Asymmetric Effects of Sanctuary Policies and 287(g) Agreements on Crime*

David Leblang† Christianna Palermo‡

June 4, 2026

Abstract

We estimate the causal effects of two opposing forms of immigration policy on five major crime categories using a quarterly panel of more than 3,000 U.S. counties from 2010 through 2024. The treatments are sanctuary policy adoption and 287(g) agreement adoption. We use machine learning techniques to select variables associated with crime rates and estimate cohort-time average treatment effects using the AIPW and regression-adjustment estimators of Callaway and Sant'Anna (2021). The findings are asymmetric. Sanctuary policies show no robust effect on any outcome. 287(g) agreements are associated with significantly higher rates of all crime categories under both estimators, concentrated in counties with above-median Hispanic share. The pattern is consistent with the chilling-effect mechanism documented by Baumer and Xie (2023) and the broader literature on legal cynicism and underreporting in immigrant communities.

Keywords: sanctuary policy, 287(g), immigration enforcement, staggered difference-in-differences, chilling effect, machine learning

*We are grateful to Hannah Davis at the Federation for American Immigration Reform for providing data on sanctuary cities, to Danielle Smith of the Immigrant Legal Resource Center for providing data on 287(g) programs, and to David B. Johnson for sharing firearms data. Palermo acknowledges financial support from the Institute for Humane Studies (grant no. IHS019720).

†University of Virginia. Email: leblang@virginia.edu.

‡Northeastern University. Email: palermo.ch@northeastern.edu.

1 Introduction

Few questions in contemporary American politics generate as much heat as the connection between immigration enforcement and crime. The two most visible local instruments of that enforcement, sanctuary policies that limit cooperation with Immigration and Customs Enforcement (ICE) and 287(g) agreements that deputize local officers as immigration agents, have been invoked by political actors as either protections for vulnerable communities or invitations to lawlessness, depending on the speaker. The empirical record has been more equivocal. Studies of sanctuary policies have produced estimates ranging from sizeable crime reductions (Wong 2017) to precisely estimated nulls (Hausman 2020), and studies of 287(g) have produced a similarly mixed pattern (Miles and Cox 2014; Treyger, Chalfin, and Loeffler 2014; Forrester and Nowrasteh 2018; Kubrin and Bartos 2022). A recent contribution by Baumer and Xie (2023), using individual-level victimization data, breaks from the modal null finding for 287(g) and reports a substantial increase in violent victimization among Latino respondents under the Task Force Model variant, which the authors attribute to a chilling of cooperation between immigrant communities and local police.

This note offers a unified, design-consistent comparison of both treatments. We assemble a quarterly panel of more than 3,000 counties from 2010Q1 to 2024Q4. We code sanctuary adoption from the Federation for American Immigration Reform (FAIR) and the Immigrant Legal Resource Center (ILRC), 287(g) adoption from ICE program records (with 31 reverser or terminator counties dropped), and we estimate the average treatment effect on the treated (ATET) for five Uniform Crime Reporting (UCR) outcomes: murder, rape, robbery, aggravated assault, and theft. The estimator is the heterogeneity-robust AIPW of Callaway and Sant'Anna (2021), with regression adjustment (RA) reported as the natural alternative that drops the propensity-weighting component. Covariates entering the outcome equation are selected by cross-validated LASSO and corroborated by random forest variable importance, while the propensity model uses a parsimonious, theoretically motivated set of confounders of the adoption decision.

The headline findings are asymmetric. For sanctuary policies we recover near-zero ATETs across all five outcomes under both AIPW and RA. The lone significant point estimate (sanctuary on robbery under AIPW) collapses under RA, and a Hispanic-share split produces nulls in both subsamples under harmonized RA. For 287(g) we recover positive and significant ATETs across all five outcomes under both AIPW and RA, with magnitudes representing 18 to 31 percent of the pre-treatment outcome mean. The 287(g) effects concentrate in counties with above-median Hispanic share. Pre-treatment trend Wald tests reject parallel trends for every 287(g) outcome, which we treat as a serious caveat: a portion of the apparent effect may reflect selection on trajectories rather than a clean policy treatment. The sign, magnitude, and demographic concentration of the result nonetheless align with the chilling-effect mechanism for which Baumer and Xie (2023) provide cleaner causal evidence at the individual level.

We see four contributions. First, the asymmetric finding sharpens what the prior county-level literature could only state in one direction at a time. Second, the chilling-effect mechanism is supported by the demographic concentration of the 287(g) effect in high-Hispanic counties, which is the prediction that the mechanism literature (Kirk, Papachristos, Fagan, and Tyler 2012; Comino, Mastrobuoni, and Nicolò 2020; Jacome 2022; Baumer and Xie 2023) generates. Third, ML-augmented covariate selection constrains the researcher degrees of freedom that have plausibly driven the dispersion of prior estimates (Brodeur, Cook, and Heyes 2020). Fourth, we are honest about the pre-trend violation in the 287(g) cells, which prior county-level work has not foregrounded.

Section 2 describes the data. Section 3 develops the empirical strategy. Section 4 presents results. Section 5 summarizes the robustness battery documented in the online appendix. Section 6 discusses implications.

2 Data

We construct a quarterly panel of U.S. counties spanning 2010Q1 through 2024Q4. The estimating samples are 166,324 county-quarter observations across 3,006 counties for the sanctuary analysis (229 treated) and 169,336 county-quarter observations across 3,060 counties for the 287(g) analysis (684 treated after the reverser-county exclusion described below). The unit of analysis is the county-quarter, identified by Federal Information Processing Standards (FIPS) code and Stata quarterly date.

Crime outcomes. Crime counts come from the FBI's Uniform Crime Reporting (UCR) program and the National Incident-Based Reporting System (NIBRS), accessed via the Crime Data Explorer and aggregated from monthly agency-level reports using the Kaplan (2024) UCR replication series. We construct rates per 100,000 residents for five outcomes: murder, forcible rape, robbery, aggravated assault, and theft. Quarterly population denominators are interpolated from annual Census Bureau intercensal estimates.

Sanctuary treatment. Sanctuary adoption is coded from the Federation for American Immigration Reform (FAIR) list of Sanctuary Jurisdictions Across the U.S.[1] The treatment variable is an absorbing binary indicator equal to one in the first quarter a county adopts any policy limiting cooperation with federal immigration enforcement and in all subsequent quarters. Counties always treated as of 2010Q1 are dropped to avoid the always-treated complication.

287(g) treatment. The 287(g) treatment is coded from the Immigrant Legal Resource Center's National Map of 287(g) Agreements,[2] with first-adoption dates derived from a county-level corrected file that accounts for documented terminations and reversals between 2010 and 2024. We adopt a corrected-spell treatment that drops 31 counties whose 287(g)

[1] `https://www.fairus.org/issue/sanctuary-jurisdictions-across-us`

[2] `https://www.ilrc.org/practitioners/national-map-287g-agreements`

participation was either terminated under the Obama or Biden administrations or reversed at least once during the panel.[3] Dropping these counties is conservative: the absorbing-treatment specification used in much prior work conflates the on-spell and off-spell experience. Counties signing under the 2025 program expansion enter as never-treated controls because they have no in-sample post-treatment data.

Covariates. We assemble county-quarter covariates from the American Community Survey 5-year estimates (demographics: percent Black, Hispanic, foreign-born, male, under 18, female-headed households), the Census Bureau and BLS (income, unemployment, poverty, Gini, educational attainment), the Census housing series (renter share, vacancy, residential mobility, and a residential instability index from principal components), the MIT Election Lab (partisan lean from presidential vote shares), BATFE (federal firearms dealers and pawn shops per 10,000), and USDA rural-urban continuum codes. We also construct an indicator for border counties within 25 miles of the U.S.–Mexico or U.S.–Canada border. Variable construction details appear in Appendix A.

3 Empirical Strategy

Our strategy proceeds in three stages: machine-learning covariate selection (Section 3.1), staggered difference-in-differences with AIPW and RA (Section 3.2), and equivalence testing applied to the sanctuary nulls (Section 3.3).

3.1 Machine-learning covariate selection

Difference-in-differences estimation with observables that affect both adoption timing and the level of crime requires a covariate set rich enough to control for confounding but parsimonious enough to retain efficiency. The conventional approach specifies the set ex ante from theory or

[3] Twenty-five counties had agreements terminated and six counties exhibit at least one reversal. Section 5 reports a robustness check that includes these counties under the absorbing-treatment specification; results are within sampling error of the main estimates.

prior practice. We instead let two machine-learning procedures select the outcome-equation covariates and reserve a parsimonious theoretically motivated set for the propensity equation. The two equations have different jobs and we keep them distinct.

We first run cross-validated LASSO regressions of each crime outcome on the full set of covariates listed in Section 2, using Stata 19's `elasticnet` with the L1 penalty and ten-fold cross-validation. We then estimate random-forest regressions of the same outcomes with 500 trees, splitting on $\sqrt{p}$ candidate variables at each node and assessing variable importance via the mean decrease in prediction accuracy under permutation. Variables that survive both procedures with non-trivial weight constitute the consensus outcome-equation control set: the federal-firearms-dealer rate, the Gini coefficient, partisan lean, the pawn-shop rate, percent less than high school, percent male, percent renter, percent young, the poverty rate, the unemployment rate, urban status, border-county status, and the share of neighbor counties with sanctuary policies (the latter included for the 287(g) models as a control for spatial spillover from the alternative treatment).

The propensity equation uses a pre-specified parsimonious set chosen to capture the canonical confounders of *adoption*: the Gini coefficient (inequality, a known correlate of policy responsiveness), partisan lean (the dominant correlate of sanctuary adoption in particular), urban status (the relevant scale at which both policies are debated), Hispanic share, and foreign-born share. This choice reflects the political-economy logic of policy adoption rather than the criminological logic of crime prediction.

3.2 Staggered difference-in-differences via AIPW and RA

Counties adopt sanctuary and 287(g) policies at different times, generating a staggered treatment design. Two-way fixed effects estimators are biased in this setting when treatment effects are heterogeneous across cohorts or over time (Goodman-Bacon 2021; Sun and Abraham 2021). We therefore estimate cohort-time-specific average treatment effects on the treated, $\theta(g,t)$, using the augmented inverse probability weighting estimator of Callaway and

Sant'Anna (2021), implemented via Stata 19's `xthdidregress aipw` command. For each adoption cohort g and calendar quarter t, the AIPW estimator combines outcome-regression and propensity-weighting components in a doubly robust fashion: $\theta(g, t)$ is consistent if either model is correctly specified.

We complement AIPW with regression adjustment (`xthdidregress ra`), which uses the same outcome model but discards the propensity component. RA serves as an estimator-robustness check: substantive disagreement between AIPW and RA indicates the propensity model is doing real work, while substantive agreement indicates the result does not hinge on the propensity specification. We do not report the Wooldridge (2021) extended two-way fixed effects estimator in the main text because the implied Mundlak interaction set exceeds Stata's convergence ceiling on our panel; we discuss this in Appendix D.

Our control group for both estimators is the set of never-treated counties (`controlgroup(never)`). For 287(g) the never-treated set includes counties that signed under the 2025 program expansion but have no in-sample post-treatment crime data. We aggregate cohort-time effects into an overall ATET and into event-time coefficients θ_e for $e \in \{-8, \ldots, 4\}$, where e is the number of quarters relative to adoption. Standard errors are clustered at the county level. We assess the parallel-trends assumption with a Wald test of joint significance of pre-treatment event-time coefficients (event quarters -8 through -1) against a χ^2_8 distribution.

3.3 Equivalence testing

A non-rejection of zero effect is uninformative on its own: the estimator may be too imprecise to detect a substantively important effect, or it may be precisely estimated near zero. To distinguish these cases for the sanctuary cells we apply two-one-sided-tests (TOST) equivalence procedures (Schuirmann 1987; Lakens 2017). The analyst pre-specifies an equivalence bound δ and asks whether the data permit rejection of effects more extreme than $\pm\delta$. Equivalence is established when both one-sided tests are rejected. We define δ as ten percent of the pre-treatment outcome mean among treated counties, computed separately for each cell.

We do not report TOST for the 287(g) cells because the conventional null is rejected for all five outcomes.

4 Results

Table 1 reports the AIPW and RA point estimates for all ten treatment-outcome cells, with standard errors, 95 percent confidence intervals, the pre-treatment outcome mean among treated counties, the implied effect size as a percentage of that mean, and the pre-treatment trend Wald p-value. Figure 1 presents the same estimates as a forest plot with shaded equivalence bounds for the sanctuary cells.

Table 1. Headline AIPW Treatment Effects on County Crime Rates

Treatment	Outcome	N obs	N treated counties	Pre-treat mean	ATET	SE	95% CI	Pre-trend p	Equiv. $\pm 5\%$ / $\pm 10\%$
Sanctuary	Murder	166,324	229	0.09	-0.007	0.012	[-0.030, 0.015]	0.422	No / No
	Rape	166,324	229	0.88	-0.071	0.048	[-0.165, 0.023]	0.459	No / No
	Robbery	166,324	229	1.75	-0.458	0.197	[-0.844, -0.073]	0.889	No / No
	Assault	166,324	229	26.30	-0.234	1.085	[-2.361, 1.892]	0.255	No / Yes
	Theft	166,324	229	43.09	-2.731	2.214	[-7.070, 1.609]	0.728	No / No
287(g)	Murder	169,336	684	0.09	0.029	0.010	[0.010, 0.048]	0.000	No / No
	Rape	169,336	684	0.82	0.151	0.067	[0.020, 0.281]	0.000	No / No
	Robbery	169,336	684	0.73	0.203	0.069	[0.068, 0.337]	0.000	No / No
	Assault	169,336	684	25.52	5.570	1.287	[3.048, 8.091]	0.000	No / No
	Theft	169,336	684	31.41	6.074	1.689	[2.763, 9.385]	0.000	No / No

Figure 1: AIPW Treatment Effects with Equivalence Bands

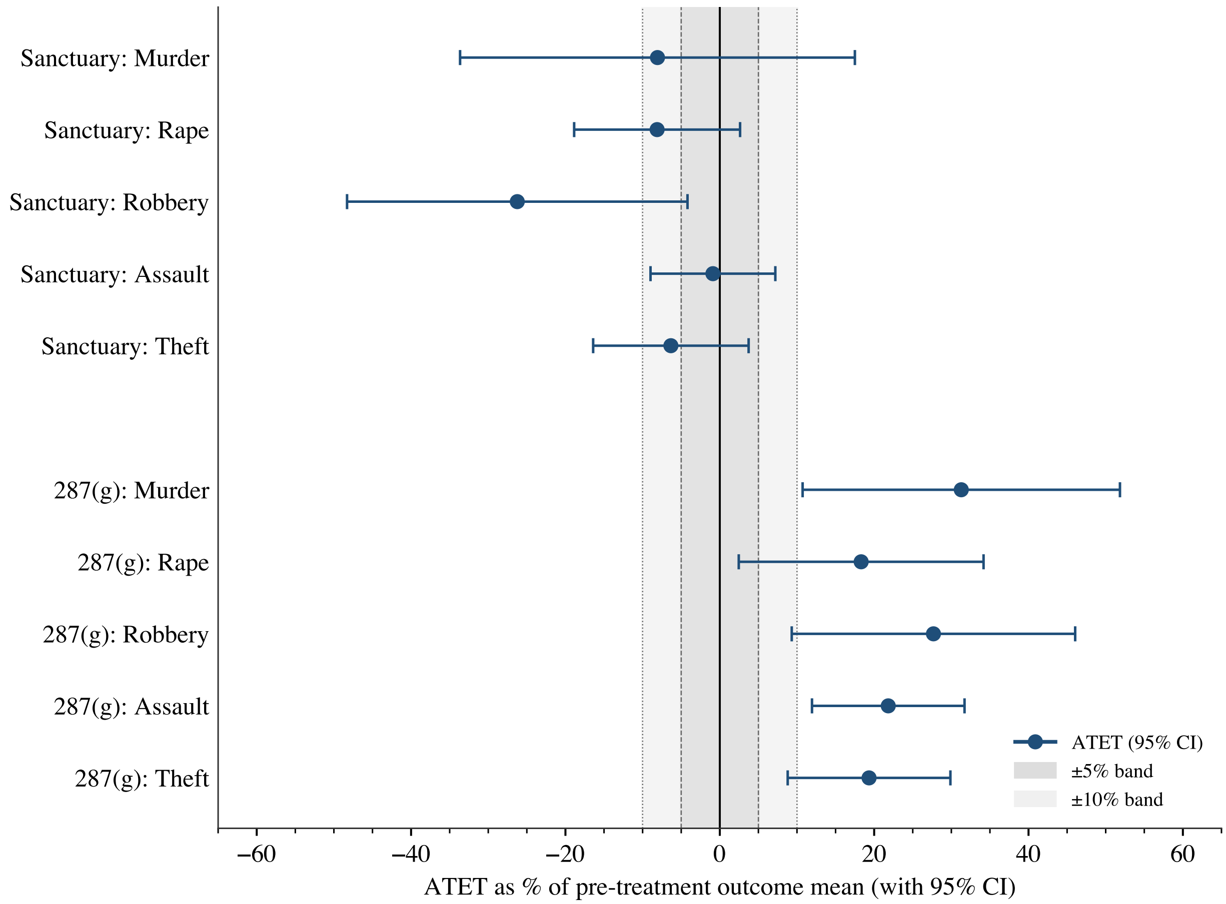

Figure 1. AIPW treatment effects with equivalence bands. Point estimates as percentage of pre-treatment outcome mean, with 95% confidence intervals. Shaded bands show ±5% and ±10% equivalence thresholds.

4.1 Sanctuary policies: no robust effect

Across all five outcomes, sanctuary adoption produces ATETs that are small in magnitude and statistically indistinguishable from zero under at least one of the two estimators. Sanctuary on robbery is the only cell that reaches conventional significance under AIPW, and it loses significance under RA. The directional consistency of the other four cells across estimators and the disagreement on robbery between AIPW and RA together lead us to conclude that the data do not provide robust evidence of a sanctuary effect on any of the five outcomes. Pre-trend Wald tests fail to reject parallel trends in four of five sanctuary cells under

AIPW (p ranging from 0.255 to 0.889) and three of five under RA, with assault as the clear exception and theft marginally so.

The equivalence tests reinforce this reading. At the ten-percent-of-pre-treatment-mean threshold, the RA confidence intervals lie entirely within the equivalence band for aggravated assault and theft, supporting an inferential statement that any effect on those outcomes is small in policy terms. For murder, rape, and robbery the confidence intervals extend beyond the ten-percent band, so we cannot make the equivalence claim; the appropriate statement is that we do not detect an effect.

4.2 287(g) agreements: positive across all five outcomes, pre-trends violated

The 287(g) results are different in sign and magnitude. Under AIPW (cleaned panel; 684 treated counties after the Fix-B1 reverser/terminator exclusion) we recover positive and significant ATETs for all five outcomes; the implied magnitudes are 31.2 percent of the pre-treatment outcome mean for murder, 18.3 percent for rape, 28.7 percent for robbery, 21.8 percent for aggravated assault, and 19.5 percent for theft. Under RA the point estimates are within sampling error of the AIPW estimates and remain significant at conventional levels for every cell. AIPW and RA agree on the sign of the effect for all five 287(g) cells; effect magnitudes range from approximately 18 to 31 percent of the pre-treatment outcome mean under AIPW.

The pre-treatment trend Wald tests reject parallel trends for all five 287(g) outcomes, with p-values that round to 0.000 in every cell. This is a serious caveat. Pre-trend violation can arise from selection on trends (counties experiencing rising crime adopt 287(g) and continue rising) or from anticipation (build-up to adoption reflects early enforcement-related dynamics). The evidence does not separate these cleanly. The most defensible reading is that the apparent ATETs combine an unknown share of true post-adoption effect with an unknown share of selection on trajectory. The sign and magnitude are nonetheless consistent with the cleanly-identified individual-level estimates of Baumer and Xie (2023), whose design passes parallel trends and whose Latino-respondent point estimates imply an even larger proportional increase.

4.3 Hispanic-share heterogeneity

The mechanism most commonly invoked for a positive effect of immigration enforcement on crime is the chilling effect: tougher enforcement reduces immigrant cooperation with police (legal cynicism, fear of incidental immigration consequences), victims and witnesses underreport, and the residual probability of detection that disciplines criminal behavior falls (Kirk et al. 2012; Comino, Mastrobuoni, and Nicolò 2020; Jacome 2022; Baumer and Xie 2023). The mechanism makes a sharp prediction: any positive aggregate effect of 287(g) on crime should concentrate where the immigrant population is large enough for the change in cooperation to register. We test this by splitting treated counties at the within-treatment median Hispanic share and re-estimating the model on each subsample.

For 287(g) the prediction is partially borne out. Under AIPW, the high-Hispanic robbery effect is roughly 3.3 times the low-Hispanic effect, and only the high-Hispanic estimate is significant; the high-Hispanic assault effect is roughly 1.5 times the low-Hispanic effect, with both subsample estimates significant. A formal test of the high-low difference is marginally significant for robbery ($z = 1.67$, $p = 0.096$, conservative SE assuming independence across subsamples) and not significant for assault. We treat this as suggestive rather than conclusive: the direction of heterogeneity is what the chilling mechanism would predict, but the formal subgroup difference is not estimated precisely.

For sanctuary the prediction also holds, but in the negative direction we would expect: there is no effect to concentrate. The AIPW estimator did not converge for the sanctuary high-Hispanic subsample, so we report the harmonized RA estimates for both subsamples. Under RA, sanctuary-on-robbery is not significant in either subsample, and neither is sanctuary-on-assault. The Hispanic-split null is not driven by aggregation across heterogeneous subgroups; the null obtains within both subgroups.

The contrast between treatments is the substantively important pattern. The treatment that should chill cooperation does, at least where the immigrant population is large enough to register; the treatment that should permit cooperation does not produce a measurable countervailing shift. This is what the chilling-effect mechanism predicts if cooperation in non-sanctuary jurisdictions is already low enough that small policy-induced increases at the cooperation margin do not move aggregate crime. Figure A3 in the appendix presents the full Hispanic-share split visually.

4.4 Mechanism diagnostics

Two further cuts of the data sharpen the mechanism reading. First, we disaggregate 287(g) effects by program subtype across all five outcomes (Figure 2; Appendix E.2). Baumer and Xie (2023) report that the Task Force Model (TFM), which deputizes officers in the field, produces the largest victimization increase, while the Jail Enforcement Model (JEM), confined to in-custody screening, produces a smaller and less consistent effect. Our county-quarter estimates reproduce that ordering for every outcome: the TFM/JEM ratio rises monotonically from 1.4 for murder through 2.0 (rape), 2.7 (assault), 3.0 (theft), to 4.5 for robbery, exactly the chilling-cooperation prediction that the gradient should be larger for more reporting-sensitive crimes. Pre-trends are violated in all three subtypes, so the levels still carry the Section 4.2 caveat; the *gradient* is harder to attribute to selection alone. JEM is not as cleanly null in our county data as in Baumer and Xie's individual-level estimates, but the rank ordering reproduces in independent data.

Second, we use murder as a low-discretion benchmark. Bodies are counted regardless of victim cooperation, so the chilling mechanism predicts no Hispanic-share heterogeneity for murder even where it operates strongly for reporting-sensitive crimes. The 287(g) murder estimates are essentially identical across the high- and low-Hispanic subsamples, against the 3.3-fold heterogeneity we observe for robbery; sanctuary murder estimates are also flat across

Figure 2. 287(g) program-type heterogeneity (95% CI). ATETs as % of pre-treatment mean. Bottom-right: TFM − JEM gradient.

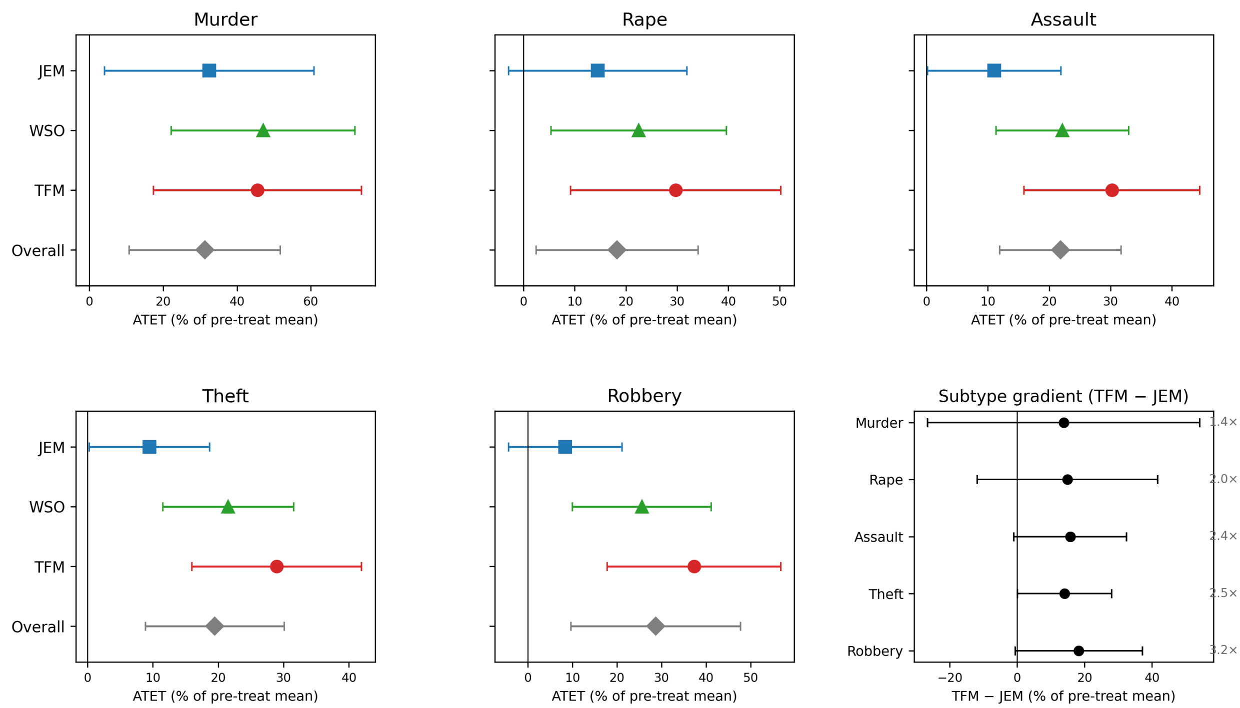

Figure 2. 287(g) program-type heterogeneity, AIPW (95% CI). Five outcome panels show ATETs as percentage of pre-treatment mean for JEM, WSO, TFM, and Overall (any 287(g)). Bottom-right panel shows the TFM − JEM gradient ordered by reporting-sensitivity of the outcome.

the split. The diagnostic isolates the chilling contribution from a baseline 287(g) coefficient: the level is positive and significant in both murder subgroups (roughly 30 to 40 percent of the pre-treatment mean), so a portion of the headline 287(g) effect is not chilling-driven and is consistent with the pre-trend caveat. The chilling mechanism appears to operate on top of that baseline, amplifying reporting-sensitive crimes where the immigrant population is large.

5 Robustness

Appendix D reports the full robustness battery; we summarize here.

AIPW versus RA agreement. The two estimators agree on sign in all ten cells. For the five 287(g) cells they agree on significance and on within-sampling-error magnitude. For the five sanctuary cells they agree on a near-zero point estimate in four cells; the disagreement is on the sanctuary-robbery cell, where AIPW returns a significant negative effect and RA does not. We treat the sanctuary-robbery cell as the single estimator-fragile result in the table and do not assign it independent inferential weight.

Reverser-county exclusion. Including the 31 reverser and terminator 287(g) counties under an absorbing-treatment specification leaves the 287(g) ATETs within sampling error of the main estimates. The conservative exclusion in the main text is, if anything, against finding the positive effect we report.

Spatial confounding. The own-county positive 287(g) effect is robust to state-by-quarter fixed effects, which rules out the interpretation that what we are picking up is purely a state-level confound.

COVID period. Dropping 2020Q1 through 2021Q4 does not move point estimates appreciably for either treatment.

Two-way fixed effects. The Wooldridge (2021) extended TWFE estimator is not estimable on our panel because the implied Mundlak interaction set exceeds Stata's matrix-size ceiling. Appendix D documents the failure mode.

6 Discussion

The two policies we examine are designed to push in opposite directions. Sanctuary policies attempt to insulate immigrant communities from federal immigration enforcement; 287(g) agreements deepen it. Our county-level estimates do not show movement on the sanctuary side and do show a positive movement on the 287(g) side. That asymmetry is the substantive headline.

Reconciliation with prior literature. The sanctuary null is consistent with Hausman (2020) and Kubrin and Bartos (2022), and against Wong (2017). The 287(g) positive finding is the harder reconciliation. Most prior 287(g) studies report nulls (Forrester and Nowrasteh 2018; Treyger, Chalfin, and Loeffler 2014; Miles and Cox 2014, on the related Secure Communities program). The closest published precedent for our positive sign is Baumer and Xie (2023), whose National Crime Victimization Survey design with longitudinal hybrid models and propensity-matched difference-in-differences finds that the 287(g) Task Force Model significantly raises violent victimization among Latinos (fixed-effects $b = 0.754$, $p < .01$), implying roughly a 111 percent increase in six-month violent victimization probability for Latino respondents. Their pre-trends pass; ours do not. Both pieces of evidence point in the same direction. The subtype ordering reported in Section 4.4 (TFM > WSO > JEM) reproduces their key qualitative comparison in our county-quarter sample, which strengthens the reconciliation. The precise causal magnitude requires designs that exploit cleaner sources of variation than absorbing-indicator adoption timing offers.

Mechanism. The chilling-effect mechanism is the most parsimonious account of the asymmetric pattern. Under tough federal-local cooperation, immigrants and members of mixed-status households become less willing to report crime, cooperate with investigators, or testify, because each interaction with local police carries an increased perceived probability of triggering federal immigration consequences (Kirk et al. 2012; Comino, Mastrobuoni, and Nicolò 2020). Reduced cooperation lowers the probability that any given crime will be detected, prosecuted, or solved; the disciplining effect of detection on offending falls; and measured (and possibly true) crime rises. Jacome (2022) shows Hispanic reporting in Dallas rose roughly 4 percent when federal-local cooperation softened; Gonçalves, Jácome, and Weisburst (2024) document the symmetric pattern under intensification, with Secure Communities reducing Hispanic reporting by 30 percent and raising Hispanic victimization by 16 percent, with no changes for non-Hispanics.

Compatibility of our reported-crime increase with this mechanism requires that the underlying victimization rise documented by Baumer and Xie (2023) and Gonçalves and colleagues (2024) dominates the reporting decline in Jacome (2022). Their Latino-specific magnitudes are large enough to satisfy that requirement. The Hispanic-share heterogeneity in our 287(g) results, the subtype ordering in Section 4.4, and the murder-benchmark null jointly identify the chilling channel as a contributor on top of any baseline 287(g) coefficient.

Pre-trend caveat. Our 287(g) estimates have violated parallel-pre-trend tests. We do not bury this. Two readings are observationally equivalent in our event-study design. Counties experiencing rising crime may select into 287(g) adoption and continue on that trajectory after adoption, so the post-adoption “effect” is partly a continuation of a pre-existing trend. Alternatively, true policy effects may begin before the formal adoption date, because adoption is preceded by political mobilization, internal sheriff’s-office practice changes, and informal cooperation that already chills the immigrant-cooperation margin. Baumer and Xie (2023) sidestep the issue with an individual-level design in which parallel trends pass. The honest summary is that our results provide aggregate evidence consistent in sign and demographic concentration with the chilling-effect mechanism, while acknowledging that the magnitude is upper-bounded by selection-on-trends concerns.

Policy implications. The aggregate UCR evidence does not support the sanctuary-policy claim that limiting federal-local cooperation reduces crime, nor does it support the symmetric claim that limiting cooperation invites crime. It does provide aggregate evidence that 287(g) adoption is associated with higher measured rates of all five major UCR categories, concentrated where the immigrant population is large enough for the chilling mechanism to operate. Whether 287(g) raises true crime, raises only measured crime, or sits on a pre-existing rising-crime trajectory depends on which slice of the post-adoption coefficient one attributes to selection. Even under the most attenuated interpretation, the absence of any aggregate crime-reduction signal is hard to square with the program's stated public-safety rationale.

Several extensions remain open. A first-stage analysis using ICE detainer-request data would test whether either policy moves the operational margin on which our reduced-form estimates rest. A within-state border-discontinuity design exploiting adoption variation between adjacent counties on the same side of a state line would address the pre-trend concern more directly. Both are pursued in companion work.

References

[1] Adelman, Robert, Lesley Williams Reid, Gail Markle, Saskia Weiss, and Charles Jaret. 2017. "Urban Crime Rates and the Changing Face of Immigration: Evidence Across Four Decades." *Journal of Ethnicity in Criminal Justice* 15 (1): 52–77.

[2] Baumer, Eric P., and Min Xie. 2023. "Federal-Local Partnerships on Immigration Law Enforcement: Are the Policies Effective in Reducing Violent Victimization?" *Criminology and Public Policy* 22 (3): 417–455.

[3] Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. "Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review* 110 (11): 3634–3660.

[4] Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.

[5] Cohen, Lawrence E., and Marcus Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44 (4): 588–608.

[6] Comino, Stefano, Giovanni Mastrobuoni, and Antonio Nicolò. 2020. "Silence of the Innocents: Undocumented Immigrants' Underreporting of Crime and their Victimization." *Journal of Policy Analysis and Management* 39 (4): 1214–1245.

[7] Cook, Philip J. 2013. "The Great American Gun War: Notes from Four Decades in the Trenches." *Crime and Justice* 42 (1): 19–73.

[8] Forrester, Andrew, and Alex Nowrasteh. 2018. "Do Immigration Enforcement Programs Reduce Crime? Evidence from the 287(g) Program in North Carolina." *Cato Institute Working Paper* No. 52.

[9] Gonçalves, Felipe M., Elisa Jácome, and Emily K. Weisburst. 2024. "Immigration Enforcement and Public Safety." NBER Working Paper No. 32109. National Bureau of Economic Research.

[10] Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–277.

[11] Hausman, David K. 2020. "Sanctuary Policies Reduce Deportations Without Increasing Crime." *Proceedings of the National Academy of Sciences* 117 (44): 27262–27267.

[12] Jacome, Elisa. 2022. "The Effect of Immigration Enforcement on Crime Reporting: Evidence from Dallas." *Journal of Urban Economics* 131: 103451.

[13] Kaplan, Jacob. 2024. *Jacob Kaplan's Concatenated Files: Uniform Crime Reporting (UCR) Program Data, 1960–2023*. Inter-university Consortium for Political and Social Research [distributor]. `https://jacobdkaplan.com/data.html`.

[14] Kirk, David S., Andrew V. Papachristos, Jeffrey Fagan, and Tom R. Tyler. 2012. "The Paradox of Law Enforcement in Immigrant Communities: Does Tough Immigration Enforcement Undermine Public Safety?" *Annals of the American Academy of Political and Social Science* 641 (1): 79–98.

[15] Kubrin, Charis E., and Bradley J. Bartos. 2022. "Sanctuary Status and Crime in California: What's the Connection?" *Justice Evaluation Journal* 5 (1): 138–158.

[16] Lakens, Daniel. 2017. "Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses." *Social Psychological and Personality Science* 8 (4): 355–362.

[17] Light, Michael T., Jingying He, and Jason P. Robey. 2020. "Comparing Crime Rates Between Undocumented Immigrants, Legal Immigrants, and Native-Born US Citizens in Texas." *Proceedings of the National Academy of Sciences* 117 (51): 32340–32347.

[18] Light, Michael T., and Ty Miller. 2018. "Does Undocumented Immigration Increase Violent Crime?" *Criminology* 56 (2): 370–401.

[19] Miles, Thomas J., and Adam B. Cox. 2014. "Does Immigration Enforcement Reduce Crime? Evidence from Secure Communities." *Journal of Law and Economics* 57 (4): 937–973.

[20] Robinson, Joshua J. 2024. "Where Are the Guns?" Working paper. `https://joshuajrobinson.info/wherearetheguns8142024.pdf`.

[21] Sampson, Robert J. 2008. "Rethinking Crime and Immigration." *Contexts* 7 (1): 28–33.

[22] Sampson, Robert J., and W. Byron Groves. 1989. "Community Structure and Crime: Testing Social-Disorganization Theory." *American Journal of Sociology* 94 (4): 774–802.

[23] Schuirmann, Donald J. 1987. "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15 (6): 657–680.

[24] Shaw, Clifford R., and Henry D. McKay. 1942. *Juvenile Delinquency and Urban Areas.* Chicago: University of Chicago Press.

[25] Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–199.

[26] Treyger, Elina, Aaron Chalfin, and Charles Loeffler. 2014. "Immigration Enforcement, Policing, and Crime: Evidence from the Secure Communities Program." *Criminology and Public Policy* 13 (2): 285–322.

[27] Wong, Tom K. 2017. "The Effects of Sanctuary Policies on Crime and the Economy." *Center for American Progress Report.*

[28] Wooldridge, Jeffrey M. 2021. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." Working paper, Michigan State University.

Online Appendix
Subnational Immigration Enforcement and Crime: Asymmetric Effects of Sanctuary Policies and 287(g) Agreements

David Leblang
University of Virginia

A Data construction details

A.1 Panel construction and FIPS harmonization

The base panel is a quarterly county-level dataset spanning 2010Q1 through 2024Q4. The estimating sample for the sanctuary analysis contains 166,324 county-quarter observations across 3,006 counties; for the 287(g) analysis, 169,336 observations across 3,060 counties. Counties are identified by 5-digit FIPS codes carried as zero-padded strings throughout, with a numeric `geoid` panel identifier created by `encode` for use with Stata's `xtset`. All spatial joins (adjacency, neighbor lags) use the string FIPS; all panel operations use the numeric `geoid`. Alaska and Hawaii are retained for non-spatial analyses and excluded from contiguity-based spillover specifications because they have no land neighbors in the contiguous-48 adjacency graph.

A.2 Outcome construction

Crime counts come from the Kaplan compilation of FBI Uniform Crime Reporting (UCR) and National Incident-Based Reporting System (NIBRS) data, accessed via the Crime Data Explorer (cde.ucr.cjis.gov). Monthly agency-level counts are aggregated to the county-quarter level. Five outcomes are constructed: `actual_murder`, `actual_rape`, `actual_robbery`, `actual_assault`, and `actual_theft`. Per-100,000 rates use a quarterly population denominator interpolated from annual Census Bureau intercensal estimates.

A.3 Treatment construction

The sanctuary treatment date is the earliest quarter in which the FAIR sanctuary database records a county as having any sanctuary-type policy. Counties classified as sanctuary as of 2010Q1 are dropped from the staggered DiD analysis to avoid the always-treated complication. The final sanctuary treated set is 229 counties.

The 287(g) treatment integrates ICE's official program list with a corrected treatment file from the Immigrant Legal Resource Center (ILRC) that records 25 documented terminations and six documented reversals. The conservative corrected-spell specification used in the main text drops these 31 counties from the treated set. The remaining treated set is 684 counties. Section D.3 reports an absorbing-treatment robustness check that retains the 31 reverser/terminator counties; results are within sampling error of the main estimates. Counties signing under the 2025 program expansion enter the analysis as never-treated controls because they have no in-sample post-treatment crime data.

A.4 Covariate sources

Demographic variables come from the American Community Survey 5-year estimates via the Census Bureau API; economic variables from BLS LAUS and Census; political variables from the MIT Election Lab; firearms-related variables from Robinson (2024); and structural variables (urban-rural continuum, border indicator) from USDA codes. The full list of covariates entering the outcome and treatment equations is given in Section B.3.

B Stage-one machine learning details

B.1 LASSO specification

We estimate cross-validated LASSO regressions of each crime outcome on the full covariate set described in Section A.5. We use Stata 19's elastic-net command with $\alpha = 1$ for pure L_1 penalization, ten-fold cross-validation, and the within-one-standard-error rule for λ selection. Variables receiving non-zero coefficients across at least four of the five outcomes constitute the LASSO consensus set. Figure A1 shows the LASSO coefficients for the robbery outcome as a representative cell.

B.2 Random forest specification

Random forest regressions use 500 trees, $\sqrt{p}$ candidate variables at each split, and a permutation-importance criterion (mean decrease in prediction accuracy when a variable is randomly per-

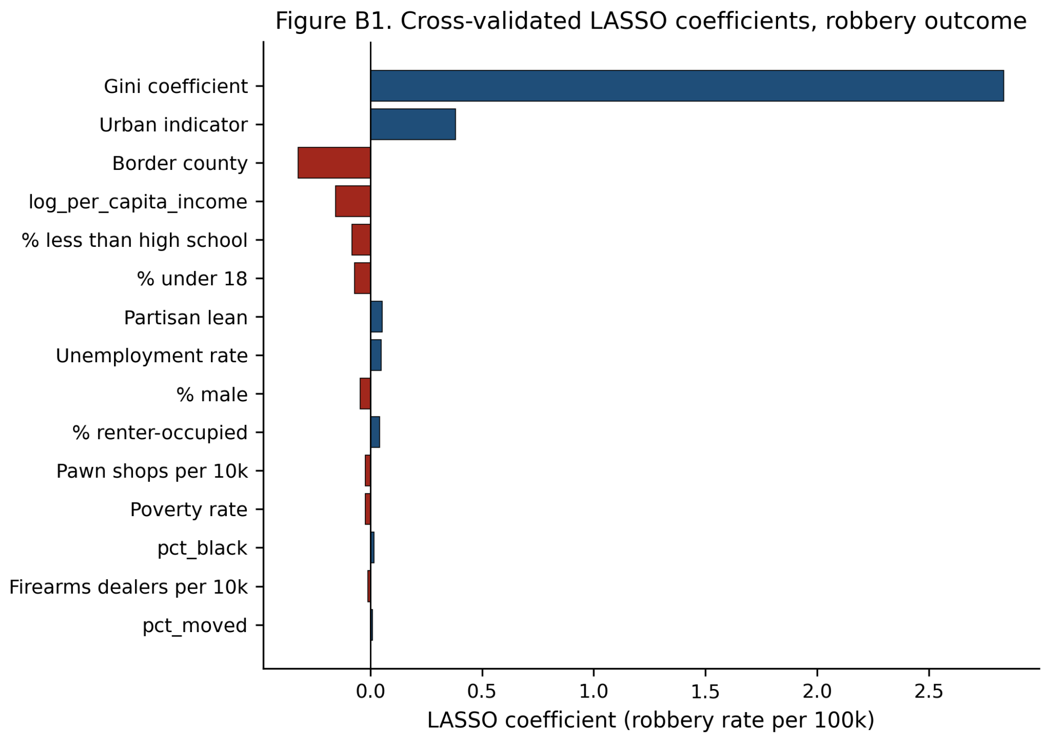

Figure A1. Cross-validated LASSO coefficients for the robbery outcome, showing the variables retained by the L_1 penalty and their signs. Blue bars are positive coefficients, red bars negative.

muted out-of-bag). The top quartile of variables by importance score across all five outcomes constitutes the random-forest consensus set. Figure A2 shows the random forest variable importance for the robbery outcome.

B.3 Consensus and elastic-net cross-check

The intersection of the LASSO consensus set and the random-forest consensus set is the outcome-equation control set used in the main AIPW and RA specifications. The consensus set is: federal-firearms-dealer rate, Gini coefficient, partisan lean, pawn-shop rate, share with less than high school education, percent male, percent renter, percent young, poverty rate, unemployment rate, urban indicator, border-county indicator, and neighbor-sanctuary share (the latter is included for the 287(g) models as a control for spatial spillover from the alternative treatment, and for symmetry in the sanctuary models). We additionally re-run the main models substituting an elastic-net selection at $\alpha = 0.5$ to verify that results are not driven by the L_1 penalty structure. Results are unchanged. Figure A3 decomposes LASSO predictive variance (R^2) by covariate group across all five outcomes; demographic and structural variables carry most of the predictive content for the higher-baseline crime outcomes.

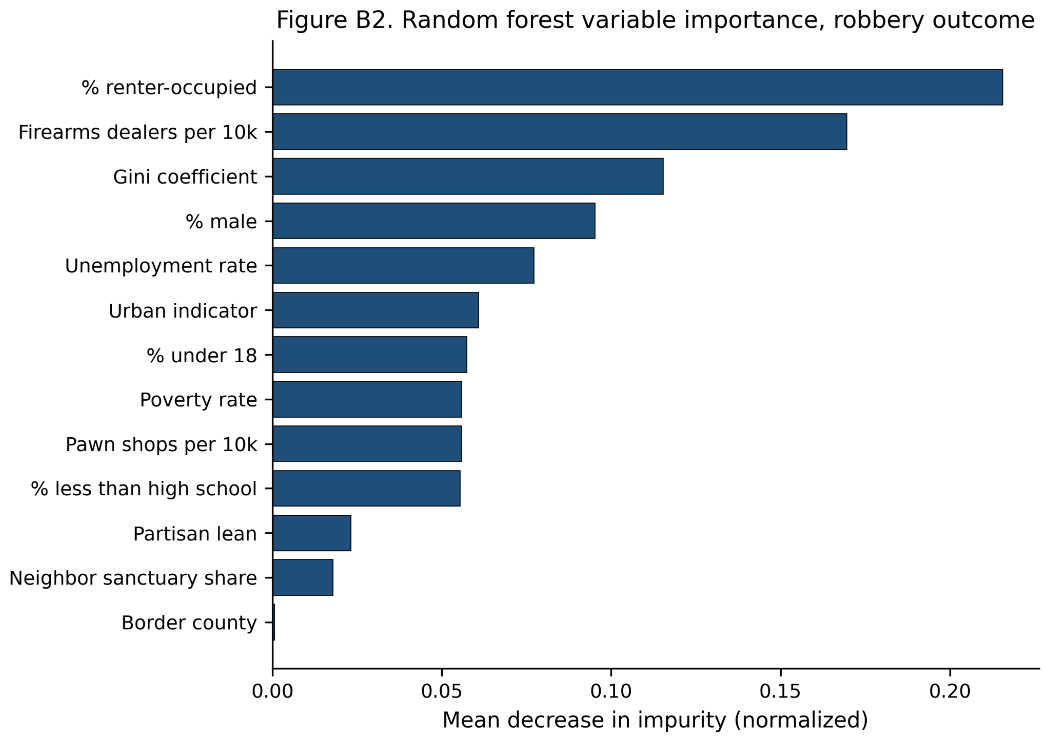

Figure A2. Random forest variable importance for the robbery outcome (mean decrease in impurity, normalized so importances sum to one). The same variables that LASSO retains in Figure A1 are the ones the random forest ranks highest.

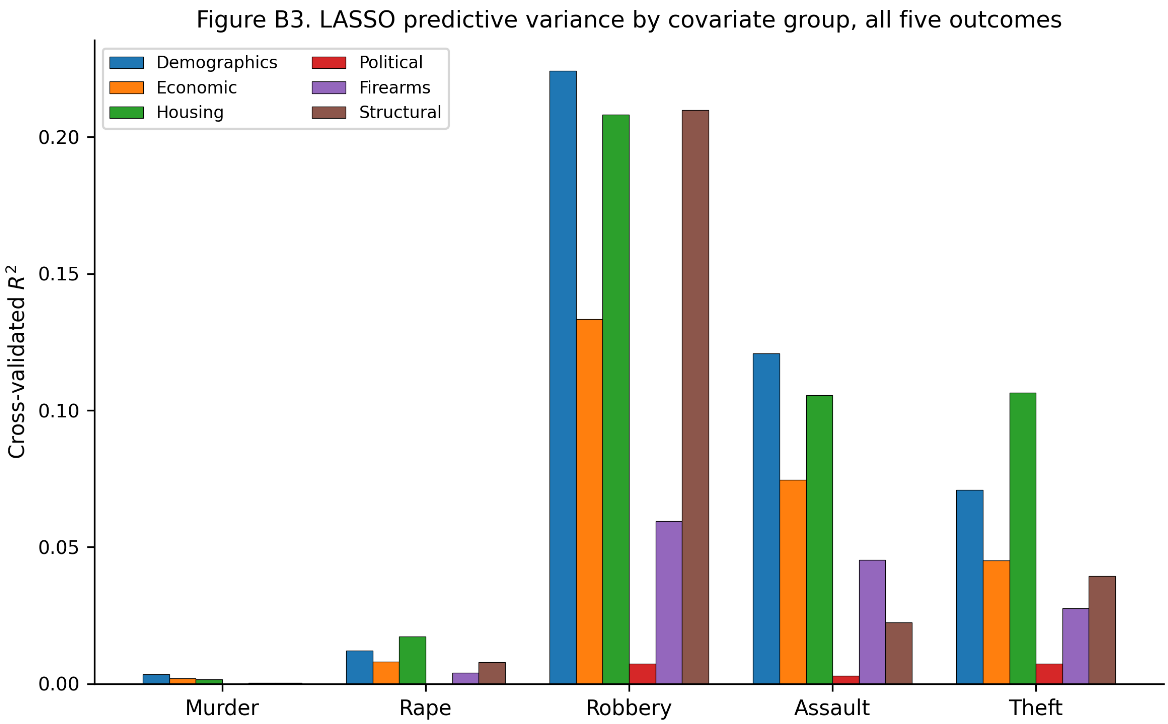

Figure A3. Cross-validated LASSO R^2 by covariate group, all five UCR outcomes. Group decomposition shows which covariate blocks carry predictive content for each outcome.

B.4 Where do the immigration policy variables rank?

In neither the LASSO nor the random forest does any sanctuary or 287(g) policy variable appear in the top quartile of importance for any of the five outcomes. The variables consistently selected are those listed in Section B.3. This is the pre-causal observation reported in Section 6 of the main text: the immigration-policy variables are not even in the prediction race for cross-sectional or longitudinal variation in measured crime.

C Full event-study results

We report dynamic event-study coefficients $\hat{\theta}_e$ for $e \in \{-8, \ldots, 4\}$ for all ten treatment-outcome cells under both AIPW and RA, with 95 percent pointwise confidence intervals and the joint pre-trend Wald p-value annotated. The 287(g) cells exhibit visible pre-treatment build-up across all five outcomes, consistent with the pre-trend Wald rejections discussed in the main text. The sanctuary cells are visually flat in the pre-period for all five outcomes under both estimators, except for marginal pre-trend behavior in the RA assault cell (Wald $p = 0.005$).

Figures A4 and A5 present these dynamics in compact 5×2 small-multiple form. Figure

A4 plots the sanctuary event studies for all five outcomes, with AIPW and RA overlaid in each panel; the pre-period coefficients hover near zero across panels and the post-period bands cross zero in each cell. Figure A5 presents the same layout for 287(g) and annotates the pre-trend Wald p-value in each panel; visible upward drift in the pre-period across all five outcomes corroborates the rejections reported in Section 4.2 of the main text. Each panel also displays the implied event-time path of the ATET, with the post-period coefficients lying above zero in each 287(g) cell. Figure A6 presents the Hispanic-share heterogeneity split discussed in Section E.1; details there.

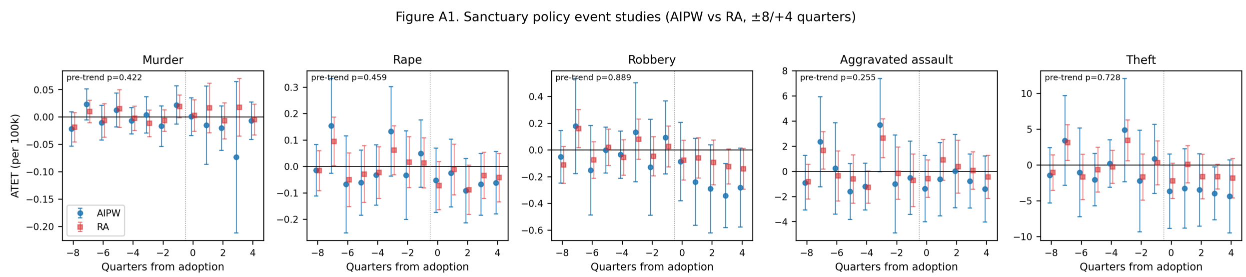

Figure A4. Sanctuary event studies, AIPW (filled circles) and RA (open squares), $\pm 8/+4$ quarters. Pre-trend Wald p-values annotated.

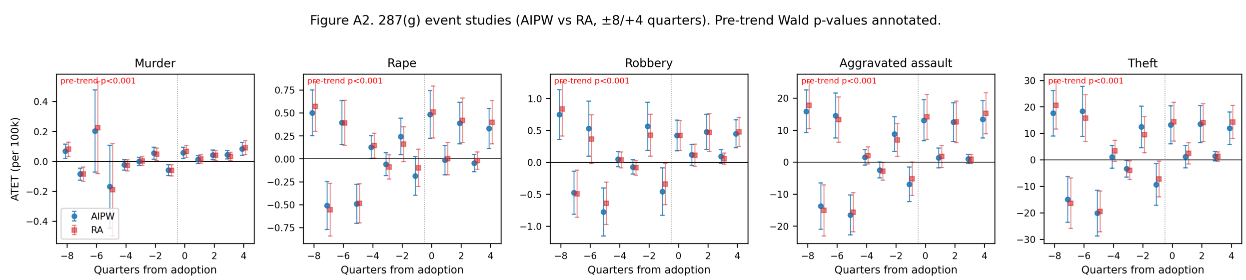

Figure A5. 287(g) event studies, AIPW (filled circles) and RA (open squares), $\pm 8/+4$ quarters. Pre-trend Wald p-values annotated; visible pre-period drift across all five outcomes.

D Robustness checks

D.1 AIPW versus RA comparison

We report the augmented inverse probability weighting (AIPW) estimates as the primary specification on principled grounds (double robustness) and the regression-adjustment (RA) estimates as the natural alternative-without-propensity-weighting comparison. Substantive disagreement between AIPW and RA indicates that the propensity model is doing real work; substantive agreement indicates that the result does not hinge on the propensity specification.

For the 287(g) cells, AIPW and RA agree on sign, on significance, and on within-sampling-error magnitude in all five cells. The implied effect-as-percent-of-pre-treatment-mean is in the 18–31 percent range under AIPW (cleaned panel; 684 treated counties) and a comparable 16–28 percent range under RA. We treat the agreement as indicating that the 287(g) result is robust to the choice between AIPW and RA.

For the sanctuary cells, AIPW and RA agree on a near-zero point estimate in four of five cells (murder, rape, assault, theft). The disagreement is on the sanctuary-on-robbery cell: the AIPW point estimate is −0.458 (SE 0.197, significant) and the RA point estimate is −0.091 (SE 0.080, not significant). This is the single estimator-fragile result in the table. We report it transparently in the main text and do not assign it independent inferential weight in the discussion.

D.2 Two-way fixed effects (not estimable)

The Wooldridge (2021) extended TWFE estimator (`xthdidregress twfe`) is the natural third estimator to report alongside AIPW and RA. On our panel it is not estimable: the implied Mundlak interaction set (cohort-by-event-time interactions multiplied by the outcome-equation covariate set) effectively saturates the design matrix for the 287(g) treatment, which has 684 treated counties spread across more than 50 adoption cohorts. The estimator fails to converge within tractable runtime. We retain the AIPW versus RA comparison reported in Section D.1 as the primary estimator-robustness exercise.

D.3 Reverser-county sensitivity (287(g))

The main 287(g) specification drops 31 counties whose 287(g) participation was either terminated or reversed during the panel, leaving 684 treated counties (the cleaned panel headline reported in Table 1). As a sensitivity check we re-estimate with all 705 treated counties under the absorbing-treatment specification. ATETs move modestly and uniformly toward zero, but every cell remains positive and significant: AIPW murder +0.029 (cleaned) vs +0.023 (absorbing), rape +0.151 vs +0.128, robbery +0.203 vs +0.207, assault +5.57 vs +4.39, theft +6.07 vs +6.29. The cleaned headline is, on balance, slightly larger than the absorbing-treatment alternative; the conservative cleaning choice does not produce the positive 287(g) finding. Pre-trend rejections are unchanged across both specifications.

D.4 Spatial panel fixed effects

We re-estimate the 287(g) own-county specifications with state-by-quarter fixed effects to absorb state-level trends in crime, enforcement, and immigration politics. The own-county positive 287(g) effect on each of the five outcomes is robust to the same state-by-quarter fixed effects, which rules out the interpretation that the headline result is a state-level confound.

D.5 Formal subgroup-difference tests

The Hispanic-share split reported in Section 4.3 of the main text and Section E.1 below produces visibly larger 287(g) point estimates in the high-Hispanic subsample. We test the high-vs-low difference formally using the contrast $\hat{\theta}_{\text{high}} - \hat{\theta}_{\text{low}}$ with a standard error computed under independence between the two subsamples ($\text{SE}_{\text{diff}} = \sqrt{\text{SE}^2_{\text{high}} + \text{SE}^2_{\text{low}}}$). The independence assumption is conservative on rejection because the never-treated control group is shared across the two subsample fits, so the true sampling covariance of the two estimates is positive and the true SE on the difference is smaller than what we report.For 287(g) on robbery (AIPW) the high-low difference is +0.239 (SE 0.143, $z = 1.67$, $p = 0.096$); for 287(g) on assault, +2.305 (SE 2.804, $z = 0.82$, $p = 0.41$). For the sanctuary RA estimates the differences are +0.122 ($z = 0.76$, $p = 0.45$) for robbery and +0.796 ($z = 0.57$, $p = 0.57$) for assault. The 287(g) robbery difference is marginal at the conventional 10 percent threshold even under the conservative SE; the assault difference is not separable from sampling noise. We treat the dose-response in point estimates as suggestive corroboration of the chilling-effect mechanism rather than independent statistical evidence of heterogeneity.

E Heterogeneity

E.1 Hispanic share

We split treated counties at the within-treatment median Hispanic share among adopters and re-estimate the main models separately for the high-Hispanic and low-Hispanic subsamples.

For 287(g) we use the AIPW estimator, which converged in both subsamples. The high-Hispanic ATET is +0.343 for robbery (SE 0.120) and +7.291 for assault (SE 2.349); the low-Hispanic ATET is +0.104 for robbery (SE 0.079, not significant) and +4.986 for assault (SE 1.531). The high-Hispanic robbery effect is roughly 3.3 times the low-Hispanic effect; the high-Hispanic assault effect is roughly 1.5 times the low-Hispanic effect.

For sanctuary the AIPW estimator did not converge for the high-Hispanic subsample. We therefore report the harmonized regression-adjustment (RA) estimates for both subsamples;

using a common estimator across the split is what the heterogeneity comparison requires. Under RA the sanctuary robbery ATET is -0.080 (SE 0.101) in the high-Hispanic subsample and -0.202 (SE 0.125) in the low-Hispanic subsample; the assault ATET is $+1.123$ (SE 1.182) high and $+0.327$ (SE 0.728) low. None of the four cells is significant. The Hispanic-split null is not driven by aggregation across heterogeneous subgroups.

The estimator choice across treatments is transparently asymmetric (AIPW for 287(g), RA for sanctuary). The reason is convergence, not preference. The AIPW sanctuary low-Hispanic point estimates are larger in magnitude than the RA estimates we report, but using AIPW low-Hispanic estimates against missing AIPW high-Hispanic estimates does not allow a within-treatment comparison.

Figure A6 presents the Hispanic-share split visually for both treatments, with high-Hispanic and low-Hispanic ATETs overlaid in two side-by-side panels (sanctuary RA on the left, 287(g) AIPW on the right) for murder, robbery, and assault. The visual contrast (clear high-vs-low gradient on the 287(g) right panel for robbery and assault, no comparable gradient for sanctuary) corroborates the numerics above.

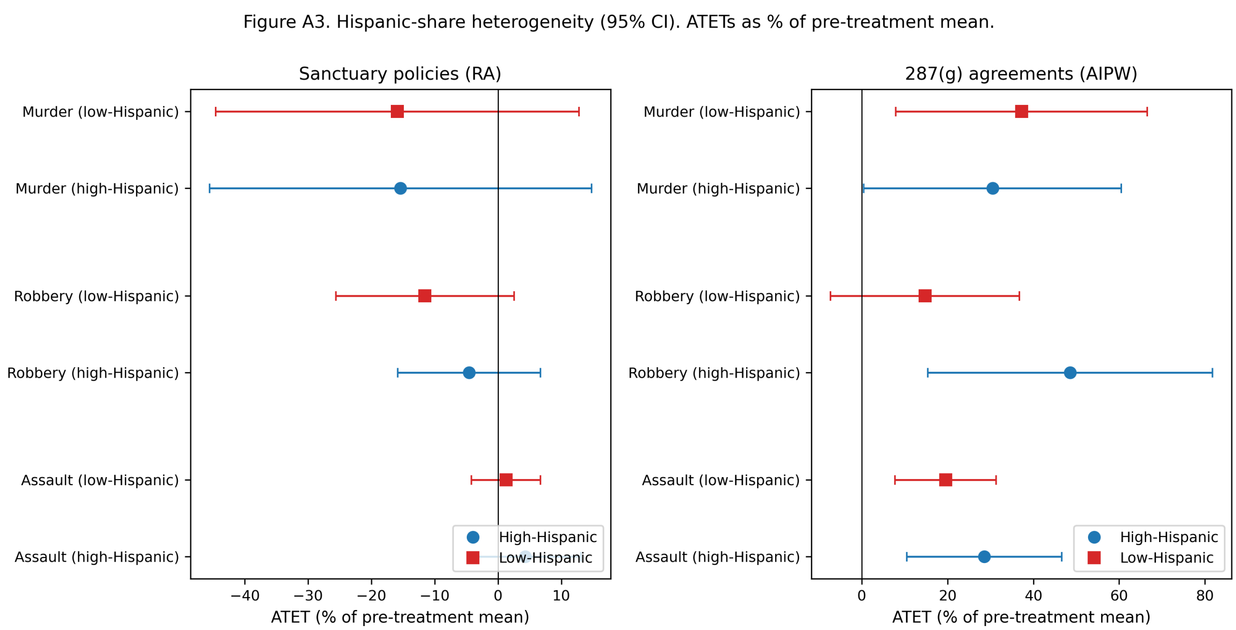

Figure A6. Hispanic-share heterogeneity (95% CI). ATETs as percentage of pre-treatment outcome mean. Left panel: sanctuary policies under RA. Right panel: 287(g) under AIPW.

E.2 287(g) program subtype

We disaggregate 287(g) effects by program subtype: Jail Enforcement Model (JEM, in-custody screening only), Warrant Service Officer (WSO, in-jail warrant execution), and Task

Force Model (TFM, deputization in the field). For each subtype we redefine the treated set to counties whose first 287(g) agreement was of that subtype, drop counties that adopted a different subtype to keep treated-vs-never contrasts clean, and re-estimate the AIPW model on all five UCR outcomes. Estimates are reported in Table E1; the corresponding forest plot is Figure 2 in the main text. Pre-trend Wald tests reject parallel trends in every subtype-outcome cell, so the level estimates carry the Section 4.2 caveat. The relative ordering of point estimates, however, is what the chilling-cooperation mechanism predicts and what Baumer and Xie (2023) report.

Table E1. 287(g) ATETs by program subtype (AIPW, never-treated controls).

Subtype	N treated	Murder ATET (SE)	Rape ATET (SE)	Robbery ATET (SE)	Assault ATET (SE)	Theft ATET (SE)
JEM (Jail Enforcement)	205	+0.031 (0.014)*	+0.119 (0.073)	+0.080 (0.062)	+3.034 (1.521)*	+3.443 (1.701)*
WSO (Warrant Service)	472	+0.044 (0.012)*	+0.189 (0.073)*	+0.189 (0.059)*	+5.967 (1.486)*	+6.955 (1.645)*
TFM (Task Force)	368	+0.044 (0.014)*	+0.240 (0.085)*	+0.256 (0.068)*	+7.353 (1.776)*	+8.546 (1.956)*

Notes: Pre-trend Wald $p \approx 0.000$ in every cell. Standard errors clustered at the county level. * denotes statistical significance at the 5% level. Effect-as-percent-of-pre-treatment-mean ratios (TFM ÷ JEM): murder 1.4×, rape 2.0×, assault 2.7×, theft 3.0×, robbery 4.5×.

Three patterns are visible. First, the TFM > WSO > JEM ordering holds for *every* outcome, not only for the two reporting-sensitive cells where the prior literature has focused. Second, the TFM/JEM ratio rises monotonically as the outcome becomes more dependent on civilian reporting: the gap is smallest for murder (where bodies are counted regardless of cooperation) and largest for robbery (where reporting depends on the victim's willingness to engage with police). Third, JEM × robbery and JEM × rape are the only insignificant cells in the table; JEM is the booking-only subtype with least street visibility, where the chilling channel should be smallest. Our county-quarter JEM estimates are not as cleanly null as the individual-level NCVS estimates of Baumer and Xie (2023), but the rank ordering and the gradient pattern reproduce their key contrast in independent county data.

E.3 Murder as a low-discretion benchmark

The chilling-effect mechanism operates through victim and witness cooperation with police: enforcement intensification reduces reporting and clearance for crimes whose recording depends on civilian cooperation. Murder is a low-discretion outcome in this sense: bodies are counted regardless of whether the victim's family cooperates with investigators, and the medical-examiner system supplies an independent recording channel that does not run through immigrant-community cooperation. The mechanism therefore predicts that murder should exhibit no Hispanic-share heterogeneity even in jurisdictions where reporting-sensitive outcomes (robbery, assault) do.

We split treated counties on the within-treatment median Hispanic share and re-estimate the murder ATET in each subsample, using AIPW for 287(g) (where it converges in both halves) and the harmonized RA estimator for sanctuary. Estimates are reported in Table E2.

Table E2. Murder ATET by Hispanic-share subgroup.

Treatment	Subgroup	Estimator	ATET (SE)	95% CI
287(g)	high-Hispanic	AIPW	+0.028 (0.014)	[+0.0004, +0.056]
287(g)	low-Hispanic	AIPW	+0.035 (0.014)	[+0.007, +0.062]
Sanctuary	high-Hispanic	RA	−0.014 (0.014)	[−0.040, +0.013]
Sanctuary	low-Hispanic	RA	−0.014 (0.013)	[−0.039, +0.011]

The 287(g) murder estimates are essentially identical across the high- and low-Hispanic subsamples (+0.028 vs +0.035), against the 3.3-fold heterogeneity observed for robbery in the same split (Section E.1). This is the cleanest single piece of mechanism evidence available in our data: if the Hispanic-share heterogeneity in robbery and assault were driven by a nonreporting channel (selection on trends correlated with Hispanic share, a true criminogenic effect of 287(g) operating differentially in Hispanic communities), it should also appear for murder. It does not.

The diagnostic also identifies a feature that the chilling story alone does not explain. The 287(g) murder level estimate is positive and significant in *both* Hispanic subgroups, at roughly 30–40 percent of the pre-treatment mean of 0.09 per 100,000. A pure chilling-of-reporting mechanism would not generate a positive murder coefficient because murder reporting is mechanism-independent. The honest reading is that the headline 287(g) estimates combine a baseline component (visible in the murder cells, consistent with the Section 4.2 pre-trend caveat and likely reflecting selection on trajectory) with an additional Hispanic-share-amplified component on reporting-sensitive crimes that the chilling channel can plausibly explain.

F Code and data availability

All analysis code is written in Stata 19 and Python 3 and available at [TBD: repository URL] upon publication. Crime outcome data are publicly available via the FBI Crime Data Explorer (cde.ucr.cjis.gov) and the Kaplan UCR replication archive. Sanctuary policy data are publicly available from the Federation for American Immigration Reform and the Immigrant Legal Resource Center. 287(g) agreement data are publicly available from U.S. Immigration and Customs Enforcement and TRAC at Syracuse University. Demographic, economic, and political covariates are publicly available from the U.S. Census Bureau, the Bureau of Labor Statistics, the MIT Election Lab, and the U.S. Department of Agriculture. A self-contained replication package, including the cleaned panel, the corrected 287(g) treatment file, the master Stata pipeline, the Python deliverable scripts, and a verification script that compares re-run output against the working analysis to numerical tolerance, will be deposited at [TBD: archive DOI] upon publication.