Reshaping the Bench:

The Effect of Case Quotas in U.S. Immigration Courts*

Catalina Amuedo-Dorantes[1] and Eduardo Polo-Muro[2]

[1]University of California, Merced
[2]University of the Basque Country and San Diego State University

Abstract

We study the DOJ's 2018 quota requiring immigration judges to complete 700 cases annually. Using nationwide EOIR data (2000–2019) and a difference-in-differences design with exposure based on pre-policy predicted completions, we find the quota raised judges' separation likelihood by 8.6 percentage points—from 4.1% to 12.7%. Effects were strongest among more-experienced, female, and non-White judges. At the court level, higher exposure increased departures, lowered average experience, and did not measurably raise overall case completions. Grant rates fell as both non-removal and removal cases declined. Overall, the quota destabilized staffing, eroded expertise, and shifted outcomes away from relief without delivering throughput gains.

Keywords: Immigration policy, immigration courts, United States.
JEL Classification: K41, K42, J15, J18.

---

*We are grateful to Ariel Binder, Jose Bucheli, Jeff Clemens, Julie Cullen, Gordon Dahl, Melissa Famulari, Alex Gelber, Neeraj Kaushal, Paul Niehaus, Susan Pozo, and participants at the 2024 APPAM meetings and the 2025 Berkeley Migration Seminar for their feedback on earlier drafts. Eduardo Polo-Muro acknowledges financial support from the Basque Government (POS-2021-1-0030) and the Department for Education, Universities, and Research of the Basque Government under research grant IT1336-19.

1 Introduction

One of the hallmarks of democratic systems is the separation of powers, which ensures judicial independence and safeguards impartiality. Immigration Courts, however, operate within the Department of Justice's (DOJ) Executive Office for Immigration Review (EOIR), under a structure that blurs those lines. Immigration judges serve as civil servants who report to the Attorney General, while their decisions remain subject to judicial branch review. This dual position puts them at the intersection of executive and judicial authority, complicating efforts to preserve independence. Importantly, immigration judges are not Article III judges with lifetime tenure but Article II appointees subject to removal by the Attorney General. Their administrative status leaves them vulnerable to executive pressure, especially under performance-based evaluation systems such as case completion quotas. These features raise concerns about how internal DOJ policies affect who serves on the bench and, ultimately, the fairness of immigration proceedings.

In October 2018, a quota policy was introduced requiring immigration judges to complete a minimum of 700 cases annually to receive a "Satisfactory Performance" rating (GAO, 2017). The policy marked a sharp departure from prior performance metrics and was presented as a response to the growing immigration case backlog. While intended to enhance efficiency and reduce delays, it quickly became a focal point of public and scholarly debate. Critics warned that quotas could incentivize rushed decision-making and compromise case quality (Nielson, 2018; Rose, 2018), while others emphasized potential harms to migrants and their families. Proponents viewed the change as necessary to address longstanding inefficiencies in immigration court operations.

Existing research has begun to document the consequences of the quota policy. Most notably, Blasingame et al. (2023) report a nationwide increase in in-absentia removal orders and removal orders on the merits following the policy's adoption. Their analysis focuses on judge-level behavior in asylum adjudication—a narrow but high-stakes subset of immigration court proceedings. Our focus is different. We study how exposure to the quota policy affected judicial departures and, in turn, how those departures reshaped court composition, productivity, and decision-making across all case types. Because departures

alter the experience and characteristics of the judicial pool, understanding this mechanism is essential for assessing broader institutional effects.

Using administrative EOIR data from 2000 to 2019, we implement a quasi-experimental difference-in-differences design that leverages variation in immigration judges' likely exposure to the policy based on their pre-policy case completions. We find that the policy increased the likelihood of judicial departure by 8.6 percentage points. The effect was disproportionately larger for female judges, who remain underrepresented on the bench, and higher among non-White judges. This differential response is consequential. Prior work links judge gender to adjudication (Cook, 1981; Boyd et al., 2010; Boyd, 2016), so gender-skewed attrition—coupled with race-skewed exits—poses institutional risks for bench composition and downstream decision patterns.

At the court level, we measure exposure as the share of judges with predicted completions below the 700-case completion threshold. Courts with greater exposure experienced higher separation rates and a decline in average judicial experience (approximately 0.21–0.27 years per 10 percentage-point increase in exposure; moving from zero to full exposure corresponds to about two fewer years on the bench), yet no significant increase in case completions. Instead, the share of cases decided in favor of respondents fell. The decline appears in both removal and non-removal proceedings.[1] However, since removal cases account for the overwhelming majority of filings, the aggregate drop is arithmetically driven by such cases.

In sum, short-term performance benchmarks such as quotas can accelerate judicial

[1]Removal proceedings, which are the dominant class of immigration cases, begin when a respondent or individual alleged by Immigration and Customs Enforcement (ICE) to be removable faces charges outlined in a Notice to Appear (NTA). The immigration judge determines whether the respondent should be ordered removed or granted relief, such as asylum or cancellation of removal. When relief is granted, the case is considered "successful" or "granted," meaning the respondent may lawfully remain in the United States. Either ICE or the respondent may appeal the decision to the Board of Immigration Appeals (BIA).

turnover and erode decision-making capacity, compromising fairness without achieving the promised efficiency gains. Despite hiring over 100 judges in FY2022, the EOIR backlog continued to rise, surpassing 2 million cases by FY2023.[2] Moreover, in early 2025, the Justice Department abruptly terminated multiple immigration judges—widely reported at the time—while the backlog reached 3.6 million cases (Ainsley et al., 2025). Reports also noted that many of the dismissed judges had been appointed under the previous administration, intensifying concerns about politicization (Ainsley et al., 2025). Together, these developments heighten the stakes for understanding how internal DOJ policies determine who remains on the bench and how cases are resolved. The choices DOJ makes about staffing and performance management will set the pace—and the fairness—of adjudication in the years to come.

2 Institutional Framework

As mentioned in the Introduction, immigration judges are civil servants who report to the Attorney General. Yet they are also expected to uphold the highest standards of judicial conduct, placing them at the intersection of the executive and judicial branches of government. Immigration judges oversee removal proceedings, during which a "respondent"—the individual accused by Immigration and Customs Enforcement (ICE) of being removable—faces charges outlined in a Notice to Appear (NTA). The judge determines whether the respondent should be ordered removed or is eligible for relief from removal, such as asylum, for which the respondent must apply. A case is considered "granted" when the judge rules in favor of the respondent, allowing them to remain in the United States under some form of legal protection.

Cases are assigned to immigration courts based on the respondent's location or detention facility, not on judge or court preferences. Within each court, case assignment

[2]Congressional Research Service. 2023. "Immigration: Staffing and Funding for Executive Office for Immigration Review (EOIR)." Available at: https://crsreports.congress.gov/product/pdf/IF/IF12212.

follows internal docketing protocols and automated scheduling systems, including the Court Automated Scheduling Engine (CASE), which allocates cases based on availability, case type, and hearing constraints. This structure introduces a high degree of randomness and limits the ability of either judges or court administrators to steer particular cases. Although staff may adjust assignments to accommodate calendar conflicts or language needs, the system in practice constrains strategic behavior—such as working slowly to reduce assigned caseload or seeking particular case types. These safeguards help ensure that, to a meaningful extent, case flow is exogenous to judge characteristics and behavior, reinforcing the validity of judge-level causal analysis.

For more than a decade, immigration courts across the United States have struggled with a mounting backlog of removal cases, driven by too few judges relative to caseload. Despite ongoing hiring efforts, the backlog has continued to expand.[3] As shown in Appendix Figure A1, the backlog remained relatively stable in the early 2000s but began to rise in the early 2010s as case receipts outpaced completions. This trend accelerated after 2015, culminating in a sharp surge around 2019, likely due to policy changes, increased enforcement, and shifts in migration patterns. This persistent growth prompted the U.S. Government Accountability Office (GAO) to issue a 2017 report recommending actions to address the backlog, including the implementation of performance measures for judges, such as case completion goals (GAO, 2017).

The Trump administration responded in October 2018 with the "EOIR Performance Plan." Among other metrics,[4] the plan required immigration judges to complete at least 700 cases per fiscal year to receive a "Satisfactory Performance" rating (GAO, 2017; Nielson, 2018; Rose, 2018). The quota policy was widely criticized as a politically driven attempt to align adjudication with broader enforcement goals (Sacchetti, 2017; Benner,

[3]By FY2016, the number of pending immigration court cases had increased by 58 percent since the end of FY2012, reaching 518,545 (EOIR, 2016).

[4]In addition to the 700-case completion quota for a "Satisfactory" rating, judges were expected to maintain an appeals-overturn rate below 15 percent, resolve motions and custody cases promptly, and meet at least half of other case-management benchmarks (EOIR, 2018).

2018; Long, 2018). The National Association of Immigration Judges (NAIJ) condemned the policy as a threat to judicial independence and due process (Bar, 2020).

Recent research has examined the implications of this policy. Blasingame et al. (2023) show that—among judges who remained on the bench—removal orders increased. Because that analysis is conditional on staying, the overall response also hinges on who remained: if quotas raised separations, shifts in bench composition could mediate aggregate effects (Alvarez, 2019; Hennessy-Fiske, 2020). Understanding who remained on the bench is critical because, unlike Article III federal judges, who are appointed for life and enjoy constitutional protections, immigration judges are Article II appointees who serve at the discretion of the Attorney General, who retains authority to hire and dismiss them. The 2018 plan made clear that failing to meet the 700-case target could result in an "Unsatisfactory" rating. While records on dismissals tied to the quota are not public, fears of removal or career repercussions were widely reported among judges and bar observers (Rose, 2018; Sacchetti, 2017; Bar, 2020; U.S. Commission on Civil Rights, n.d.; National Association of Immigration Judges, n.d.). In this context, observed effects on judicial separations likely represent a lower bound of the policy's influence. The perceived risk alone could plausibly induce early retirements, resignations, or behavioral changes, regardless of whether formal consequences were ultimately enforced.

In the sections that follow, we examine how immigration judges' departure rates responded to the quota policy and whether responses varied by gender and race. We then investigate how the policy affected the expertise of judges who remained on the bench, as well as its implications for court case completion volumes and adjudication patterns.

3 Data and Some Descriptive Evidence

To examine the impact of the quota policy, we use administrative records from the Executive Office for Immigration Review (EOIR) covering all immigration judges, cases, and hearings from 2000–2019 to avoid pandemic-year distortions. We combine three datasets—case, hearing, and judge files—merged via unique identifiers. Case-level records

provide outcomes and adjudication information; hearing files contain dates and judge identifiers; and judge-level files include names, tenure, and court location. This structure lets us measure departures, caseloads, and experience consistently over time. Additional details appear in the Data Appendix.

Our primary goal is to assess whether the quota policy altered immigration judges' willingness to remain on the bench, as captured by their departure likelihood. We build an unbalanced panel of active judges by year. Because some judges begin with very small dockets (e.g., as interns), we restrict attention to judges with at least 75 cases per fiscal year and define the departure year as the first year in which a judge's caseload falls below 75. This threshold yields a total number of judges that matches official counts over the study period. Importantly, the data indicate whether a judge is active and, by implication, whether they separated, but they do not distinguish retirements, resignations, or dismissals. Nonetheless, sharp increases in departures among judges most exposed to the quota threshold are consistent with policy-related exits rather than routine retirements.

A related objective is to examine whether the performance policy changed the composition of the bench through differential responses by gender and race. Because gender and race are not directly observable, we infer them from two sources. First, we use the University of California, Irvine Machine Learning Repository UCI (2020), based on Social Security card applications from 1880–2019, to assign the most likely gender from names. Second, we use Rosenman et al. (2022), which links names to self-reported race using voter files across six states, to classify judges into White, Black, Hispanic, Asian, and Other.[5] Of the 738 immigration judges in our sample, we assign gender to 701 and race/ethnicity to 685. Finally, we compute each judge's total years of judicial experience, a measure we later use to study changes in judicial expertise.

Table 1 provides descriptive statistics for immigration judges and courts in our sample. Panel A summarizes the characteristics of 738 immigration judges observed between 2000 and 2019, including both new hires and longer-serving judges. On average, the annual

[5]We do not separately identify Hispanic ethnicity.

departure rate was 4.85 percent. Judges handled an average of 1,040 cases per year and had 9.75 years of judicial experience. Using name-based algorithms, we estimate that 36.83 percent of judges were female and 79.28 percent were White.

Table 1: Descriptive Statistics of Immigration Judges and Courts, 2000–2019

Statistic	Mean	Std. Dev.	Min	Max
Column	(1)	(2)	(3)	(4)
Panel A: Immigration Judge Level Dataset				
Departure Rate	0.0485	0.2150	0	1
Completed Cases	1039.53	674.49	76	4159
Judicial Experience	9.75	7.52	0	34
Female	0.3683	0.4827	0	1
White	0.7928	0.4055	0	1
N		5186		
Panel B: Immigration Court Level Dataset				
Share of Judges Departing	0.0434	0.1239	0	1
Average Judicial Experience	9.7984	5.1027	1	30
Share of Judges Below Threshold	0.2243	0.3080	0	1
Case Completion Volume	5076.87	4392.61	133	34978
N		1,062		

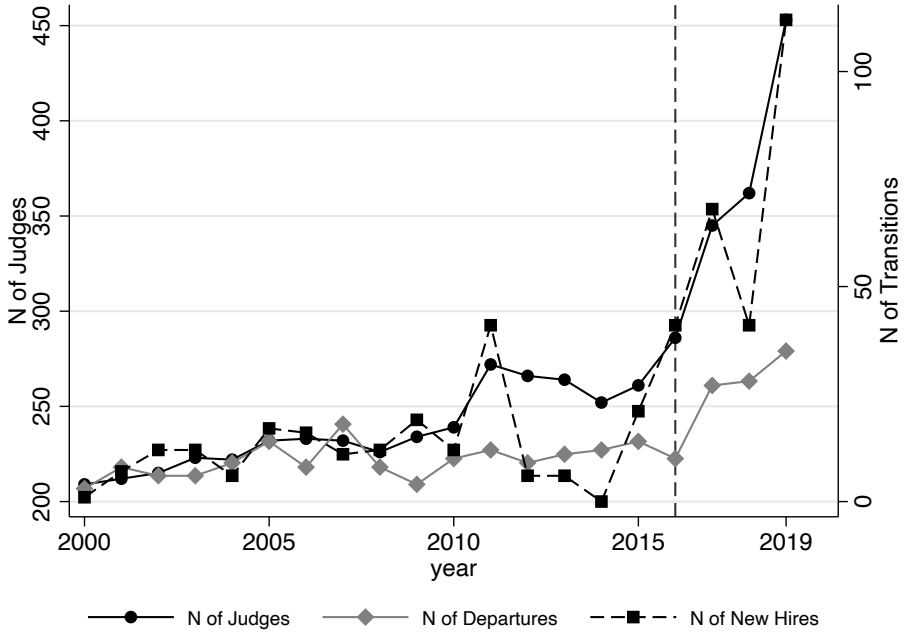
Notes: This table reports summary statistics (mean, standard deviation, minimum, and maximum) for judge-year and court-year variables from 2000 to 2019. In Panel A, *Departure Rate* is an indicator equal to 1 if a judge leaves the bench in that year; *Completed Cases* is the number of cases completed by the judge that year; *Share Below Threshold* equals 1 if the judge completes fewer than 700 cases; *Judicial Experience* is years on the bench; and *Female* and *White* are name-based classifications. In Panel B, *Share of Judges Departing* is the fraction of judges at a court who depart in that year; *Share of Judges Below Threshold* is the fraction of judges at a court completing fewer than 700 cases; *Average Judicial Experience* is the mean years on the bench across judges at the court; and *Case Completions Volume* is the total number of cases completed at the court in that year.

Panel B presents court-level statistics for the immigration courts in our sample, covering the same 2000–2019 period.[6] On average, 4.34 percent of judges at a given court left the bench each year, and tenure at the court level averaged 9.8 years. Around 22 percent of judges fell below the case-completion threshold, and courts completed 5,077 cases per year on average.

Figure 1 illustrates trends in the number of immigration judges, new hires, and

[6]We focus on consistently operating courts throughout the sample period to ensure comparability over time.

Figure 1: Number of Judges, New Hires, and Departures



Notes: This figure shows the trends in active judges, departures, and new hires. For context, the figure extends through 2022, while our empirical window ends in 2019.

departures (including both retirements and quits). The vertical dashed line marks 2016—the year before the GAO report recommending case quotas—and serves as the reference point for our event studies. The total number of active judges (dark solid line) rose steadily after 2015. Departures (light gray solid line) also increased, particularly after 2016, and accelerated around the 2018 implementation of the quota policy. The growth in active judges appears to have been driven by a surge in new hires (dashed line), which began in 2014 and accelerated after 2016, outpacing departures. Overall, the patterns in Figure 1 suggest that the policy's likely impact on departures was at least partially offset by hiring. However, hiring more judges may not have helped increase completions when many experienced judges were leaving.

To assess these relationships directly, Table 2 reports key judicial and court-level outcomes before and after the case-completion quota policy. According to the statistics in Panel A, the share of judges failing to meet the quota increased from 33.70 to 36.55 percent ($\Delta = 2.84$ percentage points, $p = 0.072$) and, perhaps not surprisingly, the departure rate nearly doubled, rising from 4.02 to 7.75 percent ($\Delta = 3.73$ percentage points, $p < 0.001$).

Table 2: Changes in Judicial and Court Outcomes Pre- and Post-Quota Policy

Column	(1)	(2)	(3)	(4)
Statistic	Mean Pre	Mean Post	(2) - (1)	P-Value
Panel A: Immigration Judge Level Dataset				
Share of Judges Below Threshold	0.3370	0.3655	0.0284*	0.072
	(0.4727)	(0.4817)		
Departure Rate	0.0402	0.0775	0.0373***	0.000
	(0.1965)	(0.2676)		
N	4026	1160		
Panel B: Immigration Court Level Dataset				
Share of Judges Below Threshold	0.2029	0.3312	0.1283***	0.000
	(0.3023)	(0.3151)		
Share of Judges Departing	0.0385	0.0681	0.0296***	0.004
	(0.1212)	(0.1341)		
Judicial Experience	10.0640	8.4707	-1.5933***	0.000
	(5.0023)	(5.3995)		
Case Completion Volume	4880.20	6060.04	1179.831***	0.000
	(4170.86)	(5270.31)		
N	885	177		

Notes: This table compares outcomes *pre-policy* (2000–2017) and *post-policy* (2018–2019). Columns (1)–(2) report means with standard deviations in parentheses; column (3) is the post–pre difference; column (4) gives two-sided p-values for equality of means. Panel A (judge-year): *Share Below Threshold* equals 1 if the judge completes fewer than 700 cases; *Departure Rate* equals 1 if a judge leaves the bench that year. Panel B (court-year): *Share of Judges Below Threshold* is the fraction of judges at the court completing fewer than 700 cases; *Share of Judges Departing* is the fraction of judges at the court who depart in that year; *Judicial Experience* is the average years on the bench across judges at the court; *Case Completion Volume* is the total cases completed at the court that year. Asterisks denote significance: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

At the court level (Panel B), the share of judges below the threshold increased from 20.29 to 33.12 percent ($\Delta = 12.83$ percentage points, $p < 0.001$), and the share of departing judges rose from 3.85 to 6.81 percent ($\Delta = 2.96$ percentage points, $p = 0.004$). At the same time, average judicial experience declined from 10.06 to 8.47 years ($\Delta = -1.59$, $p < 0.001$), and the volume of case completions rose by about 1,180 cases annually, from 4,880 to 6,060 ($p < 0.001$).

Overall, the descriptive patterns indicate that the quota policy was associated with higher separation rates, greater workload pressure, and lower average judicial experience. These patterns motivate the court-level analysis that follows on workforce dynamics, court efficiency, and adjudicatory outcomes.

4 Empirical Strategy

Our goal is to estimate how the case-quota policy affected immigration judges' separation rates and, by extension, the composition, experience, case completion, and decision-making capacity of immigration courts. We begin by studying judge departures using a difference-in-differences (DiD) design:

$$Y_{j,t} = \alpha + \beta_1 \, BelowThreshold_{j,t} + \beta_2 \left(BelowThreshold_{j,t} \times Post_t \right) + \zeta_j + \psi_t + \epsilon_{j,t}, \quad (1)$$

where $Y_{j,t}$ is an indicator equal to 1 if judge j separates (departs) in year t and 0 otherwise. $Post_t$ equals 1 in the post-policy period, and ζ_j and ψ_t are judge and year fixed effects, respectively. The variable $BelowThreshold_{j,t}$ equals 1 if, absent the policy, judge j would be expected to complete fewer than 700 cases in year t (construction detailed below). In this specification, β_2 is the parameter of interest: it captures the change in separation rates for judges predicted to fall below the 700-case threshold after the policy relative to judges predicted to meet it.

4.1 Constructing Predicted Exposure to the Quota

A key challenge is that realized case completions are endogenous to the quota—after all, the policy aimed to raise completions. Therefore, we derive the *predicted* (counterfactual) case completion for each judge ($\widehat{P}_{j,t}$) based only on pre-policy information to construct the $BelowThreshold_{j,t}$ indicator.

Let $C_{j,t}$ denote *realized* annual case completions for judge j in year t. To reduce transitory noise, we form a centered five-year moving average, $\widetilde{C}_{j,t} \equiv \mathrm{MA}_5(C_{j,t})$ (two lags and two leads). We then estimate judge-specific linear trends using only pre-policy years (2000–2016):

$$\widetilde{C}_{j,t} = \alpha + \zeta_j + \theta_j \, t + u_{j,t} \qquad (t \leq 2016), \quad (2)$$

where ζ_j are judge fixed effects and θ_j are judge-specific slopes. Using the estimates from (2), we generate for all years the *predicted counterfactual* (absent the quota) completions

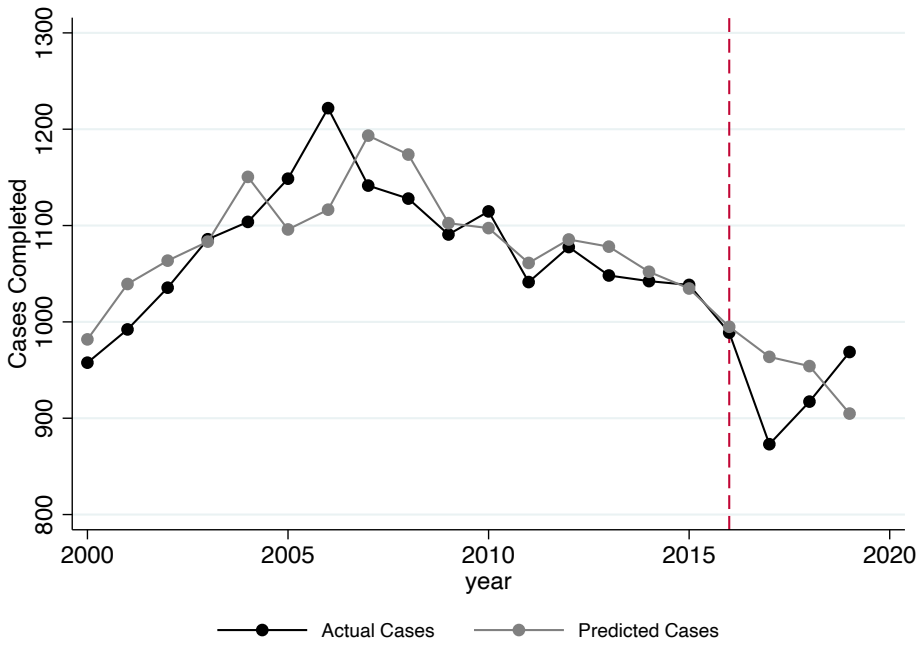
$$\widehat{P}_{j,t} \equiv \widehat{\alpha} + \widehat{\zeta}_j + \widehat{\theta}_j\, t,$$

and define exposure as

$$BelowThreshold_{j,t} \equiv 1\{\widehat{P}_{j,t} < 700\}.$$

That is, $C_{j,t}$ stands for actual completions; $\widetilde{C}_{j,t}$ is the smoothed moving-average of $C_{j,t}$ used to estimate pre-policy trends; and $\widehat{P}_{j,t}$ are out-of-sample predictions from those pre-policy trends.

Figure 2: Average Number of Cases Completed: Actual vs. Smoothed and Predicted



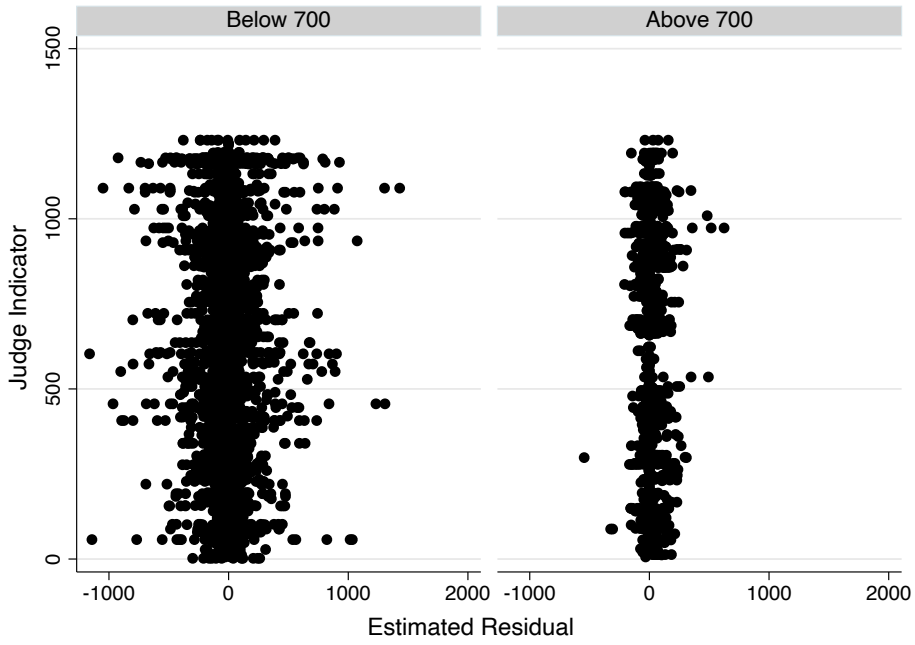
Notes: The figure overlays (i) actual completions $C_{j,t}$ and (ii) the centered five-year moving average $\widetilde{C}_{j,t}$.

4.2 Assessing Prediction Quality

We validate $\widehat{P}_{j,t}$ by examining *pre-policy* prediction errors, $e_{j,t} \equiv \widetilde{C}_{j,t} - \widehat{P}_{j,t}$, separately for judges predicted to be *below* vs. *above* the quota threshold (i.e., $BelowThreshold_{j,t} = 1$ vs. 0).

Figure 3 shows that the error distributions are centered near zero and look similar across groups, indicating no systematic over- or under-prediction for treated judges before the policy. This supports the credibility of our counterfactual predictions and helps rule out that our exposure measure simply reflects pre-existing forecast bias rather than genuine post-policy responses.

Figure 3: Pre-Policy Prediction Errors by Treatment Status



Notes: The figure plots the distribution of $e_{j,t} \equiv \widetilde{C}_{j,t} - \widehat{P}_{j,t}$ in the pre-policy period (2000–2016), separately for judges with $BelowThreshold_{j,t} = 1$ (predicted completions < 700) and $= 0$ (predicted ≥ 700). $\widetilde{C}_{j,t}$ is the centered five-year moving average of actual completions used to estimate the pre-policy trends in equation (2); $\widehat{P}_{j,t}$ are out-of-sample predictions from that pre-policy fit. Distributions centered near zero and similar across groups indicate no systematic, group-specific forecast bias prior to the quota.

Additionally, we assess the pre-policy covariate balance between judges predicted to be above and below the threshold using 2000–2016 data (Table 3). As shown therein, treated and control judges appear similar with respect to gender, race/ethnicity, and experience. Observed differences are small and not statistically significant. The balanced covariates, combined with the use of judge fixed effects, supports the comparability of the two judge groups pre-policy.

Table 3: Differences in Characteristics between Treatment and Control Groups, 2000–2016

Characteristic	Control	Treatment	Difference	P-Value
Female	0.3429	0.3422	-0.0007	0.691
			(0.0018)	
White	0.8620	0.8622	0.0002	0.803
			(0.0009)	
Black	0.0454	0.0453	-0.0000	0.717
			(0.0001)	
Hispanic	0.0798	0.0798	-0.0000	0.939
			(0.0007)	
Asian	0.0126	0.0125	-0.0001	0.686
			(0.0003)	
Experience	10.1492	10.1035	-0.0460	0.441
			(0.0597)	
N	2968	1058		

Notes: This table presents covariate differences between judges predicted to fall above or below the case quota threshold during the pre-policy period. Each row is estimated via a separate OLS regression with year fixed effects. Standard errors are clustered at the judge level.

4.3 Predicted Below-Threshold Descriptive Statistics

Table A1 mirrors Tables 1 and 2 using the predicted-below-threshold measure that we use in the estimations. Importantly, because predictions are built only from pre-policy (2000–2016) completions, they are not mechanically affected by the 2018 quota.

Panels A–B report unconditional averages over 2000–2019. At the judge-year level, 31.70 percent of observations are predicted to fall below the 700-case quota. At the court-year level, the mean predicted share of judges below the threshold is 18.95 percent.

Panels C–D contrast pre-policy (2000–2016) and post-policy (2017–2019) periods. The judge-level predicted-below-threshold rate rises from 26.27 percent to 50.51 percent ($\Delta = 24.23$ percentage points, $p < 0.001$). At the court level, the predicted share increases from 13.12 percent to 48.09 percent ($\Delta = 34.96$ percentage points, $p < 0.001$).

In what follows, we rely on the constructed measures to gauge the responsiveness to the quota policy by judges, as well as its impact on the composition and functioning of immigration courts.

5 Main Findings

5.1 Judicial Departures Following the Quota Policy

Our primary goal is to assess how the adoption of the quota policy affected the departure of immigration judges. To that end, we estimate equation (1) with judge fixed effects (ζ_j), which net out time-invariant judge characteristics (e.g., adjudication style), and year fixed effects (ψ_t), which absorb nationwide shocks, including contemporaneous federal policies and macroeconomic conditions.[7] Standard errors are clustered at the judge level to allow for serial correlation within judges over time.

We estimate equation (1) on three nested samples to make the policy comparison transparent. Table 4 reports effects on the probability that a judge separates from the bench in a given year (percent per year). Treatment is defined as *Predicted Below 700* $\times$ *Post* (2017–2019), where the "predicted" component uses only each judge's pre-policy completions. In this manner, we fix treatment status before the quota and avoid mechanically baking post-2017 outcomes into the regressor.

Column 1 includes all active judges (75 cases/year), matching EOIR reporting and providing the broadest view of how the bench responded when the quota arrived. The estimate is +7.36 percentage points per year off a 4.02 percent per-year baseline, moving separations from roughly 1 in 25 judge-years per year to about 11.38 percent per year (1 in 9 per year). An important limitation of this sample, however, is the fact that it pools long-tenured judges with "quota-era" hires whose selection and assignments may be correlated with the policy environment.

To address this shortcoming, column 2 restricts to judges with at least three years of experience. This cut reduces learning-curve (ramp-up) effects that depress early throughput and improves the quality of the pre-policy prediction that defines treatment.

[7] Some may worry that estimated effects reflect broader Trump-era dynamics. Year fixed effects absorb national-level changes affecting all judges simultaneously. Identification relies on within-year differences between judges predicted to be below versus above the quota.

Table 4: Quota Policy Impact on Immigration Judge Departures

Specification / Column	(1)	(2)	(3)
Below Threshold × Post	0.0736**	0.1071**	0.0861**
	(0.0352)	(0.0448)	(0.0398)
Min. 3 Yr. Experience	No	Yes	No
Started Before 2016	No	No	Yes
DV Mean	0.0402	0.0432	0.0406
Observations	5,058	3,806	4,628
R-squared	0.1765	0.1728	0.1650

Notes: Unit of analysis is a *judge–year*. The dependent variable equals 1 if judge j separates from the bench in fiscal year t. Estimates are linear probability models with judge and year fixed effects (equation (1)). The treatment indicator equals 1 in 2017–2019 for judges predicted—using only *pre-policy* completion histories (2000–2016)—to complete fewer than 700 cases; construction of the counterfactual prediction and the *BelowThreshold* indicator is described in Section 4. Column (1) includes all active judges (defined as ≥75 completions/year); column (2) restricts to judges with at least three years of experience; column (3) restricts to judges who started before 2016. Coefficients are reported in *percentage points*. "DV Mean" reports the unconditional mean separation rate for each column's sample. Standard errors are two-way clustered by judge and year. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The trade-off is potential selectivity: conditioning on survival to three years may tilt the composition toward more stable judges. With that caveat, the estimate rises to +10.71 percentage points per year off a 4.32 percent per-year baseline, implying about 15.03 percent per year (1 in 7 per year). In other words, among experienced judges predicted ex ante to fall below 700, the post-quota annual risk of separation is more than triple the baseline.

Finally, in column 3, we experiment with keeping those judges who started before 2016. Fixing entry well ahead of the 2018 quota—before performance metrics and backlog pressures were widely foregrounded (e.g., GAO discussions)—limits endogeneity from quota-era hiring. This is our preferred sample since it allows for the most policy-exogenous comparison. The estimate is +8.61 percentage points per year off a 4.06 percent per-year baseline, taking separations from 1 in 25 per year to 1 in 8 per year.

Overall, the empirical evidence points to judges whose predicted case completion fell below the 700-case benchmark becoming substantially more likely to separate from the bench after the announcement of the performance metric. The impact was largest among experienced judges and remained strong when we focus on judges hired before 2016, for

whom concerns about quota-era selection and reassignment should be minimal.

5.2 Identification and Robustness Checks

5.2.1 Event Study Analysis: Timing and Dynamics of Departures

A key identifying assumption of our empirical strategy is that departure trends for judges above and below the quota threshold would have evolved similarly in the absence of the policy. To examine the plausibility of this assumption and explore the timing of effects, we estimate a dynamic event study specification given by:

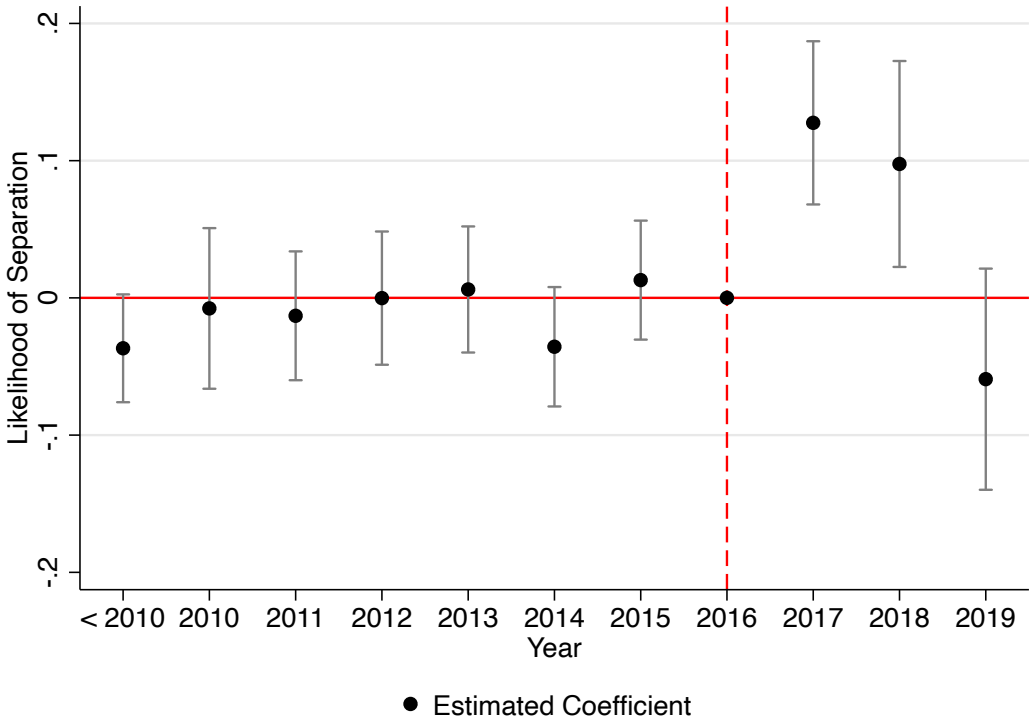
$$Y_{j,t} = \alpha + \sum_{\kappa=2000, \kappa \neq 2016}^{2019} \beta_\kappa BelowThreshold_{j,t} \times 1\{t = \kappa\} + \zeta_j + \psi_t + \epsilon_{j,t} \tag{3}$$

where β_κ captures the year-specific differential in departure rates between treated and control judges, relative to the reference year 2016. All models include judge and year fixed effects, and standard errors are clustered at the judge and year level. For robustness, the continuous specification also includes judge-specific linear trends.

Figure 4 reports dynamic treatment effects using the binary predicted-below-700 indicator. The specification includes judge and year fixed effects, excludes judges appointed after 2016, and uses FY2016 as the reference year. As shown therein, pre-policy coefficients (2000–2016) are small and statistically indistinguishable from zero. A joint test fails to reject equality to zero ($p = 0.44$), supporting the assumption of parallel trends. In contrast, beginning in 2017, the separation probability for treated judges rises by about 12 percentage points per year. That likelihood peaks in 2018 at roughly 15–18 percentage points per year, and remains elevated in 2019 at about 10–13 percentage points per year (with less precision in 2019). Given a baseline departure rate of approximately 4 percent per year, the 2018 peak implies separation rates in the mid-teens (about 1 in 6–7 per year), moderating in 2019 but staying well above baseline (about 1 in 8–10 per year).

Overall, the event study shows a sharp post-policy discontinuity, with an immediate jump in the first full year and a persistent, though smaller, elevation thereafter. The results are consistent with the performance metrics shock being followed by continued

Figure 4: Event Study: Quota Policy Impact on Immigration Judge Departures



Notes: Points plot event–time coefficients from equation (3) using the binary treatment (predicted below–700 cases). Judges appointed after 2016 are excluded. FY2016 is the omitted year; the "<2010" bin averages 2000–2009. Models include judge and year fixed effects; treatment status is defined using only pre-policy completions (2000–2016). Vertical lines show 95 percent confidence intervals based on two-way clustered standard errors (judge and year).

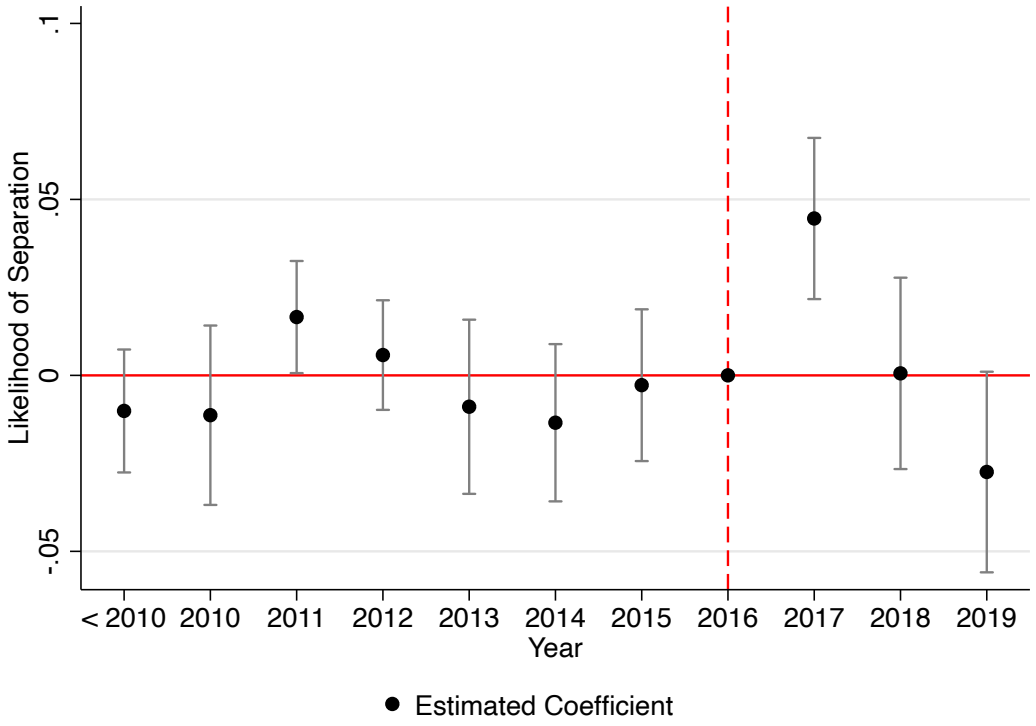
pressure rather than a one-time adjustment.

5.2.2 Robustness Check: Using a Continuous Measure of Caseload Shortfall

We replicate this dynamic analysis using a continuous measure of treatment intensity, defined as the number of standard deviations a judge's caseload falls below or above the 700-case quota. Figure 5 plots the estimated event-time coefficients, where the treatment effect captures the marginal increase in the likelihood of departure per standard deviation shortfall.

Pre-policy coefficients (2010–2016) hover around zero and, jointly, are not statistically different from it, supporting parallel trends. Following the GAO's recommendation to adopt performance metrics in 2017, the first post-period shows an immediate and sizable response. Appendix Table A2 indicates that a one-standard-deviation larger shortfall is

Figure 5: Continuous Event Study: Quota Policy Impact on Immigration Judge Departures



Notes: Points plot event–time coefficients $\{\beta_\kappa\}$ from equation (3) using a *continuous* exposure measure—the standardized caseload shortfall relative to the 700–case quota (constructed with pre-policy completions, 2000–2016; one unit = one s.d.). Judges appointed after 2016 are excluded. Fiscal year 2016 is the omitted reference year; the "<2010" bin averages 2000–2009. Models include judge and year fixed effects and judge-specific linear trends. Vertical bars show 95 percent confidence intervals from standard errors two-way clustered by judge and year. Coefficients are percentage-point effects for a one-s.d. larger shortfall.

associated with a 4.45 percentage points per-year increase in the separation probability in 2017. This marginal effect is large relative to the sample's baseline of roughly 4 percent per year (about 1 in 25 judge-years). By 2018, the continuous-treatment effect is essentially zero, and by 2019, it turns negative. This pattern is consistent with *dynamic selection*. The 2017–2018 shock prompts the marginal leavers—judges most sensitive to performance pressure—to exit first. By 2019, the remaining treated judges (those predicted, ex ante, to fall below 700) are a selected set of stayers with lower underlying separation hazards: greater attachment to the bench, different assignment/court mixes, or better adaptation to performance management—an explanation we examine next. Because each event-time coefficient is estimated on the *risk set* that survives to that year, the changing composition

naturally attenuates the average effect and can flip it negative,[8] as we observe by 2019.

5.2.3 Assessing the Adaptation Among Judges Who Remain

Having shown that judges predicted to fall below the quota were more likely to separate, we turn to those who remained. A stated aim of the policy and a mechanism highlighted by the continuous-treatment results. Figure 6 plots actual and predicted case completions for continuing judges, where predictions are constructed from pre-policy performance trends. By construction, "treated" judges are those whose *predicted* completions fall below the 700-case threshold.

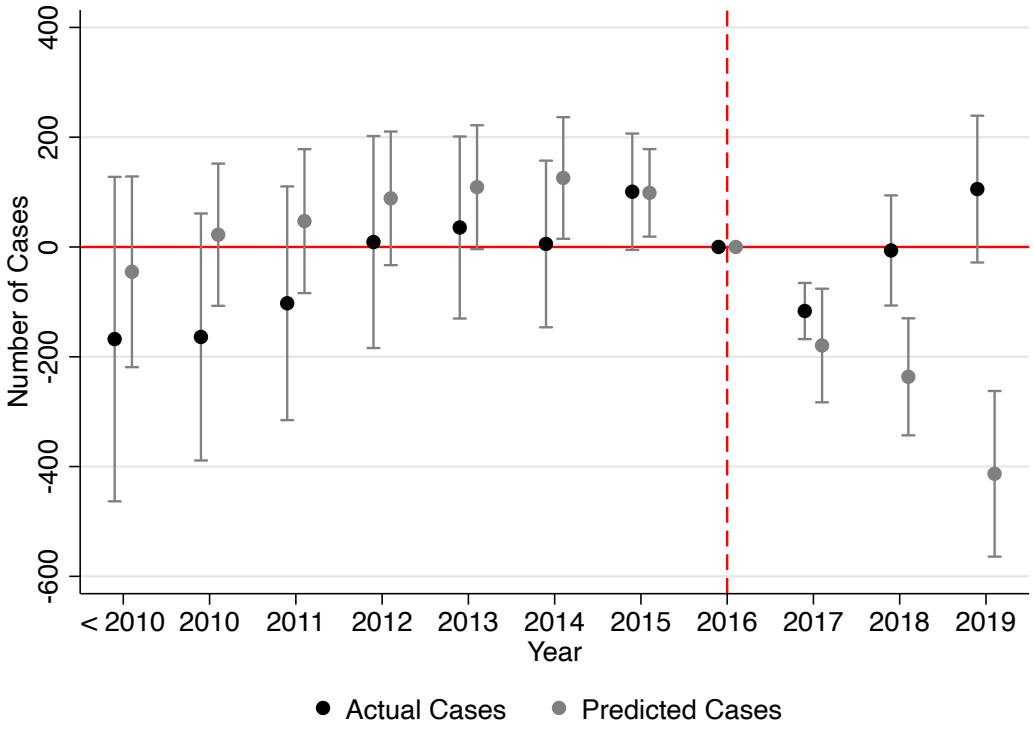
Two patterns stand out. First, before the quota (through 2016), the actual and predicted series move together and their confidence bands overlap, signaling no meaningful pre-policy gaps. Second, beginning in 2018, the lines diverge: actual completions rise while predicted completions fall, with statistically meaningful gaps in 2018–2019. Importantly, judges predicted to miss the threshold either left the bench (our separation results) or, if they stayed, increased throughput once performance metrics were in place. This within-judge adjustment helps explain the event-study time profile for separations: a sharp spike when the quota arrives, followed by attenuation as the remaining treated judges adapt. In addition, the observed pattern justifies the construction of our exposure measure based on each judge's pre-policy performance and motivates the placebo that follows, which test for comparable divergences absent a policy shock.

5.2.4 Placebo Test

To conclude, we probe identification using a falsified-timing test. Specifically, we set the policy year to FY2014 and re-estimate the event-time coefficients using the binary exposure measure. The goal of this exercise is to examine whether our main results could be driven by spurious correlations in the data rather than the policy itself. If our design

[8]Similar attenuation—and occasional negatives—appear for some subgroups in the heterogeneous analyses, which is exactly what early exits would generate.

Figure 6: Treatment Effects on Actual and Predicted Cases

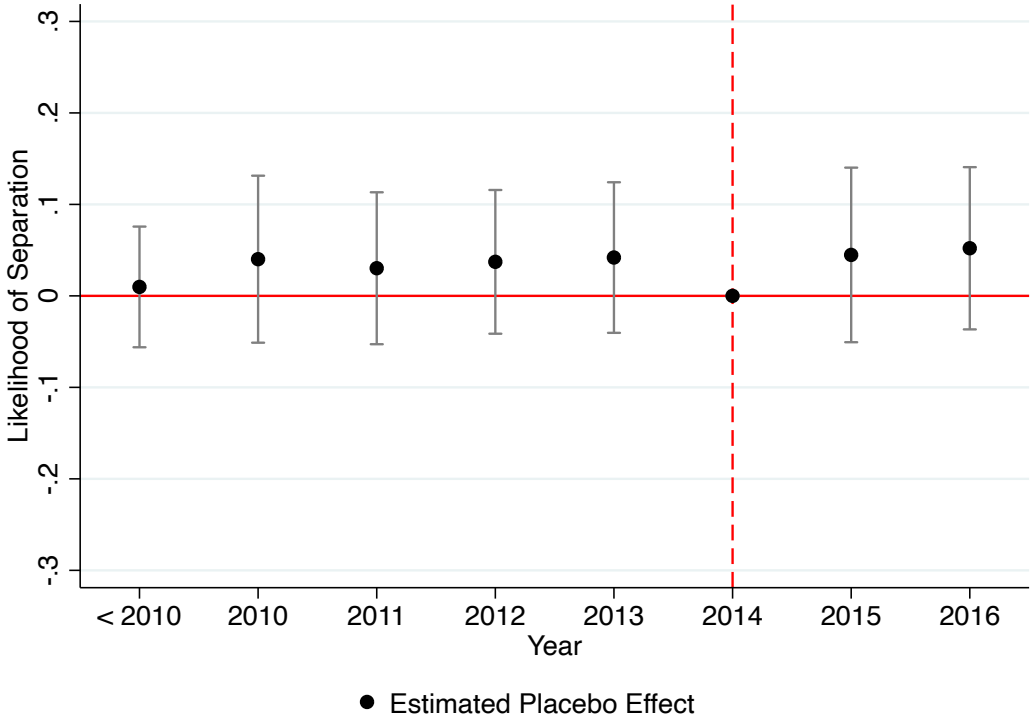


Notes: Points plot event–time coefficients from equation (3) estimated separately for two outcomes: realized case completions $C_{j,t}$ (black) and counterfactual predictions $\widehat{P}_{j,t}$ (gray). $\widehat{P}_{j,t}$ are obtained by extrapolating judge–specific pre-policy (2000–2016) trends estimated on the centered five-year moving average of completions; treatment status (predicted below 700 cases) is defined using only pre-policy data. The sample includes judges who remain active in year t; judges appointed after 2016 are excluded. Fiscal year 2016 is the omitted reference year; "<2010" averages 2000–2009. Models include judge and year fixed effects. Vertical bars show 95 percent confidence intervals from standard errors two-way clustered by judge and year.

were simply capturing random variation or pre-existing trends, we would expect to see an apparent "effect" emerge even when the policy year is artificially shifted.

As shown in Figure 7, the estimated coefficients are flat and statistically indistinguishable from zero throughout the entire pre- and post-2014 period. There is no visible discontinuity around the placebo treatment date, in contrast to the clear break observed after the actual policy introduction in 2017. The absence of spurious effects around the false intervention year provides reassurance that the dynamics documented in our main analysis are not artifacts of chance timing or unrelated shocks. As such, Figure 7 strengthens the credibility of our identification strategy and supports the interpretation of the post-2017 estimates as causal effects of the policy.

Figure 7: Placebo Test: False Timing for Policy Introduction



Notes: This placebo re-assigns the policy start to FY2014 and re-estimates the event-time regression using the binary exposure. The point labeled "< 2010" averages 2000–2009. The outcome is the likelihood of separation. The sample excludes judges appointed after 2016. Vertical bars show 95% confidence intervals computed with two-way clustered standard errors (judge and year). A flat series with no break at 2014 indicates no spurious effect under false timing.

5.3 Heterogeneous Effects by Gender and Minority Status

To conclude, we explore whether the quota policy had differential impacts on the departure likelihood of immigration judges by gender and race. Understanding these heterogeneities is critical, as both theoretical frameworks and empirical studies suggest that demographic characteristics shape judicial behavior and institutional vulnerability.

Several theories help explain why female and minority judges may differ from their male and White counterparts in decision-making and institutional experiences. The *different voice* theory (Gilligan, 1982; Sherry, 1986) posits that men and women reason and communicate differently, which may influence how they adjudicate cases. *Information* theory (Boyd et al., 2010) suggests that female and Black judges bring distinct life experiences—particularly regarding discrimination—that shape their legal reasoning.

Relatedly, representative theory posits that female and minority judges may act as group representatives, seeking to advance group interests in their rulings (Farhang and Wawro, 2004; Kastellec, 2013).

Empirical studies lend support to these perspectives. For instance, female judges have been found to be more sympathetic to women litigants (Cook, 1981), while Black judges are more responsive in racially salient cases (McClain et al., 2009). More broadly, Boyd (2016) documents that both female and Black judges tend to issue rulings more favorable to plaintiffs.

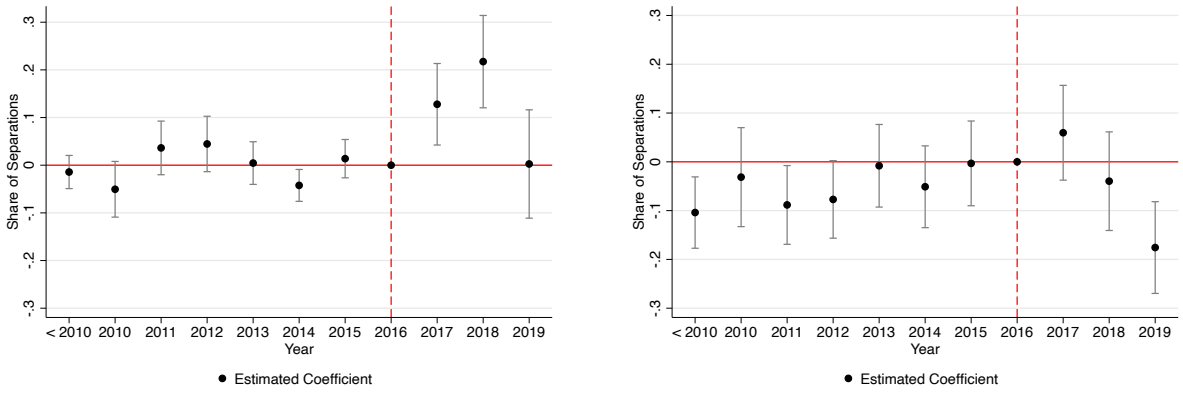
In the immigration context, prior research shows that female adjudicators are 44 percent more likely to grant asylum than male judges, even after controlling for observable characteristics (Menkel-Meadow, 2009). Similar patterns have been documented in Beougher (2016), Chand et al. (2017), Keith et al. (2013), Ramji-Nogales et al. (2007), GAO (2008), and GAO (2016). These findings underscore the importance of examining whether the quota policy disproportionately impacted underrepresented groups on the immigration bench.

5.3.1 Gender Differences

The quota appears to have resulted in selection effects on the extensive margin for women and adjustment on the intensive margin for men. In Figure 8, Panel A, separations among treated *female* judges rose immediately after implementation, reached a peak in 2018, and then eased. In contrast, the corresponding *male* series in Panel B shows no comparable increase in 2017–2018, with the likelihood of separation among remaining male judges turning negative by 2019. According to the estimates in Appendix Table A2, during the first two years after the policy's announcement, the separation of female immigration judges rose from a very low baseline (about 2.5 percent per year) into the mid–teens and, later, low twenties. For male immigration judges, the post-2017 estimates hovered around zero before dropping below baseline by 2019.

Figure 9 complements the separation outcomes by shedding some light on how remaining judges adapted to the policy. Before the policy, actual completions track

Figure 8: Event Study: Quota Policy Impact on Departures by Judge Gender

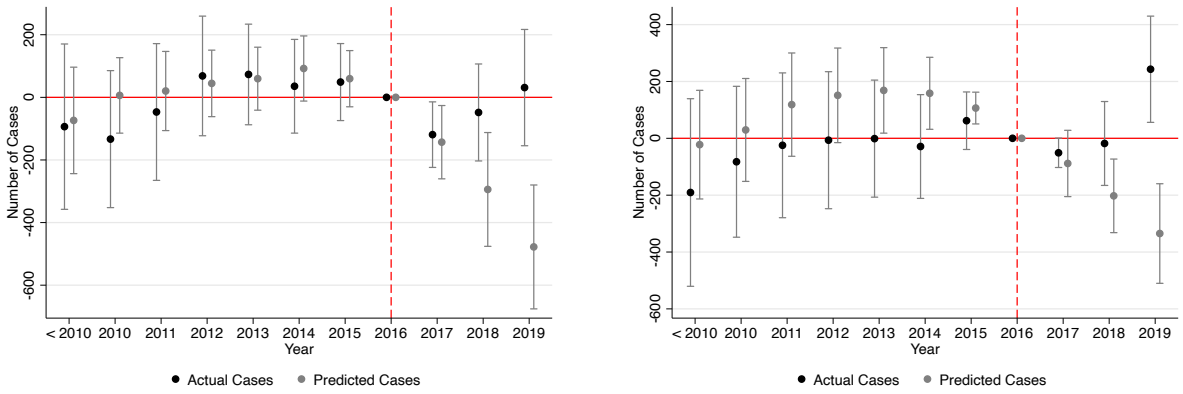


(a) Female Judges

(b) Male Judges

Notes: Points plot event–time coefficients from equation (3) estimated separately by gender, using the *binary* exposure (predicted below 700 cases, defined with pre-policy completions 2000–2016). Fiscal year 2016 is the omitted reference year; the "<2010" bin averages 2000–2009. Models include judge and year fixed effects; judges appointed after 2016 are excluded. Vertical bars show 95 percent confidence intervals from two-way clustered standard errors (judge and year). Coefficients are percentage-point effects on the probability of separation.

Figure 9: Event Study: Quota Policy Impact on Case Completion by Judge Gender



(a) Female Judges

(b) Male Judges

Notes: Points plot event–time coefficients from equation (3) estimated separately by gender for two outcomes: realized case completions $C_{j,t}$ (black) and counterfactual predictions $\widehat{P}_{j,t}$ (gray). $\widehat{P}_{j,t}$ extrapolate judge–specific pre-policy trends estimated on 2000–2016 completions (see Section 4); treatment is the *binary* exposure (predicted below 700 cases, defined using only pre-policy data). Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include judge and year fixed effects; judges appointed after 2016 are excluded. Vertical bars show 95 percent confidence intervals from two-way clustered standard errors (judge and year). Coefficients are changes in the number of cases per judge–year.

pre-policy predictions for both genders (Panels A and B). After 2017, actual completions exceeded predicted levels while the predictions themselves fell, consistent with within-judge adjustment to the benchmark. The divergence is present for women, but is significantly larger and more sustained for men.

Overall, the same performance pressure appears to map into different costs by gender, generating two margins: many women exit early (with those who remain adjusting upward), whereas men largely remain and increase output more sharply and persistently. This differential response shifts bench composition—fewer women, more men who adapted—and mechanically produces the spike in female separations and the subsequent dip for men in Figure 8, with potential consequences for adjudication and institutional outcomes (Gilligan 1982; Sherry 1986; Farhang and Wawro 2004; Kastellec 2013; Cook 1981; Boyd et al. 2010; Boyd 2016).

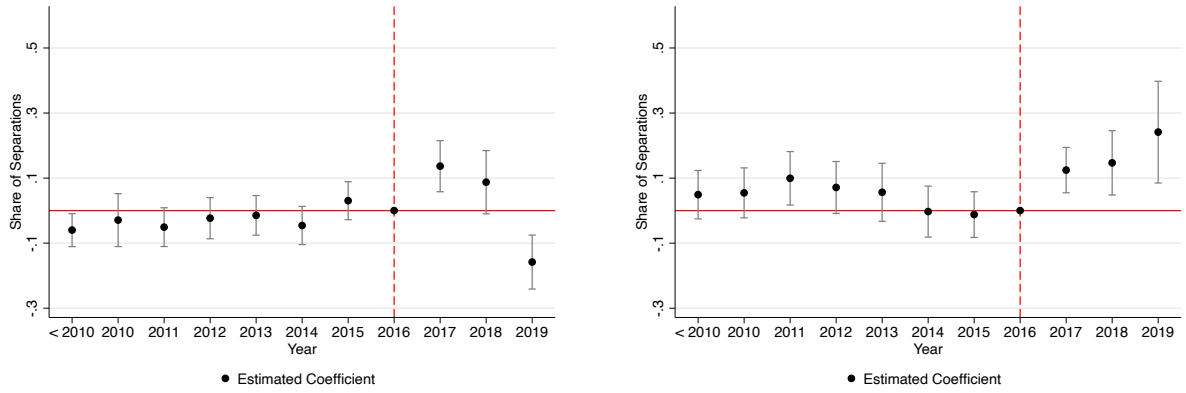
5.3.2 Racial Differences

We next examine whether the quota's effects vary by race. Figure 10 shows two trajectories. Among White judges (Figure 10, Panel A), separations rose at implementation but fell below baseline by 2019—an early exit response followed by retention among those who remained. In levels, the first post-policy year rose separations from roughly 3 percent per year to the mid-teens (1 in 6 judge-years) before easing and turning negative by 2019 (Appendix Table A2, Panel D).

Among non-White judges (Figure 10, Panel B), separations rose after the policy and continued to rise through 2019. In levels, this represented a shift from about 4–4.5 percent per year to the high-teens in 2018 and around the high-20s by 2019 (1 in 3–4 according to Appendix Table A2, Panel E).

Looking at the behavioral response of stayers in Figure 11, we find evidence of both groups increasing case completions relative to their pre-policy predictions, with a larger late-period divergence for non-White judges. A consistent interpretation is dynamic selection for White judges —early exits by those most sensitive to the target, leaving a lower-hazard group that adapts— and persistent pressure for non-White judges

Figure 10: Event Study: Quota Policy Impact on Departures by Judge Race



(a) White Judges

(b) Non-White Judges

Notes: Points plot event–time coefficients from equation (3) estimated separately for White (panel a) and Non-White judges (panel b), using the *binary* exposure (predicted below 700 cases, defined with pre-policy completions 2000–2016). Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include judge and year fixed effects; judges appointed after 2016 are excluded. Vertical bars show 95 percent confidence intervals from two-way clustered standard errors (judge and year). Coefficients are percentage-point changes in the probability of separation.

—adaptation occurs, but the effective costs of meeting the benchmark remain high enough that exits continue.

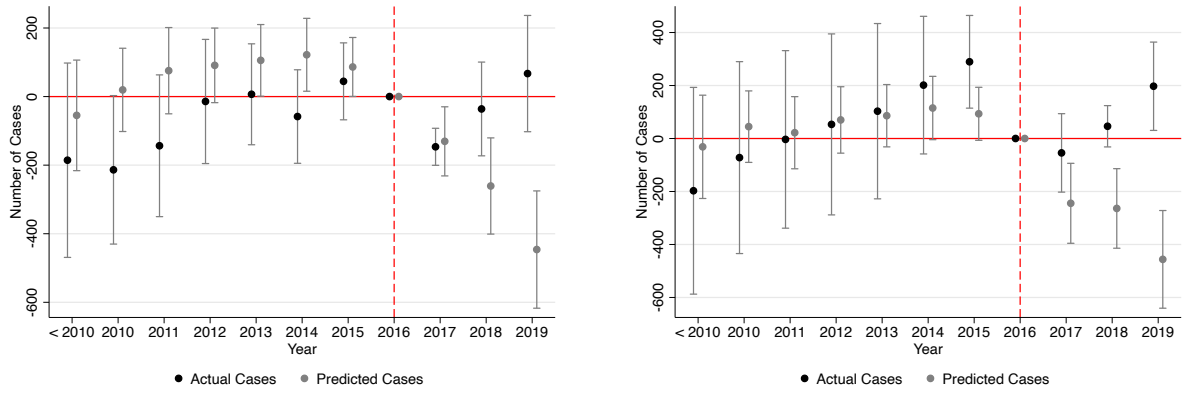
As noted earlier, the composition of the immigration bench appears to have shifted: treated White judges who adapted were likelier to be retained, while treated non-White judges continued to exit. Such re-sorting can affect both representation and institutional performance—by shrinking informational/experiential diversity on panels and thereby influencing deliberation quality, consistency of rulings, error rates, backlog growth, and overall throughput (Farhang and Wawro 2004; Kastellec 2013; see also Boyd 2016).

6 Implications for Justice Delivery

6.1 Judge Separations and Loss of Experience

We next examine how quota-induced separations reshaped court capacity. We focus on the *expertise* channel: if exits disproportionately involved experienced judges, courts lose institutional knowledge that can affect throughput and adjudicatory outcomes. We

Figure 11: Event Study: Quota Policy Impact on Case Completion by Judge Race



(a) White Judges

(b) Non-White Judges

Notes: Points plot event–time coefficients from equation (3) estimated separately by race—White (panel a) and Non-White (panel b)—for two outcomes: realized case completions $C_{j,t}$ (black) and counterfactual predictions $\widehat{P}_{j,t}$ (gray). $\widehat{P}_{j,t}$ extrapolate judge-specific pre-policy trends estimated on 2000–2016 completions (see Section 4); treatment is the *binary* exposure (predicted below 700 cases, defined using only pre-policy data). Fiscal year 2016 is the omitted reference year; "$<$ 2010" averages 2000–2009. Models include judge and year fixed effects; judges appointed after 2016 are excluded. Vertical bars show 95 percent confidence intervals from two-way clustered standard errors (judge and year). Coefficients are changes in the number of cases per judge–year.

measure expertise as average years on the bench at the court–year level and trace its event–time response to exposure; we then relate exposure-driven changes in experience to completions and decision patterns.

We aggregate the data to the court–year level and track four outcomes: (i) the share of judges separating, (ii) the average judicial experience of judges who remain, (iii) total case completions, and (iv) the share of cases decided for respondents. Exposure is measured continuously as the share of judges in court c whose *predicted* completions fell below the quota threshold in year t, denoted $Share_{c,t}$. Our event–study specification is:

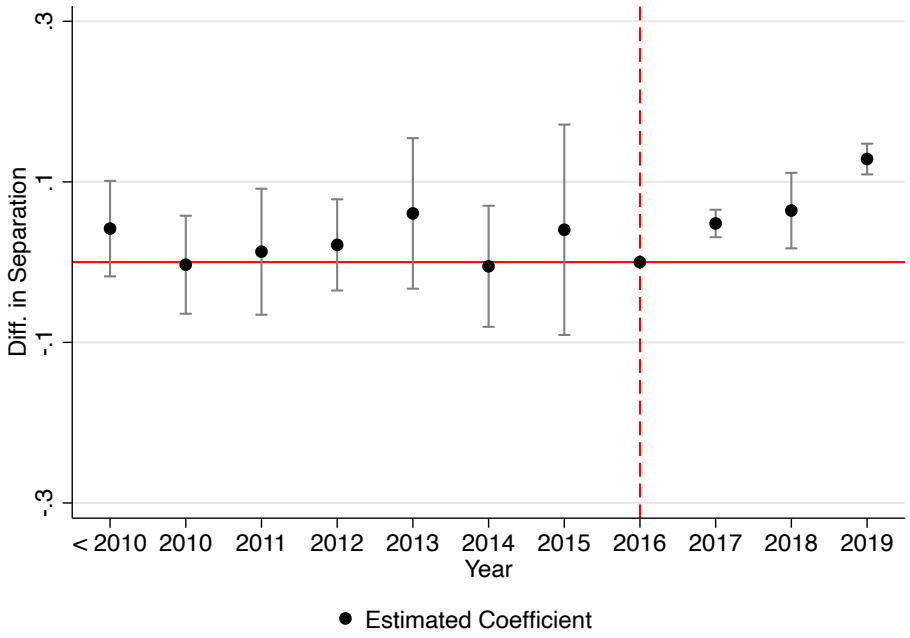
$$Y_{c,t} = \alpha + \sum_{\kappa=2000,\ \kappa\neq 2016}^{2019} \beta_\kappa \left(Share_{c,t} \times \mathbf{1}\{t = \kappa\} \right) + \zeta_c + \psi_t + \epsilon_{c,t}, \tag{4}$$

where $Y_{c,t}$ is each court-level outcome. The coefficients β_κ trace the effect of exposure in fiscal year κ relative to 2016 (the omitted year); court and year fixed effects absorb time-invariant court traits and common shocks, respectively.

6.1.1 Court-Level Judge Separations

We estimate court-level effects of the quota on separations using equation (4). Figure 12 plots event–time coefficients where the outcome is the court's annual separation rate and exposure is the court's *pre-policy* share of judges predicted to complete fewer than 700 cases.

Figure 12: Event Study — Quota Policy Impact on Court-Level Departures



Notes: Points plot event–time coefficients from equation (4) for the court-level separation rate (share of judges who separate in court c and year t). Exposure is the court-level share of judges predicted—using only pre-policy completions (2000–2016)—to complete fewer than 700 cases (see Section 4). Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include court and year fixed effects. Vertical bars show 95 percent confidence intervals from two-way clustered standard errors (court and year).

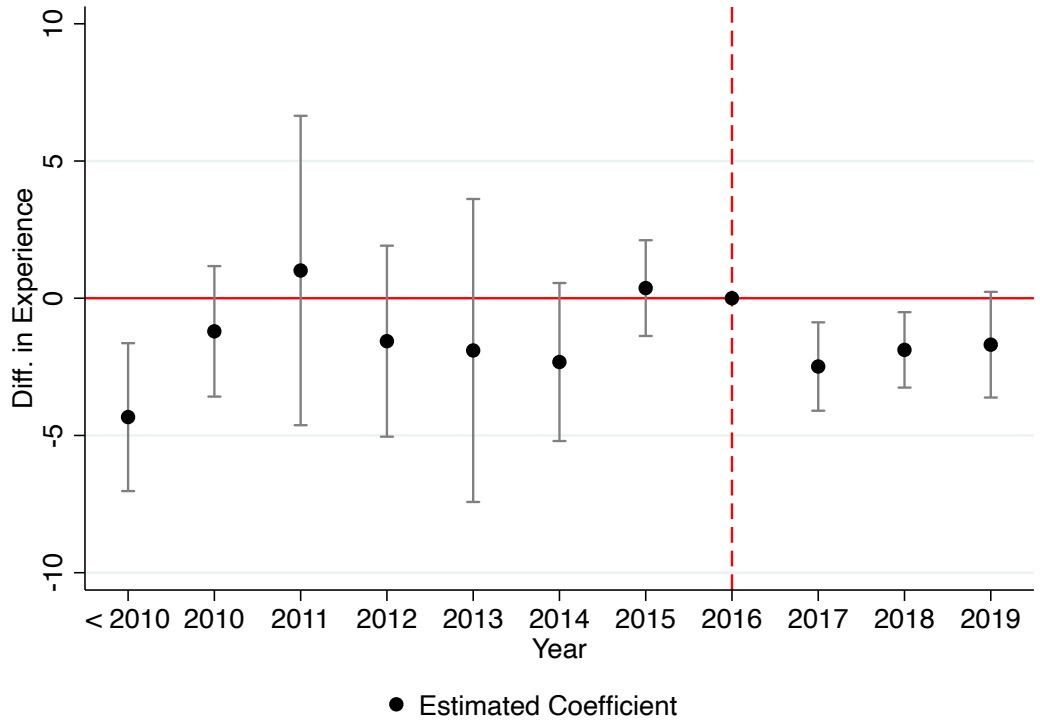
As documented therein, there is no evidence of differential pre-trends, but a sustained post-policy divergence with attrition rising from FY2017 through FY2019. Pre-policy coefficients are flat. After 2016, separations rise steadily. Using the TWFE levels in Appendix Table A3, Panel A, moving a court from zero to full exposure (with the *Share* rising from 0 to 1) would increase the separation rate by 4.8 percentage point in 2017 (3.85 percent $\rightarrow$ 8.65 percent), 6.4 percentage point in 2018 (to 10.25 percent), and 12.8 percentage point in 2019 (to 16.65 percent) relative to the pre-policy mean of 3.85 percent (DV mean, Panel A). Equivalently, a more modest 10 percentage point increase in

exposure is associated with +0.48 percentage point (2017), +0.64 percentage point (2018), and +1.28 percentage point (2019). In sum, where the quota was most binding, courts experienced a higher exit hazard that persisted within the study window.

6.1.2 Judicial Experience

Of particular relevance when understanding how departure rates may have impacted court functioning relates to the impact of having a higher share of "treated" immigration judges on the court's judicial experience. Figure 13 plots event–time estimates for the average years of experience among sitting judges at the court level. Pre-policy coefficients are flat—consistent with parallel trends—and average experience declines steadily after implementation, remaining depressed throughout the post-policy period.

Figure 13: Event Study: Quota Policy Impact on Judicial Experience at the Court Level



Notes: Points plot event–time coefficients from equation (4) for the court-level average judicial experience (years). Exposure is the court-level share of judges predicted—using only pre-policy completions (2000–2016)—to complete fewer than 700 cases (see Section 4). Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include court and year fixed effects. Vertical bars show 95 percent confidence intervals from two-way clustered standard errors (court and year). Coefficients are changes in the court's average years of experience.

Before the policy, courts averaged about 10 years of judge experience (Table 1). In the TWFE estimates (Appendix Table A3, Panel B), the $Share \times Post$ coefficient ranges from -2.72 (2017) to -2.05 (2019) when moving exposure from 0 to 1. This implies that going from zero to full exposure lowers a court's average experience by about 2.1–2.7 years—roughly one-fifth of the pre-policy mean. Similarly, a 10 percentage-point increase in exposure would lower average experience by 0.21–0.27 years (2.5–3.2 months).

In sum, according to Figure 12 and Figure 13, exposure to the quota raised separations and reduced average experience, pointing to accelerated turnover and diminished capacity. In what follows, we explore the implications for court throughput and adjudicatory outcomes.

6.2 Quota Policy Implications on Case Completion

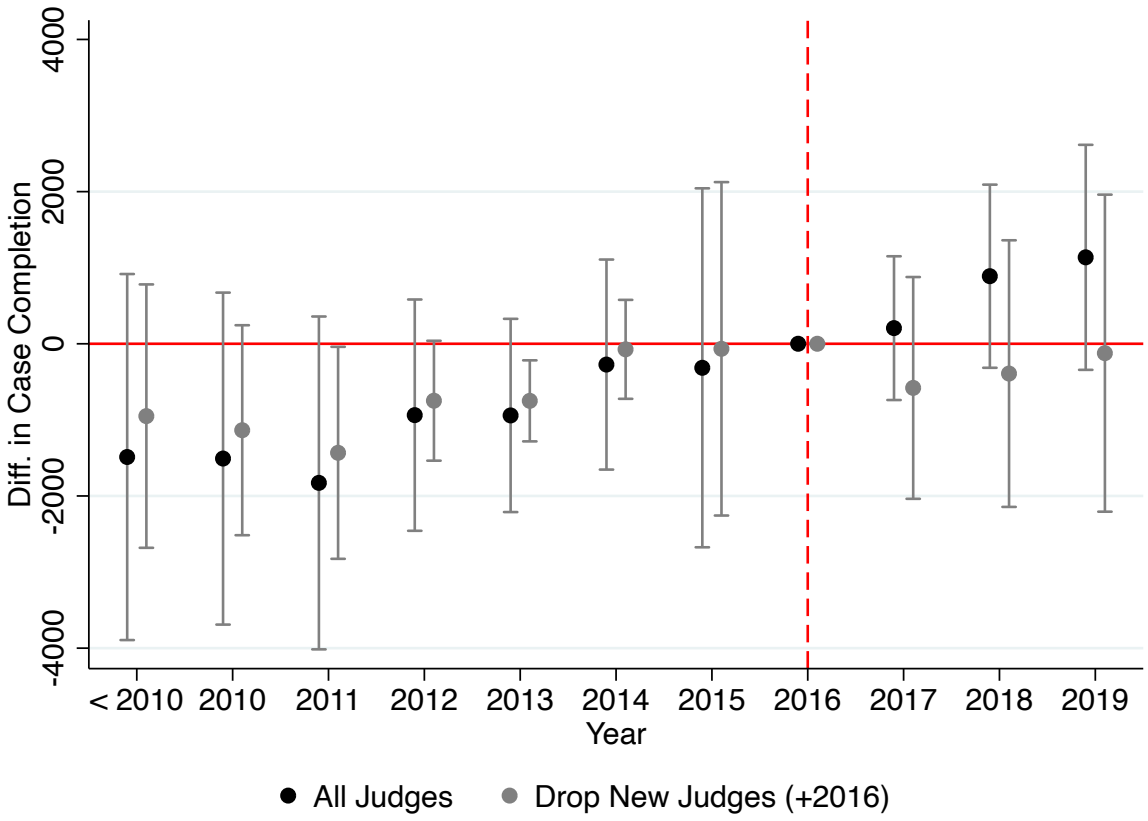
A natural question is whether any unintended staffing consequences were at least offset by higher throughput. In other words: did the quota raise case completions as intended? Figure 14 reports event–study estimates from equation (4) with court-level completions as the outcome, plotting two series: *(i)* all judges (black dot series) and *(ii)* excluding newly appointed judges to abstract from onboarding/learning effects (gray dot series). In both instances, pre-policy coefficients are flat consistent with the assumption of parallel trends. In addition, after implementation, neither series differs statistically from zero. Courts with higher exposure did not complete more cases than otherwise similar courts, whether or not we exclude new judges appointed after 2016 from the calculation.

Appendix estimates provide us with the information needed to scale the effects. According to the figures in Table A3, Panel C, moving from zero to full exposure would translate to roughly -399 cases in 2017, $+235$ in 2018, and $+522$ in 2019. A more modest 10 percentage point change in exposure would result in -40, $+24$, and $+52$ cases in each of those post-policy years. Relative to the pre-policy mean volume per court (Table 1), these shifts are small in magnitude and statistically indistinguishable from zero. Estimates are similarly imprecise when we exclude newly appointed judges (gray dot series), as shown in Panel D of the same table. A full-exposure change corresponds to roughly -580 in 2017,

−392 in 2018, and −123 in 2019, and the corresponding 10 percentage point change in exposure would result in −58, −39, and −12 cases in each of those years, respectively. None of those coefficients is statistically different from zero.

Altogether, the flat event–study paths indicate that the quota did not deliver measurable throughput gains at the court level within our sample window. Any intensive-margin adjustments by remaining judges appear to be offset at the aggregate court level—plausibly due to learning curves for new judges and constraints outside the judge's control (*e.g.*, staffing, interpreter availability, scheduling with DHS counsel)—so higher exposure did not translate into reliably higher completions.

Figure 14: Event Study: Quota Policy Impact on Case Completion



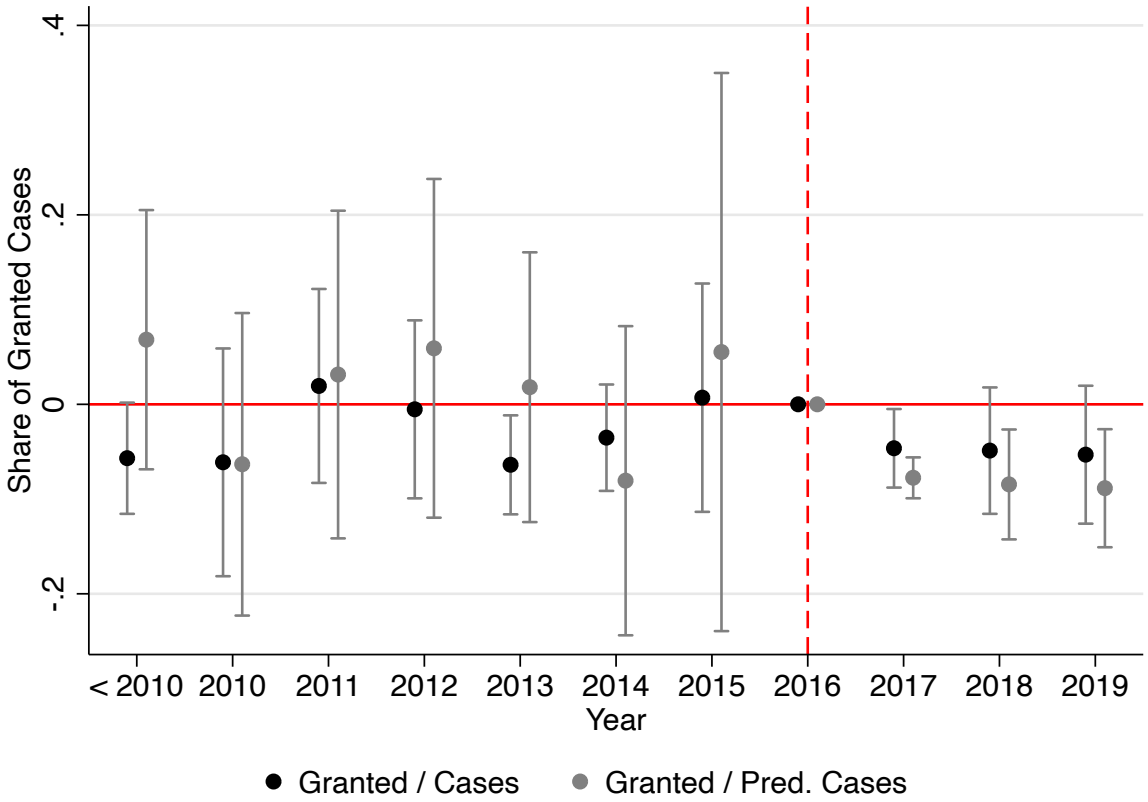
Notes: Points plot event–time coefficients from equation (4) for court-level case completions. Black dots: all judges. Gray dots: excluding newly appointed judges (post-2016 entrants). Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include court and year fixed effects. Vertical bars show 95% confidence intervals from two-way clustered standard errors (court and year).

6.3 Quota Policy Implications on Granted Cases

Having shown that exposure raised separations and reduced experience without commensurate gains in completions, we ask if adjudication itself shifted. To that end, we examine the share of cases granted—a discretionary margin likely to be sensitive to staffing strain.

Figure 15 plots event–time estimates from equation (4) using two denominators: *(i)* actual completions (black dot series) and *(ii)* predicted completions (gray dot series) —the latter fixing throughput at its pre-policy trend. In both instances, pre-policy coefficients are flat consistent with parallel trends. After 2016, grant shares decline with the share of judges impacted by the policy, also in both series.

Figure 15: Event Study: Quota Policy Impact on Granted Cases



Notes: Points plot event–time coefficients from equation (4) for the court-level share of granted cases. Black dots use *actual* completions as the denominator; gray dots use *predicted* completions (constructed from pre-policy trends, 2000–2016). Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include court and year fixed effects. Vertical bars show 95% confidence intervals from two-way clustered standard errors (court and year).

Panels E and F of Appendix Table A3 put scale on these changes. Panel E (Granted/Actual)

yields Post×Share estimates around −0.046, −0.049, and −0.053 in FY2017–FY2019 (only the first is precisely estimated), implying that a 10 percentage point increase in exposure reduces the grant share by roughly 0.5 percentage points. Panel F (Granted/Predicted) produces larger and precisely estimated effects of about −0.078, −0.085, and −0.089, implying a 0.8–0.9 percentage point decline in the grant share per 10 percentage point increase in exposure (all significant).

Altogether, along with the near-zero completion effects in Figure 14 and the imprecise throughput estimates in Panels C and D, the evidence points to a substantive reduction in the probability of granting—rather than a mechanical change in the denominator.

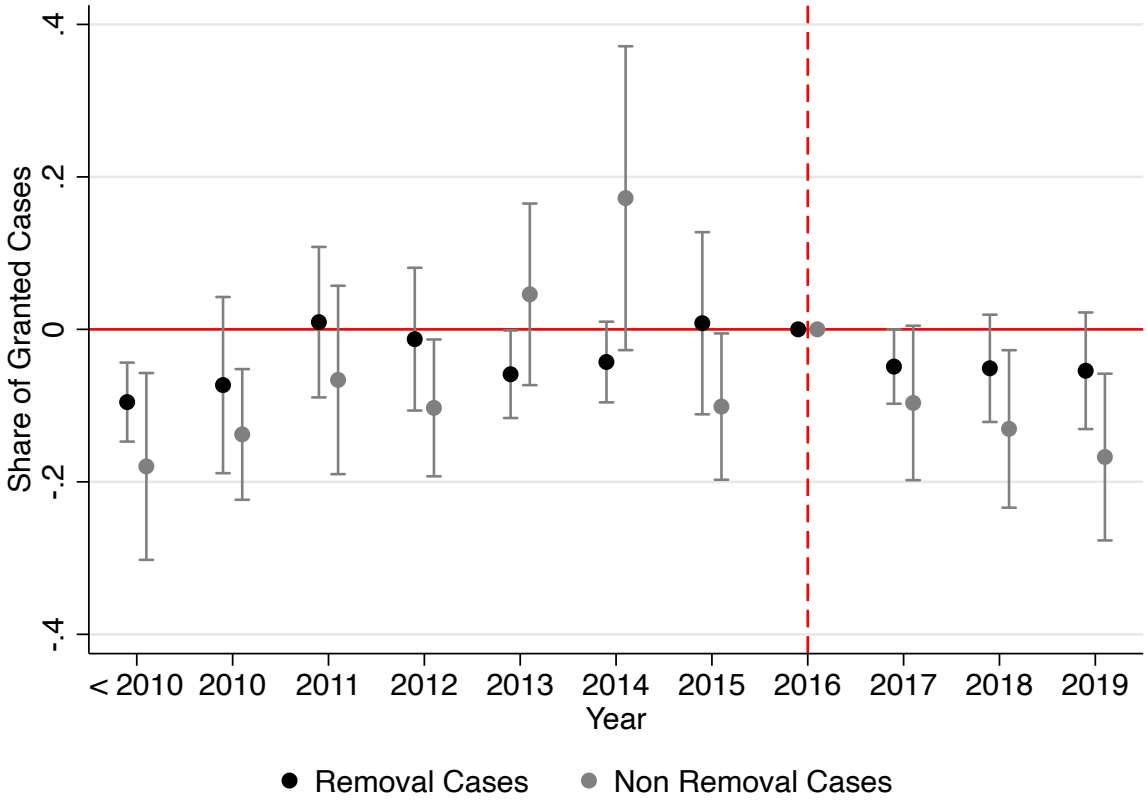
To gauge where the shift occurs in the docket, Figure 16 disaggregates grant shares by case type. Post-2016, grant shares decline with exposure in both removal and non-removal matters. The non-removal series (gray) shows larger negative *point* estimates in 2018–2019 but with wide intervals; the removal series (black) exhibits steadier negatives. Because removal proceedings constitute the overwhelming majority of filings,[9] the aggregate court-level decline in Figure 15 is arithmetically driven by removal cases even though the per–case-type drop is at least as large for non-removal matters. The court-level pattern aligns with judge-level evidence in Blasingame et al. (2023), who document lower asylum grant rates under quotas.

It is worth noting that the observed decline is not a mechanical byproduct of case completions. The previous subsection showed no reliable effect of exposure on the number of cases resolved, and the gray series in Figure 15 (which fixes the denominator at predicted completions) also falls. Hence, the evidence points to a substantive shift in adjudication: migrants became less likely to receive relief even as overall throughput did not rise.

Two mechanisms can generate these patterns: changes in who decides cases (composition) and changes in how remaining judges decide them (behavior). We cannot rule out

[9]See Executive Office for Immigration Review (EOIR). 2023. "Adjudication Statistics: Pending Cases." U.S. Department of Justice; and TRAC Immigration. 2023. "Immigration Court Backlog Tool," Syracuse University.

Figure 16: Event Study: Quota Policy Impact on Granted Cases by Case Type



Notes: Points plot event–time coefficients from equation (4) for the court-level share of granted cases, disaggregated by case type. Black dots: removal cases. Gray dots: non-removal cases. Fiscal year 2016 is the omitted reference year; "< 2010" averages 2000–2009. Models include court and year fixed effects. Vertical bars show 95% confidence intervals from two-way clustered standard errors (court and year).

behavioral adjustments, but two facts point to composition as central. First, the standardized series using *predicted* completions declines alongside the raw series (Figures 15 and 16), so the result is not a denominator artifact. Second, more-exposed courts lost experience (Figure 13) without offsetting gains in throughput (Figure 14). Together, these findings indicate that the policy reshaped the deciding bench, with the reallocation consistent with lower relief rates in the most consequential segment of the docket.[10]

[10]Full two–way fixed-effects estimates with standard errors are reported in Table A3; they mirror the event–study results discussed in the text.

7 Summary and Conclusions

In 2016, with backlogs mounting, the federal government moved toward a quantitative fix for immigration adjudication. The following year the Government Accountability Office urged the adoption of productivity metrics and, in FY2018, the Department of Justice implemented a case-completion quota that conditioned a "Satisfactory Performance" rating on closing at least 700 cases per judge per year. This placed an executive-branch target inside a judicial setting, raising first-order concerns about independence and due process. For our purposes, the quota represents a policy shock—an instance of executive control intruding into the judicial process—allowing us to trace how this type of intrusion affects court composition and, in turn, court functioning. The evidence shows tangible institutional costs—with no offsetting efficiency gains—speaking directly to current concerns about executive encroachment on judicial functions.

We leverage quasi-experimental variation in exposure to the quota threshold based on pre-policy completions to study how the policy reshaped the bench and, in turn, court performance. Three sets of results stand out. *First*, the quota increased exits. Judges most exposed to the threshold became substantially more likely to separate; our preferred estimate implies an 8.6 percentage-point rise in the annual separation probability. Furthermore, the effects proved disproportionately larger among experienced, female, and non-White judges.

Second, courts lost expertise without gaining throughput. At the court level, a 10 percentage-point increase in exposure is associated with roughly 0.5–1.3 percentage points higher separation rates over 2017–2019 and a 0.21–0.27 year decline in average experience. Yet, despite the policy's goal of reducing caseloads by increasing courts' throughput, we find no measurable increase in total completions.

Third, adjudication tilted against granting relief. A 10 percentage-point increase in exposure reduces grant rates by about 0.77–0.90 percentage points (5–6 percent of the 15.6 percent pre-policy mean), with declines visible in both removal and non-removal cases.

Overall, these patterns are consistent with a change-in-court-composition channel. Namely, the quota reshuffled the deciding bench—away from more experienced judges—while

any intensive-margin effort by those who stayed was offset by a learning curve and process constraints. The policy thus delivered neither higher output nor improved access to relief. Instead, it reduced institutional capacity and shifted outcomes against applicants, aligning with judge-level evidence on lower asylum grants under quotas (Blasingame et al., 2023).

Two caveats are important. Our window covers the early quota years; longer-run adjustments or subsequent managerial changes could alter effects. Additionally, while the evidence points to composition as central, we cannot fully separate composition from behavioral responses with available data. These limitations do not change the core conclusion: numerical targets in adjudication can erode capacity and fairness in tandem. In systems that already sit at the boundary of executive control and judicial function, effective reform should emphasize retention of experience, investments in court operations, and evaluation of decision quality—not just higher case counts. Ongoing personnel actions amid record backlogs—see Ainsley et al. (2025)—underscore the stakes and reinforce the central lesson: efficiency reforms must retain experience and evaluate quality, not just volume.

References

Ainsley, Julia, Aaron Gilchrist, Nnamdi Egwuonwu, and Megan Lebowitz, "Justice Department fires multiple immigration judges amid case backlog," NBC News 2025. Accessed: February 15, 2025.

Alvarez, Priscilla, "Immigration Judges Quit in Response to Administration Policies," *CNN, December 27*, 2019, *Available at: https://www.cnn.com/2019/12/27/politics/immigration-judges-resign/index.html.*

Bar, New York City, "Report on the Independence of the Immigration Courts," *Immigration and Nationality Law Committee Task Force on the Rule of Law Task Force on the Independent of Lawyers and Judges. Available at www.nycbar.org*, 2020.

Benner, Katie, "Immigration Judges Express Fear that Sessions's Policies Will Impede Their Work," *New York Times, June 13. https://www.nytimes.com/2018/06/12/us/politics/immigration-judges-jeff-sessions.html*, 2018.

Beougher, Mark Richard., "Who Controls Immigration Judges?: Towards a Multi-Institutional Who Controls Immigration Judges?: Towards a Multi-Institutional Model of Administration Judge Behavior Model of Administration Judge Behavior," *PhD diss. Western Michigan University*, 2016.

Blasingame, Elise N., Christina L. Boyd, Roberto F. Carlos, and Joseph T. Ornstein, "How the Trump Administration's Quota Policy Transformed Immigration Judging," *American Political Science Review*, 2023, pp. 1–16.

Boyd, Christina L., "Representation on the Courts? The Effects of Trial Judges' Sex and Race," *Political Research Quarterly*, 2016, *69* (4), 788–799.

_ , Lee Epstein, and Andrew D. Martin, "Untangling the Causal Effects of Sex on Judging," *American Journal of Political Science*, 2010, *54* (2), 389–411.

Chand, Daniel E., William D. Schreckhise, and Marianne L. Bowers, "The Dynamics of State and Local Contexts and Immigration Asylum Hearing Decisions," *Journal of Public Administration Research and Theory: J-PART*, 2017, *27* (1), 182–196.

Cook, Beverly B, "Will women judges make a difference in women's legal rights? A prediction from attitudes and simulated behaviour," *Women, power, and political systems*, 1981, pp. 216–39.

EOIR, "FY 2016 Statistics Yearbook," 2016.

_ , "Immigration Judge Performance Metrics," Memorandum March 2018.

Farhang, Sean and Gregory Wawro, "Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making," *The Journal of Law, Economics, and Organization*, 10 2004, *20* (2), 299–330.

GAO, "U.S. ASYLUM SYSTEM Significant Variation Existed in Asylum Outcomes across Immigration Courts and Judges," *Report to Congressional Requesters GAO-08-940, September*, 2008.

_ , "ASYLUM Variation Exists in Outcomes of Applications Across Immigration Courts and Judges," *Report to Congressional Requesters GAO-17-72, November*, 2016.

_ , "U.S. Government Accountability Office . Immigration Courts: Actions Needed to Reduce Case Backlog and Address Long-Standing Management and Operational Challenges," *Report to Congressional Requesters GAO-17-438, June*, 2017.

Gilligan, Carol, "In a Different Voice: Psychological Theory and Women's Development.," *Cambridge: Harvard University Press.*, 1982.

Hennessy-Fiske, Molly, "Immigration judges are quitting or retiring early because of Trump," *Los Angeles Times, January 27*, 2020, *Available at: https://www.latimes.com/world-nation/story/2020-01-27/immigration-judges-are-quitting-or-retiring-early-because-of-trump.*

Kastellec, Jonathan P., "Racial Diversity and Judicial Influence on Appellate Courts," *American Journal of Political Science*, 2013, *57* (1), 167–183.

Keith, Linda Camp, Jennifer S. Holmes, and Banks P. Miller, "Explaining the Divergence in Asylum Grant Rates among Immigration Judges: An Attitudinal and Cognitive Approach," *Law & Policy*, 2013, *35* (4), 261–289.

Long, Colleen, "Immigration Judges Say New Quotas Undermine Independence," *Associated Press, September 21.*, 2018, *https://apnews.com/article/d8008f7a66a54562b612bd74156f2bed.*

McClain, Paula D., Jessica D. Johnson Carew, Eugene Walton, and Candis S. Watts, "Group Membership, Group Identity, and Group Consciousness: Measures of Racial Identity in American Politics?," *Annual Review of Political Science*, 2009, *12* (Volume 12, 2009), 471–485.

Menkel-Meadow, Carrie, "Asylum in a Different Voice: Judging Immigration Claims and Gender," *Georgetown Law Faculty Publications and Other Works. 364.*, 2009.

National Association of Immigration Judges, "By the Numbers: Why Production Quotas for Immigration Judges Won't Work."

Nielson, Victoria, "DOJ Requires Immigration Judges To Meet Quotas," *Catholic Legal Immigration Network, April 27. Available at: https://cliniclegal.org/resources/doj-requires-immigration-judges-meet-quotas*, 2018.

Ramji-Nogales, Jaya, Andrew I. Schoenholtz, and Philip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication," *Stanford Law Review*, 2007, *60* (2), 295–411.

Rose, Joel, "Justice Department Rolls Out Quotas For Immigration Judges," *NPR, April 3. Available at: https://www.npr.org/2018/04/03/599158232/justice-department-rolls-out-quotas-for-immigration-judges*, 2018.

Rosenman, Evan, Santiago Olivella, and Kosuke Imai, "Race and ethnicity data for first, middle, and last names," *Harvard Dataverse, V9*, 2022.

Sacchetti, Maria, "Immigration Judges Say Proposed Quotas from Justice Dept. Threaten Independence," *The Washington Post, October 12.*, 2017, *https://www.washingtonpost.com/local/immigration/immigration-judges-say-proposed-quotas-from-justice-dept-threaten-independence/2017/10/12/3ed86992-aee1-11e7-be94-fabb0f1e9ffb*$_{s}$*tory.html?utm*$_{t}$*erm = .ccfbf*1*bd*5*d*7*f.*

Sherry, Suzanna, "Civic Virtue and the Feminine Voice in Constitutional Adjudication," *Virginia Law Review*, 1986, *72* (3), 543–616.

UCI, "Machine Learning Repository," *https://doi.org/10.24432/C55G7X.*, 2020.

U.S. Commission on Civil Rights, "Statement on Department of Justice Immigration Case Quotas."

Appendix

Data Appendix

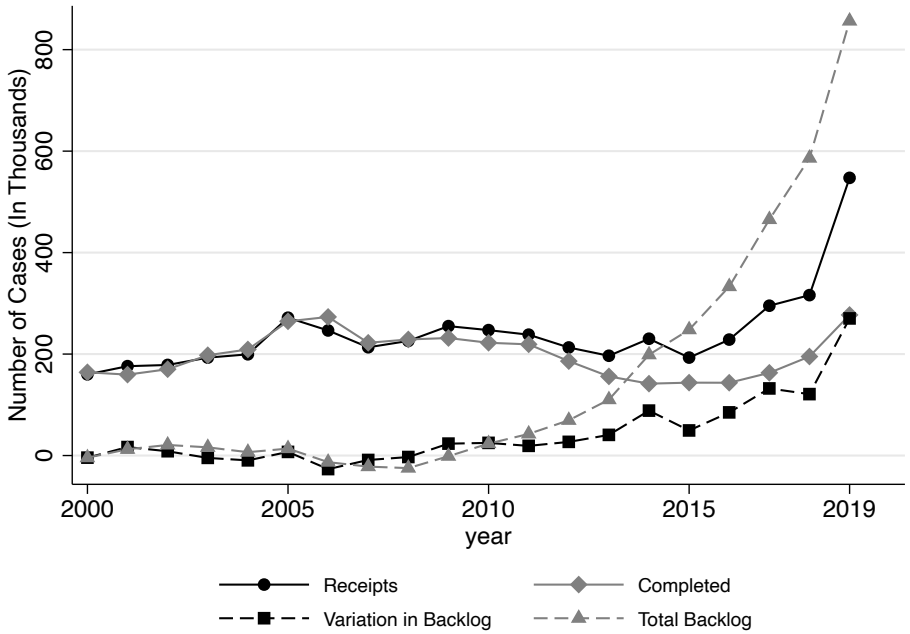
The Executive Office for Immigration Review (EOIR) dataset provides comprehensive administrative records on all cases processed in U.S. immigration courts. It is organized into multiple components, each capturing distinct dimensions of court operations and judicial performance. As with most large administrative datasets, transcription errors and missing values may occur; standard cleaning and validation procedures are applied to ensure analytic reliability. For this analysis, we rely on three core datasets.

The *first* contains case-level information, including the assigned court, final decision date, and case type. It supports analyses of processing trends such as case completion volumes and the share of decisions favoring respondents. The *second* provides hearing-level data, detailing the presiding judge, hearing date, and court location. It also includes judges' caseloads—critical for evaluating how performance quotas affected judicial productivity. The *third* dataset contains judge-level information, including names, unique identification codes, court assignments, and professional backgrounds such as years of experience. Names are used to infer demographic characteristics like gender and race, which inform analyses of judicial departures and court composition over time.

All three datasets are linked via case and judge identifiers, enabling a unified analysis of case outcomes alongside judge- and court-level attributes. This structure facilitates our examination of departures, changes in judicial experience, and shifts in adjudication patterns.

Appendix Figures and Tables

Figure A1: Case Backlog Trends, 2000–2019



Notes: Annual flows of new case receipts and case completions (solid lines) and backlog measures, FY2000–FY2019. The dashed black line is the net change in backlog in year t (receipts$_t$ – completions$_t$). The dashed gray line is the cumulative backlog (end-of-year pending cases). Counts are in thousands. Data: U.S. Department of Justice, Executive Office for Immigration Review, "Workload and Adjudication Statistics" (https://www.justice.gov/eoir/workload-and-adjudication-statistics).

Table A1: Actual Treatment vs Predicted Treatment Comparison

Column	(1)	(2)	(3)	(4)
Statistic	Mean	Std. Dev.	Min	Max
Panel A: Immigration Judge Level Dataset				
Pred. Share Below Threshold	0.3170	0.4653	0	1
Share Below Threshold	0.3434	0.4748	0	1
Panel B: Immigration Court Level Dataset				
Pred. Share of Judges Below Threshold	0.1895	0.2950	0	1
Share of Judges Below Threshold	0.2243	0.3080	0	1

Statistic	Mean Pre	Mean Post	(2) – (1)	P-Value
Panel C: Immigration Judge Level Dataset				
Pred. Share Below Threshold	0.2627	0.5051	0.2423***	0.000
	(0.4727)	(0.5002)		
Share of Judges Below Threshold	0.3370	0.3655	0.0284*	0.072
	(0.4727)	(0.4817)		
Panel D: Immigration Court Level Dataset				
Pred. Share Below Threshold	0.1312	0.4809	0.3496***	0.000
	(0.3023)	(0.3151)		
Share of Judges Below Threshold	0.2029	0.3312	0.1283***	0.000
	(0.3023)	(0.3151)		

Notes: Panels A–B report 2000–2019 summary statistics. "Pred. Share Below Threshold" is constructed solely from *pre-policy* completion histories (2000–2016) and is therefore predetermined with respect to the quota: at the judge–year level (Panel A) it is an indicator equal to 1 if a judge is predicted to complete < 700 cases; at the court–year level (Panel B) it is the share of judges in the court predicted to fall below 700. "Share of Judges Below Threshold" denotes the *actual/realized* share based on contemporaneous completions < 700 (assigned to each judge–year in Panel A; aggregated in Panel B). Panels C–D compare pre-policy (2000–2017) and post-policy (2018–2019) means; column (3) reports post–pre differences and column (4) two-sided p-values. Values in parentheses are standard deviations. Asterisks denote significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A2: Quota Policy Impact on Judge Departures Using a Continuous Treatment

Specification/Column	(1)	(2)	(3)
Year	2017	2018	2019
Panel A: Effect on All Judge Departures			
Continuous Treatment*Post	0.0445***	0.0005	-0.0274***
	(0.0102)	(0.0127)	(0.0127)
DV Mean		0.0402	
Observations		4628	
R-squared		0.0845	
Panel B: Effects on the Departure of Female Judges			
Continuous Treatment*Post	0.1278***	0.2172***	0.0025
	(0.0383)	(0.0434)	(0.0510)
DV Mean		0.0250	
Observations		1537	
R-squared		0.1800	
Panel C: Effects on the Departure of Male Judges			
Continuous Treatment*Post	0.0595	-0.03967	-0.1757***
	(0.0435)	(0.0453)	(0.0421)
DV Mean		0.0486	
Observations		2887	
R-squared		0.1702	
Panel D: Effects on the Departure of White Judges			
Continuous Treatment*Post	0.1366***	0.0873**	-0.1581***
	(0.0352)	(0.0436)	(0.0372)
DV Mean		0.0305	
Observations		3660	
R-squared		0.1773	
Panel E: Effects on the Departure of Non-White Judges			
Continuous Treatment*Post	0.1245***	0.1468**	0.2413***
	(0.0312)	(0.0444)	(0.0701)
DV Mean		0.0427	
Observations		968	
R-squared		0.1666	

Notes: Unit of analysis is a *judge–year*. The dependent variable equals 1 if judge j separates in fiscal year t. Displayed coefficients and standard errors result from the estimation of equation (3). Columns (1)–(3) report coefficients from interacting a *continuous* treatment—the judge-level standardized caseload shortfall relative to the 700-case threshold, constructed using only pre-policy completions (2000–2016; one unit = one s.d.; larger values = larger shortfall below 700)—with indicators for FY2017, FY2018, and FY2019 (relative to FY2016). Panel B restricts to female judges; Panel C to male judges; Panel D to White judges; Panel E to Non-White judges. "DV" is the unconditional mean separation rate in the corresponding subsample. Coefficients are percentage-point effects for a one-s.d. increase in the shortfall. Standard errors are two-way clustered by judge and year. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A3: Quota Policy Impact on Court-Level Outcomes

Specification/Column	(1)	(2)	(3)
Year	2017	2018	2019
Panel A: Share of Departures			
Share*Post	0.0481***	0.0642**	0.1284***
	(0.0076)	(0.0211)	(0.0085)
DV Mean		0.0385	
Observations		1057	
R-squared		0.1167	
Panel B: Judicial Experience			
Share*Post	-2.4879***	-1.8816***	-1.6922***
	(0.7225)	(0.6165)	(0.8641)
DV Mean		10.0640	
Observations		1057	
R-squared		0.4879	
Panel C: Case Completion Volume			
Share*Post	205.7100	888.3483	1136.1380
	(423.9634)	(540.1293)	(663.6380)
DV Mean		4880.2090	
Observations		1057	
R-squared		0.8050	
Panel D: Case Completion Volume (Dropping Judges Appointed After 2016)			
Share*Post	-579.9955	-391.704	-123.2191
	(654.0756)	(786.286)	(935.0906)
DV Mean		4880.2090	
Observations		1057	
R-squared		0.8052	
Panel E: Share of (Granted / Actual Cases)			
Share*Post	-0.0464**	-0.0489	-0.0531
	(0.0186)	(0.0299)	(0.0326)
DV Mean		0.2156	
Observations		1057	
R-squared		0.8115	
Panel F: Share of (Granted / Predicted Cases)			
Share*Post	-0.0775***	-0.0845***	-0.0885***
	(0.0096)	(0.0260)	(0.0279)
DV Mean		0.1556	
Observations		1057	
R-squared		0.5546	

Notes: Unit of analysis is a *court–year*. Displayed coefficients and standard errors result from the estimation of equation (4). "Share×Post(y)" interacts the court's exposure— the share of judges predicted (using only pre-policy completions, 2000–2016) to complete < 700 cases—with indicators for FY2017, FY2018, and FY2019 (reported in columns (1)–(3)). Row "DV" reports the pre-policy mean of the dependent variable for each panel. Standard errors are two-way clustered by court and year. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.