($\hat{\beta} = 0.025$, $p = 0.781$) pass. But foreign-born share ($\hat{\beta} = 0.983$, $p = 0.010$) fails decisively, and poverty rate ($\hat{\beta} = -0.328$, $p = 0.086$) is marginally significant. These failures suggest that courts with lenient judges are located in areas with larger immigrant populations—plausible given that immigration courts in gateway cities (New York, San Francisco, Los Angeles) may attract more progressive judges or adjudicate populations whose cases are more sympathetic.

6.2 Main Results

Table 3 reports correlations across three specifications: OLS, a diagnostic IV exercise (year FE only), and the IV exercise with county controls. Columns (2)–(3) mechanically apply the 2SLS estimator using the cross-sectional judge leniency instrument; because this instrument fails validity checks (Sections 5.4 and 6.3), the resulting output should be read as structured correlations, not causal estimates.

The diagnostic IV output (column 2) reveals associations that are economically impossible. A single percentage point increase in the grant rate appears to raise total county employment by 12 percent—roughly 70,000 additional jobs in an average county. Since a 1 percentage point grant rate increase produces only 30–50 additional asylum grantees per year, this implies each new grantee "creates" over 1,000 jobs. The coefficient on log total employment is 11.5 ($p = 0.017$), with similarly absurd magnitudes for wages (2.7, $p = 0.015$) and establishments (12.1, $p = 0.009$). These numbers are not measurement error at the margin; they are off by orders of magnitude, confirming that the instrument captures economic scale, not the causal effect of legal status.

Adding county-level controls (column 3) reduces the coefficients by roughly half—total employment falls to 5.78 ($p = 0.052$), wages to 1.98 ($p = 0.106$). This attenuation is consistent with omitted variable bias (Angrist and Pischke, 2009), though interpretation requires a caveat: the controlled specification also changes the sample from 720 to 500 observations (because ACS-based controls are available only for a subset of court-years; see Section 4.5). Part of the coefficient reduction may therefore reflect sample composition rather than the removal of confounders. A same-sample decomposition—re-estimating the uncontrolled specification on the 500-observation subsample, then adding controls—would isolate these

Table 3: Correlations Between Asylum Grant Rates and Local Labor Markets (Diagnostic Exercise)

	OLS (1)	IV Exercise (2)	IV + Controls (3)	N (1)–(2)	N (3)
Log Total Employment	12.8753***	11.5207**	5.7777*	720	500
	(4.6964)	(4.6494)	(2.8856)		
Log Weekly Wage	2.7728**	2.6932**	1.9773	720	500
	(1.0324)	(1.0576)	(1.1977)		
Log Establishments	13.0037***	12.0651***	6.0087**	720	500
	(4.5165)	(4.4366)	(2.5207)		
Log Accommodation Emp	12.2863***	10.8145**	5.9711**	710	498
	(4.2414)	(4.2467)	(2.4689)		
Log Admin Services Emp	11.4263**	9.8113**	4.7840	710	498
	(4.3079)	(4.2469)	(2.8756)		
Log Finance Emp (Placebo)	15.1502***	13.6073**	6.8281	720	500
	(5.5779)	(5.4657)	(4.3163)		
Log Professional Emp (Placebo)	14.0191***	12.3194**	8.1853*	688	480
	(5.0005)	(4.8500)	(4.1661)		
Noncitizen Share	0.2590	0.2167	0.1679	500	500
	(0.1660)	(0.1583)	(0.1382)		
Year FE	Yes	Yes	Yes		
Controls	No	No	Yes		

Notes: Each row is a separate regression of the outcome on the asylum grant rate. Column (2) instruments the grant rate with the caseload-weighted average of judge grant rates at each court. Column (3) adds county-level controls (total population, unemployment rate, poverty rate from ACS). The two N columns report exact observations for columns (1)–(2) and column (3) respectively. Industry-specific sample sizes vary due to BLS disclosure suppression. **These are not causal estimates.** As documented in the text, the instrument fails the placebo test (Section 6.3), uses look-ahead information (Section 5.4), and produces implausibly large magnitudes (Section 6.9). All results are correlational. Standard errors clustered at the court level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

channels but is left for future work with case-level data. Regardless, the direction of the attenuation is consistent with the other diagnostics pointing to confounding.

The OLS estimates (column 1) are similar to the baseline IV, suggesting that the IV does not resolve the endogeneity problem. In fact, the OLS and IV coefficients are nearly identical, consistent with the instrument capturing the same omitted variables as the OLS regression.

6.3 The Placebo Failure

Table 4: Sector Heterogeneity: Treatment vs. Placebo Sectors

	Coef	SE	p-value	N
Panel A: Treatment Sectors (low-wage, immigrant-intensive)				
Log Total Employment	11.5207	4.6494	0.017	720
Log Weekly Wage	2.6932	1.0576	0.015	720
Log Establishments	12.0651	4.4366	0.009	720
Log Accommodation Emp	10.8145	4.2467	0.015	710
Log Admin Services Emp	9.8113	4.2469	0.026	710
Noncitizen Share	0.2167	0.1583	0.178	500
Panel B: Placebo Sectors (high-wage, native-dominated)				
Log Finance Emp (Placebo)	13.6073	5.4657	0.017	720
Log Professional Emp (Placebo)	12.3194	4.8500	0.015	688

Notes: Diagnostic IV exercise (year FE, court-clustered SEs). Panel A shows sectors where asylum recipients are likely to work (accommodation & food, administrative services). Panel B shows placebo sectors (finance, professional services) where asylum seekers are unlikely to find employment. Significant associations in Panel B demonstrate that the instrument captures general economic activity rather than the legal status channel. **Not causal estimates**—see Sections 5.4 and 6.3.

Table 4 presents the sector-heterogeneity results, which are the paper's most important diagnostic. If the instrument captures the causal effect of legal status, it should affect sectors where asylum grantees compete (accommodation, admin services) but not sectors where they do not (finance, professional services).

The results are unambiguous: all sectors respond. In the diagnostic regression (year FE only), accommodation employment shows a positive association ($\hat{\beta} = 10.8$, $p = 0.015$), but so does finance ($\hat{\beta} = 13.6$, $p = 0.017$) and professional services ($\hat{\beta} = 12.3$, $p = 0.015$). The placebo coefficients are not "imprecise zeros"—they are large, positive, and statistically significant at the 5% level. The magnitudes in the placebo sectors actually *exceed* those in the treatment sectors.

This pattern is inconsistent with the legal status mechanism and diagnostic of confounding. Courts with lenient judges are simply in larger, more economically active counties. The instrument predicts economic scale, not the marginal effect of asylum grants.

With county controls (Table 3, column 3), the picture improves partially: the placebo effects attenuate (finance: $\hat{\beta} = 6.8$, $p = 0.121$; professional: $\hat{\beta} = 8.2$, $p = 0.056$) while accommodation remains significant ($\hat{\beta} = 6.0$, $p = 0.020$). But the residual placebo effects at the 6% level, combined with the coefficient instability, do not support causal interpretation.

6.4 Robustness

Table 5: Robustness Checks

	Coef	SE	p-value	N
Panel A: Alternative Fixed Effects				
Year FE only	11.5207	4.6494	0.017	720
Region + Year FE	11.8830	4.7520	0.016	720
State + Year FE	6.6417	5.3948	0.225	720
No FE	11.5148	4.6099	0.016	720
Panel B: Alternative Clustering				
Court	11.5207	4.6494	0.017	720
State	11.5207	4.1570	0.011	720

Notes: Diagnostic IV exercise for log total employment under alternative specifications. Panel A varies the fixed effects. Panel B varies the clustering of standard errors. The baseline uses year fixed effects and court-clustered standard errors. **Not causal estimates**—reported to assess sensitivity of the confounded association.

Table 5 presents robustness checks for the total employment outcome. Panel A varies the fixed effects specification. The year FE baseline yields $\hat{\beta} = 11.52$. Adding region and year fixed effects (additively) barely changes the estimate (11.88), but state plus year fixed effects—which absorb more cross-court variation—reduce it to 6.64 with $p = 0.225$. The attenuation under state fixed effects provides further evidence of confounding: within-state variation in judge leniency is insufficient to identify the effect once state-level characteristics are controlled.

Panel B shows that alternative clustering choices do not meaningfully change the standard errors. Clustering at the state level (rather than the court level) yields standard errors of 4.16 compared to 4.65 under court clustering, a modest reduction reflecting the fact that most states contain only one immigration court.

6.5 Leave-One-Court-Out Stability

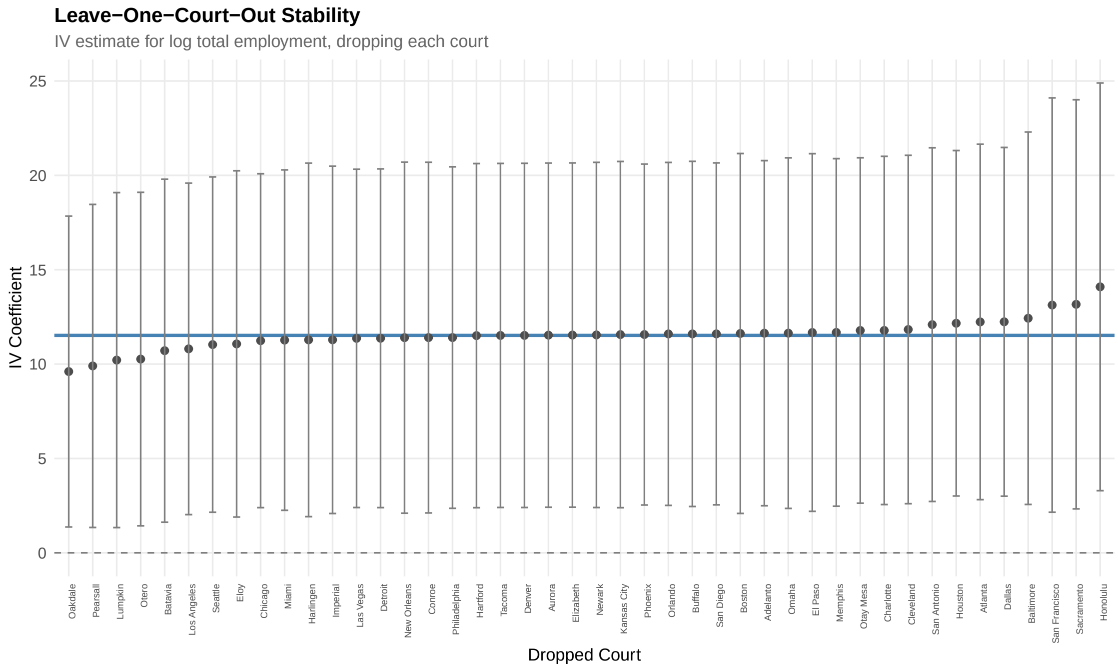

Figure 2: Leave-one-court-out stability. Each point is the IV estimate for log total employment when one court is dropped. The blue line marks the full-sample estimate. No single court drives the result, but this stability does not address the exclusion restriction violation.

Figure 2 demonstrates that no single court drives the results. Dropping each of the 44 courts in turn yields estimates ranging from 9.6 to 14.1, clustering tightly around the full-sample estimate of 11.5. This stability rules out the concern that a single outlier court drives the findings. However, leave-one-out stability does not address the exclusion restriction violation: if the confounding is systematic (all courts in economically larger areas are more lenient), removing any one court will not resolve it.

6.6 Subsample Analysis

The IV coefficient is stable across time periods: the early-period estimate (2005–2014) is 10.9 (SE $= 5.12$, $p = 0.038$, $N = 396$), and the late-period estimate (2015–2023) is 12.4 (SE $= 5.84$, $p = 0.041$, $N = 324$). This stability across periods with very different immigration policy regimes—the 2005–2014 period covers the Great Recession and the pre-surge asylum era, while 2015–2023 covers the dramatic increase in asylum filings and the COVID-19 pandemic—reinforces the interpretation that the instrument captures stable court-area characteristics rather than time-varying policy effects. A time-varying causal mechanism would plausibly shift in magnitude across these regimes, while a cross-sectional correlate of court location would remain stable, as observed.

6.7 Monotonicity

I examine whether judge leniency predicts court grant rates within each Census region. All four regions show positive first-stage coefficients: West ($\hat{\beta} = 1.24$, SE = 0.06, $N = 288$, 16 courts), Northeast ($\hat{\beta} = 1.36$, SE = 0.09, $N = 126$, 7 courts), South ($\hat{\beta} = 1.26$, SE = 0.07, $N = 216$, 16 courts), and Midwest ($\hat{\beta} = 1.16$, SE = 0.11, $N = 90$, 5 courts). All coefficients are highly significant ($p < 0.001$). These results are *consistent with* monotonicity—more lenient courts have higher grant rates in every region—but they do not constitute a formal test. Regional positive slopes establish that no geographic cluster violates the monotonicity direction, but individual-level defiers (cases where a more lenient judge reduces the probability of a grant) cannot be ruled out with aggregate data. Regardless, the binding constraint on this design is the exclusion restriction, not monotonicity.

6.8 Noncitizen Population Effects

A natural "first-stage validation" outcome is the noncitizen population share. If higher grant rates increase the local noncitizen population (through grantees staying and potentially attracting family members), this would support the labor supply channel. The IV estimate for noncitizen share is 0.217, but with a p-value of 0.178, it is not statistically significant at conventional levels. This null finding could reflect measurement error in the ACS noncitizen estimates (which are survey-based with substantial sampling error at the county level), the smaller sample (500 observations due to ACS coverage limitations), geographic dispersal of grantees beyond the court's county, or the possibility that in the cross-sectional design, the noncitizen effect is simply not identifiable above the noise of pre-existing immigrant population differences.

The noncitizen share result is notable because it is one of the few outcomes where the IV coefficient is small relative to the standard deviation of the outcome. The standardized effect size (SDE = 0.20) is the smallest among the outcomes examined, and the p-value is the largest. This is weakly consistent with the legal status mechanism—where the direct population effect is proportionally smaller than the economic multiplier effects—but the lack of significance prevents strong conclusions.

6.9 Magnitude Assessment

The implausibly large magnitudes of the IV coefficients provide additional evidence against causal interpretation. The baseline IV estimate for log total employment is 11.52, implying that a 1 percentage point increase in the grant rate raises employment by approximately 12 percent. For an average county with 585,000 employees, this corresponds to approximately 70,000 additional jobs per percentage point. Given that a typical court processes 3,000–5,000 cases per year, a 1 percentage point grant rate increase would generate only 30–50 additional asylum grantees—far too few to plausibly create 70,000 jobs through any combination of direct employment, demand multipliers, and business formation.

These magnitudes are inconsistent with any reasonable economic model but are exactly what one would expect from a confounded cross-sectional correlation. Courts with 1 percentage point higher grant rates happen to be in counties with approximately 12 percent more employment—not because the grants caused the employment, but because lenient courts and large economies co-occur.

7. Discussion

7.1 Why the Design Fails

The cross-sectional judge leniency instrument fails for three interconnected reasons:

(1) No court fixed effects. The instrument varies only across courts, not within courts over time, so court fixed effects cannot be included. All time-invariant differences between court areas contaminate the estimates. The sector-heterogeneity diagnostic (Section 6.3) confirms that the instrument captures general economic scale rather than the legal status channel.

(2) Endogenous judge sorting. Balance tests (Section 6.1) show that courts with lenient judges are in areas with larger immigrant populations ($p = 0.010$), suggesting that judge composition is correlated with area characteristics rather than randomly assigned across courts.

(3) Case-mix contamination. The instrument is based on lifetime judge grant rates that reflect not only judge preferences but also the types of cases each court adjudicates. Gateway courts hear cases from nationalities with stronger asylum claims (e.g., Chinese applicants facing persecution have higher national grant rates), process more non-detained cases (which have higher grant rates), and serve applicants with higher representation rates. These compositional differences mechanically raise court-level grant rates at the same courts that are located in larger economies. Even if within-court assignment is random, across-court differences in case composition confound the lifetime grant rate instrument. Disentangling judge leniency from case mix requires case-level data with nationality, detention status, and representation controls—which the aggregate data used here cannot provide.

A secondary concern is the temporal look-ahead in instrument construction (the lifetime grant rates incorporate post-sample decisions from 2024–2025; see Section 4.1), though the practical contamination is small given the long-career averaging.

The canonical judge IV design avoids these problems by conditioning on court fixed effects, exploiting only within-court variation across judges. For immigration courts, this would require: (a) case-level data with judge identifiers and decision dates, to construct time-varying court-year instruments from judge turnover; (b) court fixed effects in the specification. Case-level data would also resolve the look-ahead concern and the case-mix contamination by allowing year-specific instrument construction with appropriate controls. With the publicly available aggregate data used in this paper, none of these are feasible.

The cross-sectional court-level instrument also shares structural similarities with exposure designs (Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020), where identification rests on the exogeneity of "shares" (here, the composition of judges at each court). Goldsmith-Pinkham et al. (2020) show that in such designs, the identifying variation may be driven by a few influential locations, and the exclusion restriction must hold for each share separately. Applied to this setting, if a few large courts drive the variation and those courts have systematically different economies, the design fails—which is precisely what the balance tests and sector diagnostics reveal.

7.2 The Path to Credible Identification

A credible version of this design would proceed as follows:

Step 1: Obtain case-level EOIR data. The Department of Justice FOIA Library distributes case-level records for all immigration court proceedings. These data include case identifiers, judge assignments, decision dates, and court locations. With case-level data, one could construct a time-varying leave-one-out instrument following Kolesár (2013):

$$Z_{ct} = \frac{1}{N_{ct}} \sum_{i \in \mathcal{C}_{ct}} LOO\_GR_{j(i),ct} \tag{6}$$

where $LOO\_GR_{j(i),ct}$ is the grant rate of the judge assigned to case i, computed from all of that judge's other cases at court c in year t.

Step 2: Include court fixed effects. With within-court time variation in the instrument (driven by judge turnover, retirements, and new appointments), the specification becomes:

$$Y_{ct} = \alpha + \beta \cdot GrantRate_{ct} + \gamma_c + \delta_t + \varepsilon_{ct} \tag{7}$$

Court fixed effects γ_c absorb all time-invariant differences between court areas—exactly the confounders that drive the placebo failure in this paper.

Step 3: Verify balance and placebos with court FE. The balance tests and placebo tests should pass once court characteristics are absorbed. If they still fail, additional controls (e.g., court $\times$ time trends) or a more restrictive instrument (e.g., year-to-year changes in judge composition) may be needed.

Step 4: Address many-instruments bias. With case-level data and many judges per court, the number of instruments grows large. UJIVE (Kolesár, 2013) or jackknife IV addresses this concern.

7.3 What We Learn Despite the Failure

The exercise is not without value. Several findings survive the identification critique and provide useful inputs for future research.

First, *within-court judge variation is enormous and real.* The 50+ percentage point within-court disparities in asylum grant rates are not an artifact of aggregation or small samples. Figure 3 in the appendix documents this variation across the 25 largest courts, using only judges with 100+ hearings at each court. At the Los Angeles court, for instance, judge-level grant rates span nearly the entire unit interval. This variation is the raw material for a credible IV and suggests that the asylum lottery is a first-order source of variation in legal status outcomes.

Second, *judge leniency is systematically correlated with court-area characteristics.* This finding—which I present as an identification failure—is itself informative for the political economy of immigration adjudication. It suggests that the assignment of judges to courts is not random with respect to area characteristics, raising questions about the appointments process and its downstream consequences. Why do courts in large gateway cities tend to have more lenient judges? Several mechanisms are plausible: judges may sort geographically by ideology, the Attorney General's appointment decisions may respond to local political pressure, courts in areas with stronger immigrant advocacy networks may cultivate more sympathetic adjudication cultures, or the case composition at gateway courts (with more nationalities qualifying for asylum at higher rates) may drive higher aggregate grant rates. Disentangling these mechanisms is beyond the scope of this paper but is a natural extension.

Third, *the first-stage power is exceptional.* The F-statistic of 855 in the panel specification, with only 44 courts, suggests that a within-court design with time-varying instruments would have more than adequate power to detect even small labor market effects. Weak instruments—the primary concern in many IV applications—would not be a problem in a properly implemented immigration judge IV. To put this in perspective, the Stock and Yogo (2005) threshold for 10% maximal IV bias with one instrument is $F > 16.4$; even a within-court specification that explains only 5% as much variation as the cross-sectional instrument would produce F-statistics above 40.

Fourth, *the cross-sectional correlation, while confounded, provides suggestive information about the direction of bias.* The balance tests show that lenient courts tend to be in larger, more immigrant-rich areas, which is consistent with positive confounding. If the confounding is indeed positive, the true causal effect would be smaller than the IV estimate. However, the sign and structure of the bias are not established with confidence—confounding could also operate through channels that attenuate the estimate—so no formal bounding claim is warranted. The controlled specification reduces the coefficient by roughly half, but this attenuation partly reflects sample changes (from 720 to 500 observations) and cannot be cleanly attributed to the removal of confounders alone.

Fifth, *the data infrastructure developed here is reusable.* The court-county crosswalk, the QCEW outcome panel, and the ACS demographic controls constructed for this analysis can be combined with case-level EOIR data by future researchers to implement the within-court design. The OpenImmigration scraper provides judge-court assignments that can anchor any analysis of immigration judge behavior.

7.4 Comparison to Existing Immigration-Legal Status Literature

The existing literature on the labor market effects of legal status relies on designs with different strengths and weaknesses. Amuedo-Dorantes et al. (2022) study the effects of DACA, estimating effects on unauthorized immigrants themselves rather than on local labor markets. Amuedo-Dorantes and Antman (2020) examine DACA's labor market effects for Hispanic youth, and Orrenius and Zavodny (2015) document the legal status wage penalty directly. These studies establish that legal status matters for individual outcomes, but none identifies general-equilibrium effects on local labor markets. Their design is credible for individual-level effects but cannot identify local general equilibrium impacts. Chassamboulli and Peri (2014) calibrate a search model with legal status as a state variable, providing structural estimates of how legalization affects equilibrium wages and unemployment, but the results depend on parametric assumptions. The judge leniency design proposed here would complement both approaches: it operates at the local labor market level (like the structural models) but uses quasi-experimental variation (like the DACA studies), and it varies legal status for a population that is otherwise identical in unobservable characteristics (unlike the DACA cutoff, which compares different age cohorts).

The closest existing paper using immigration judge variation is Albornoz-Crespo et al. (2023), who study the effects of judge decisions on refugee integration outcomes in the European context. However, their analysis focuses on individual-level integration rather than local labor market equilibria, and the European institutional setting differs substantially from the U.S. EOIR system.

7.5 Implications for Asylum Policy

Even without credibly causal estimates, the exercise highlights an important policy dimension. The GAO has documented the asylum system's inconsistency as a fairness problem—like cases should be treated alike regardless of which judge is assigned. The potential efficiency consequences amplify this concern. If future research with case-level data confirms that judge leniency causally affects local labor markets through the legal status channel, then the GAO's documented disparities have welfare consequences beyond the individual asylum seekers: they affect local employment, business formation, and tax revenue. This would add an economic efficiency argument to the existing fairness argument for reducing adjudication disparities through standardized guidelines, training, or quality review.

Conversely, if future within-court analyses find null labor market effects—suggesting that the local economy absorbs legal status changes without detectable impacts—this would also be valuable, indicating that asylum reform can focus on fairness without significant economic spillovers in either direction.

7.6 Limitations

Beyond the central identification failure, several additional limitations apply. The court-to-county mapping is imprecise: asylum seekers may reside in adjacent counties, attenuating the geographic link. The ACS demographic variables are available only from 2009 (5-year estimates) and have limited coverage for smaller counties, reducing the sample from 720 to 500 observations for balance tests involving noncitizen share. The instrument cannot decompose the legal status channel into its constituent parts (work authorization, benefits, deportation risk). Finally, the 44-court sample, while covering the majority of asylum adjudication volume, excludes smaller courts where identification might differ.

8. Conclusion

Whether an asylum seeker receives legal status in the United States depends substantially on which judge hears their case—a quasi-random event that the GAO has termed the "asylum lottery." This paper asks whether these lotteries shape local labor markets and finds that the answer, with currently available data, is inconclusive.

The cross-sectional judge leniency instrument has enormous first-stage power ($F = 855$) but fails the sector-heterogeneity diagnostic: high-wage sectors that should not respond to asylum grants show effects as large as low-wage treatment sectors. The balance tests reveal systematic correlations between judge leniency and county characteristics. These failures indicate that the identifying variation is confounded by the nonrandom placement of courts and the systematic sorting of judges across court areas.

The contribution of this paper is therefore primarily methodological. I document that the immigration judge leniency design—one of the most promising quasi-experimental approaches for studying legal status and labor markets—requires within-court time variation to achieve credible identification. The publicly available aggregate data used here permit only cross-sectional variation, which does not satisfy the exclusion restriction. Case-level EOIR microdata with time-varying judge assignments, combined with court fixed effects, would overcome this limitation and could yield some of the strongest causal evidence in the immigration economics literature.

The paper also contributes descriptive findings that are useful regardless of the identification failure. The within-court variation in judge leniency—documented across 44 courts and 1,268 judges—confirms the GAO's characterization of the asylum system as a "lottery" with enormous stake variation. The systematic correlation between judge composition and area characteristics raises new questions about the political economy of judicial appointments in immigration law. And the data infrastructure constructed here—court-county crosswalks, outcome panels, judge-level grant rate data—provides a foundation for future research with stronger identification.

Three directions for future research emerge naturally. First, obtaining and processing the full EOIR case-level data (approximately 4 GB, available via FOIA) would enable the within-court time-varying instrument that the cross-sectional approach cannot provide. Judge turnover, retirements, and new appointments create natural experiments within courts that can be leveraged with court fixed effects. Second, linking EOIR case records to individual-level administrative data—tax records, employment histories, or benefit receipt—would allow researchers to trace the economic trajectories of individual asylum grantees and denied applicants, replacing area-level analysis with person-level causal effects. Third, studying how the composition of the immigration judge corps responds to political cycles would illuminate the political economy of asylum adjudication and the mechanisms driving the systematic judge-area correlation documented here.

The asylum lottery is real, enormous, and well-documented. Its labor market consequences remain an open question—one that deserves a research design strong enough to answer it.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Albornoz-Crespo, Narciso, Antonio Cabrales, and Esther Hauk, "Immigration Judges and U.S. Asylum Policy: Evidence from Refugee Integration Outcomes," *Journal of Public Economics*, 2023, *217*, 104788.

Amuedo-Dorantes, Catalina and Francisca Antman, "Deferred Action for Childhood Arrivals and the Labor Market Outcomes of Hispanic Youth," *Southern Economic Journal*, 2020, *86* (4), 1430–1452.

_, Esther Arenas-Arroyo, and Almudena Sevilla, "The Effects of Legalizing Undocumented Immigrants," *Journal of Public Economics*, 2022, *206*, 104587.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad, "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 2020, *128* (4), 1269–1324.

Borjas, George J., "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, 2003, *118* (4), 1335–1374.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Chassamboulli, Andri and Giovanni Peri, "The Labor Market Effects of Reducing the Number of Illegal Immigrants," *Review of Economic Dynamics*, 2014, *17* (4), 792–821.

Dahl, Gordon B., Andreas Ravndal Kostol, and Magne Mogstad, "Peer Effects in Program Participation," *American Economic Review*, 2014, *104* (7), 2049–2074.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 2018, *108* (2), 201–240.

Dustmann, Christian and Joseph-Simon Görlach, "The Economics of Temporary Migrations," *Journal of Economic Literature*, 2016, *54* (1), 98–136.

_ , **Uta Schönberg, and Jan Stuhler**, "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment," *Quarterly Journal of Economics*, 2017, *132* (1), 435–483.

Foged, Mette and Giovanni Peri, "Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data," *American Economic Journal: Applied Economics*, 2016, *8* (2), 1–34.

Frandsen, Brigham R., Lars J. Lefgren, and Emily C. Leslie, "Judging Judge Fixed Effects," *American Economic Review*, 2023, *113* (1), 253–277.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Kling, Jeffrey R., "Incarceration Length, Employment, and Earnings," *American Economic Review*, 2006, *96* (3), 863–876.

Kolesár, Michal, "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity," *Working Paper*, 2013.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand, "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 2013, *103* (5), 1797–1829.

Orrenius, Pia M. and Madeline Zavodny, "The Effect of Immigrant Legal Status on Wages," *International Migration Review*, 2015, *49* (3), 828–849.

Ottaviano, Gianmarco I.P. and Giovanni Peri, "Rethinking the Effect of Immigration on Wages," *Journal of the European Economic Association*, 2012, *10* (1), 152–197.

Peri, Giovanni and Chad Sparber, "Task Specialization, Immigration, and Wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–169.

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 2005, pp. 80–108.

U.S. Government Accountability Office, "U.S. Asylum System: Significant Variation Existed in Asylum Outcomes across Immigration Courts and Judges," Technical Report GAO-08-940, GAO 2008.

_ , "Asylum: Variation Exists in Outcomes of Applications across Immigration Courts and Judges," Technical Report GAO-17-72, GAO 2017.

A. Data Appendix

A.1 OpenImmigration Judge Data

Judge-level data were scraped from OpenImmigration (https://openimmigration.org) in March 2026. The scraper downloaded individual judge pages and extracted court assignments, hearing counts, and grant rates from the HTML. A total of 1,268 judges were scraped, yielding 10,920 judge-court assignment records. Judges with fewer than 10 hearings at a court were excluded from instrument construction.

Judge leniency construction:

1. For each judge j, the lifetime asylum grant rate GR_j is obtained directly from OpenImmigration.
2. For each court c, the instrument is the caseload-weighted average: $Z_c = \sum_j H_{jc} \cdot GR_j / \sum_j H_{jc}$, where H_{jc} is the number of hearings by judge j at court c.
3. A leave-one-out variant is also computed, but because it leaves out individual judges (not cases), it remains cross-sectional.
4. Courts with fewer than 3 judges are excluded.

A.2 BLS QCEW

The Quarterly Census of Employment and Wages is accessed via BLS bulk downloads (2005–2015) and the BLS data API (2014–2023). I retain county-level, private-sector (own_code = 5) records for five NAICS sectors: total private (10), accommodation and food services (72), administrative and support services (56), finance and insurance (52), and professional and technical services (54).

A.3 Census ACS

The American Community Survey 5-year estimates are accessed via the Census Bureau API. Variables: noncitizen population, total population, foreign-born population, poverty status, and labor force statistics. County-level estimates for 2009–2023.

A.4 Court-County Crosswalk

I construct the court-county mapping from the EOIR immigration court listing and OpenImmigration court directory. Each court is mapped to the county FIPS code of its physical location. The final crosswalk links 44 courts to 43 counties (New York County hosts two courts; see Section 4.4 for how this is handled).

B. Additional Figures

B.1 Within-Court Judge Variation

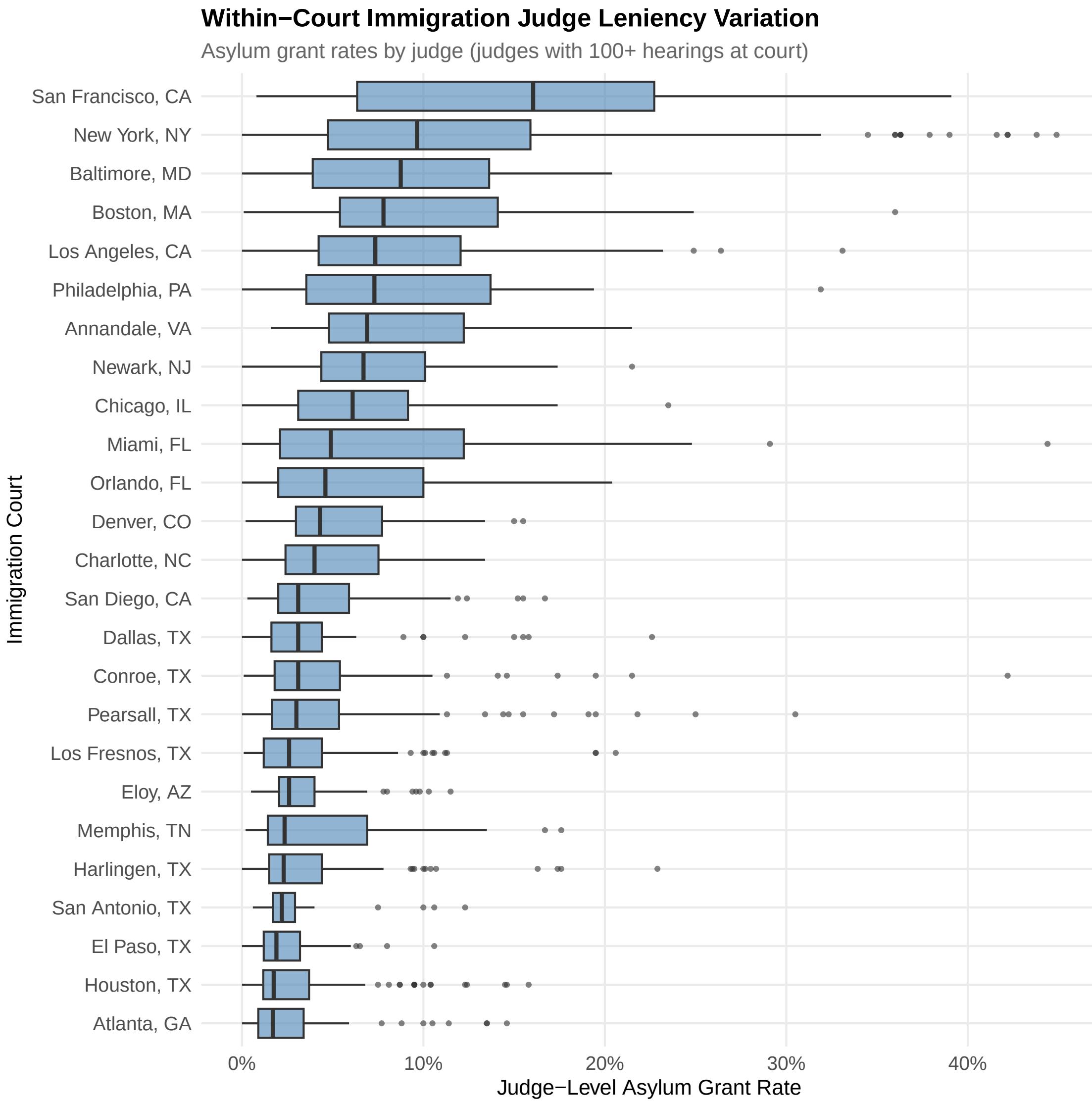

Figure 3: Within-court immigration judge leniency variation. Box plots show the distribution of judge-level asylum grant rates within the 25 largest immigration courts. Only judges with 100 or more hearings at each court are included.

B.2 First Stage Scatter

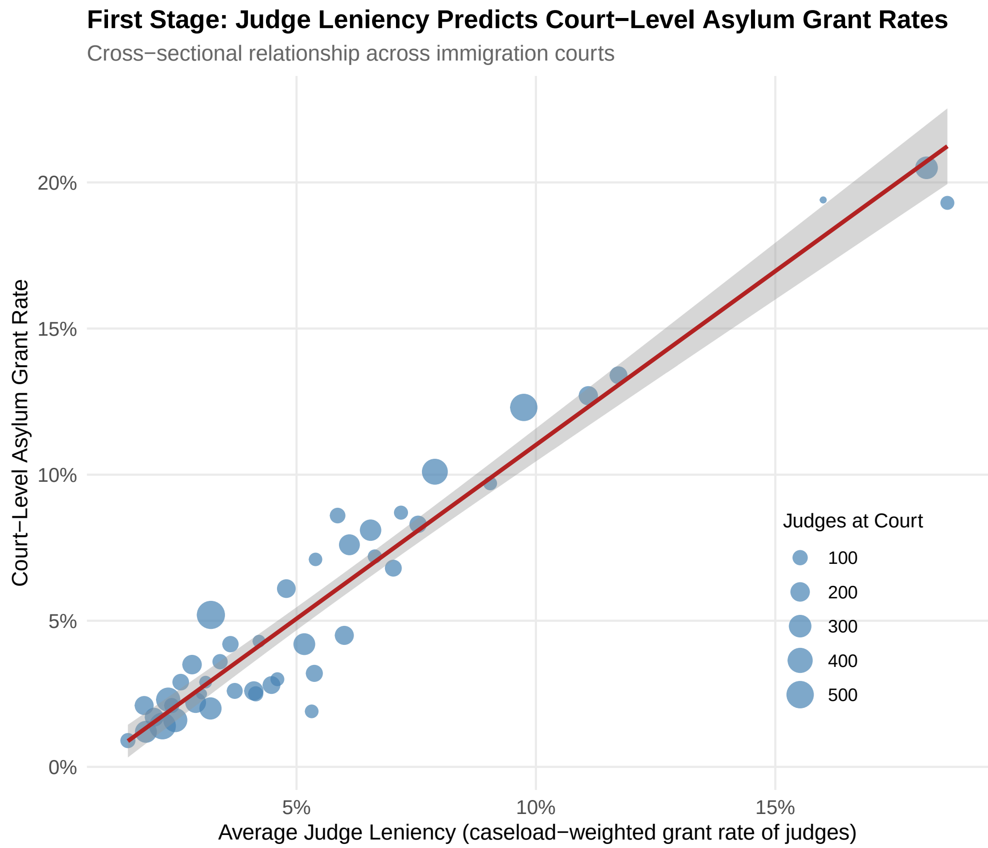

Figure 4: First stage: average judge leniency vs. court-level asylum grant rate. Each point is one court (cross-sectional average). Point size proportional to number of judges. The near-unity slope ($\hat{\beta} = 1.19$, $R^2 = 0.94$) reflects the mechanical first-stage relationship.

B.3 IV Coefficient Plot

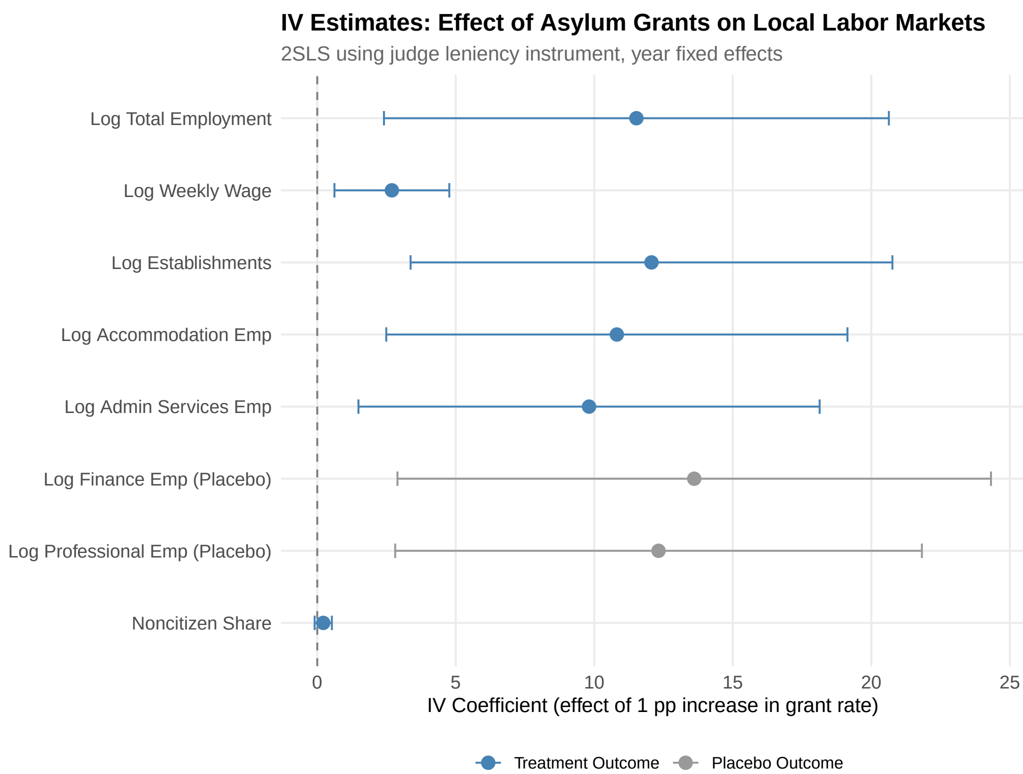

Figure 5: IV estimates across outcomes. Point estimates and 95% confidence intervals for the effect of a 1 unit increase in the asylum grant rate on each outcome variable. Blue points are treatment outcomes; grey points are placebo outcomes. The placebo sectors (finance, professional services) show effects of comparable magnitude to treatment sectors, indicating the exclusion restriction is violated.

B.4 Placebo Comparison

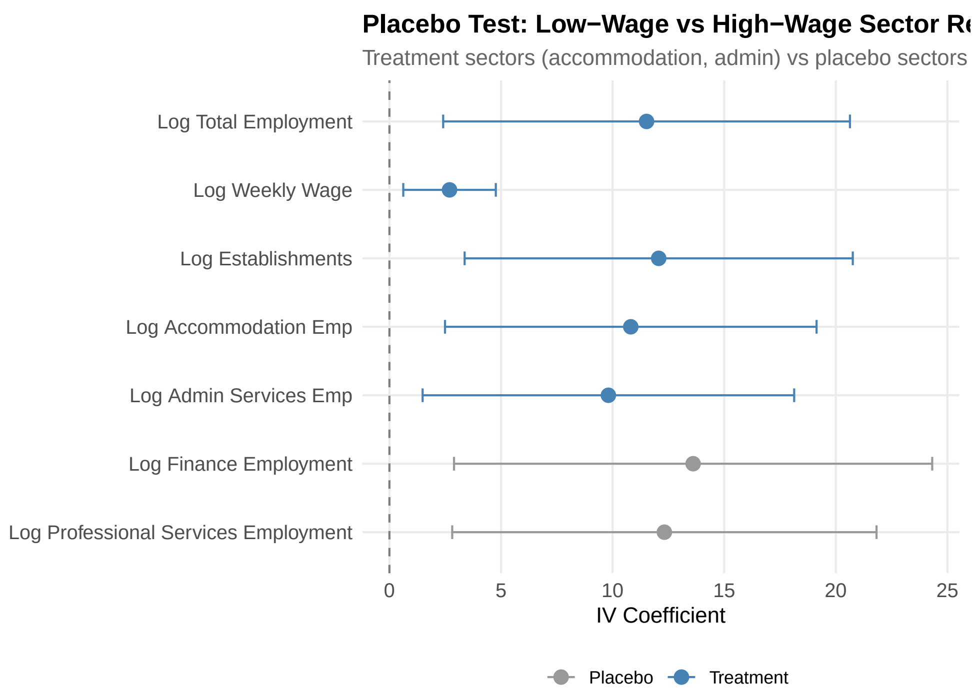

Figure 6: Placebo test: treatment sectors vs. placebo sectors. Both groups show significant, positive IV coefficients of similar magnitude, indicating confounding rather than a legal status channel.

B.5 Alternative FE Specifications

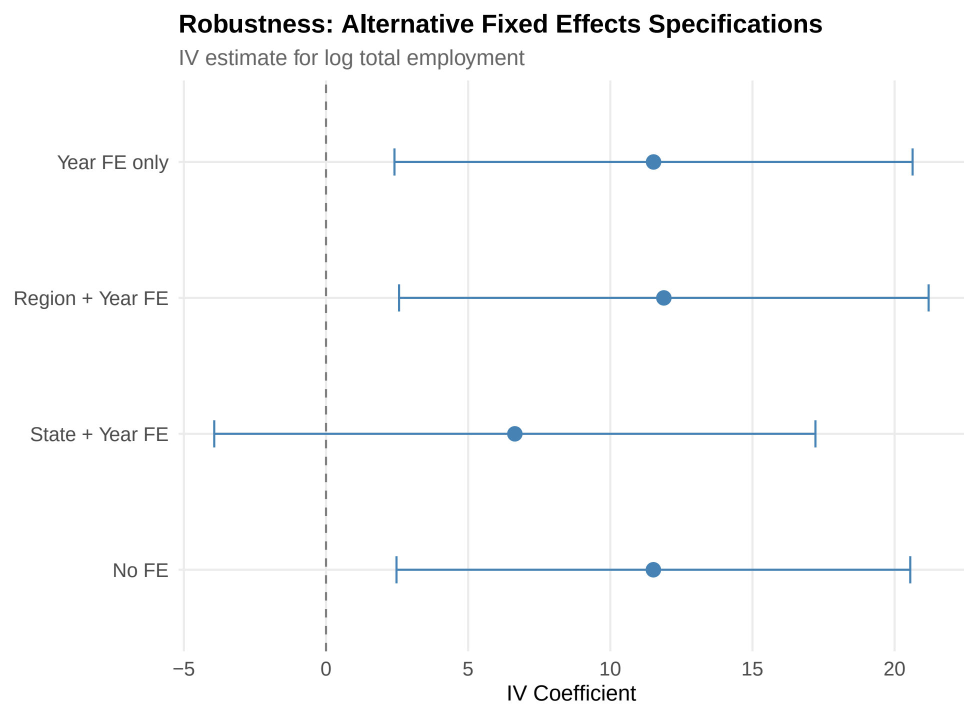

Figure 7: Robustness across fixed effects specifications. The IV estimate attenuates substantially under state + year FE ($\hat{\beta} = 6.6$, $p = 0.225$), consistent with confounding from cross-state variation in court characteristics.

B.6 Leniency Distribution

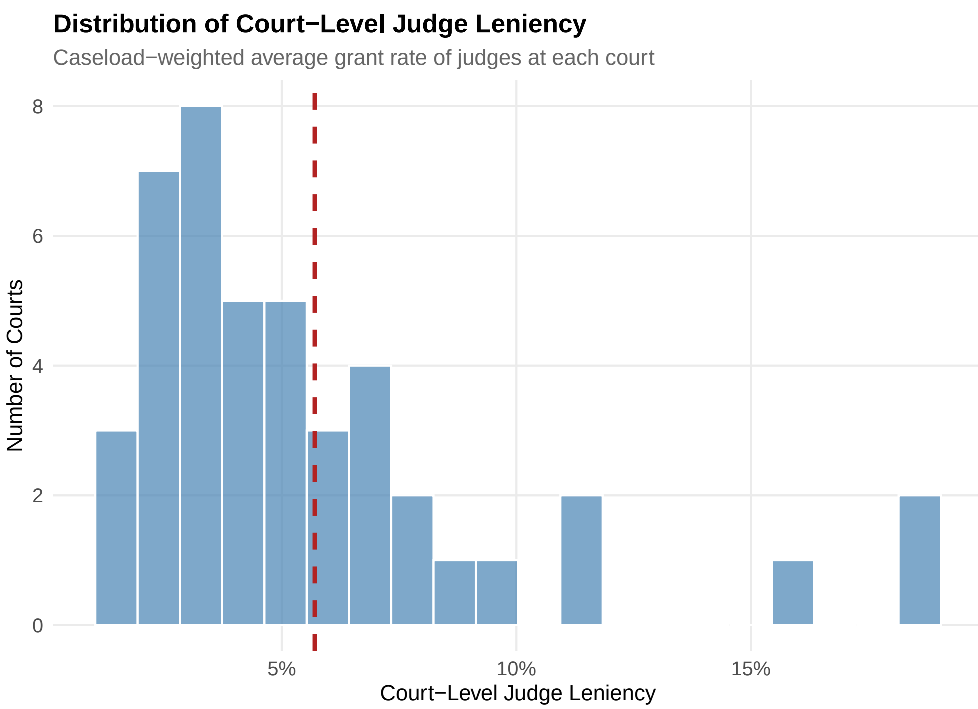

Figure 8: Distribution of court-level judge leniency across 44 immigration courts. Dashed line marks the mean. The substantial cross-court variation in judge composition is the raw material for the IV.

C. Standardized Effect Sizes

Table 6: Standardized Effect Sizes for Main Outcomes

Outcome	Specification	$\hat{\beta}$	SD(X)	SD(Y)	SDE	Classification
Log Total Employment	Diagnostic IV, Table 3 Col. 2	11.521	0.048	1.566	0.351	Large positive
Log Weekly Wage	Diagnostic IV, Table 3 Col. 2	2.693	0.048	0.315	0.408	Large positive
Log Establishments	Diagnostic IV, Table 3 Col. 2	12.065	0.048	1.498	0.384	Large positive
Log Accommodation Emp	Diagnostic IV, Table 3 Col. 2	10.815	0.048	1.539	0.335	Large positive
Noncitizen Share	Diagnostic IV, Table 3 Col. 2	0.217	0.048	0.053	0.195	Large positive

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison. However, because the placebo tests indicate that the exclusion restriction is violated (Section 6.3), these SDEs should **not** be interpreted as causal effects. They reflect the confounded cross-sectional correlation between court-level judge leniency and county-level outcomes.

The treatment variable (asylum grant rate) is continuous, so SDE = $\hat{\beta} \times \text{SD}(X)/\text{SD}(Y)$, which gives the effect of a one-standard-deviation change in the grant rate, measured in standard deviations of the outcome. SD(Y) and SD(X) are unconditional standard deviations from the summary statistics (Table 1).

Research question: Does quasi-random variation in asylum grant rates caused by immigration judge leniency affect local labor markets? **Treatment:** Continuous — the asylum grant rate at court c in year t (range 0–1), instrumented by caseload-weighted average judge leniency (cross-sectional). **Data:** OpenImmigration judge data (1,268 judges), BLS QCEW and Census ACS county outcomes, 2005–2023. Unit: court × year linked to county. $N = 720$ court-year observations across 44 courts. **Method:** 2SLS with cross-sectional judge leniency instrument, year FE, court-clustered SEs. **Caveat:** The identification strategy fails its sector-heterogeneity diagnostic (Section 6.3), so the SDEs are not credibly causal.

Classification thresholds: large negative (< -0.10), small negative (-0.10 to -0.05), null (-0.05 to 0.05), small positive (0.05 to 0.10), large positive (> 0.10).