Does Employer-Based Immigration Enforcement Reduce the Undocumented Immigrant Population? Re-examining the Effects of LAWA*

Hugo Sant'Anna   Samyam Shrestha

May 12, 2026

Abstract

This paper replicates Bohn, Lofstrom, and Raphael (2014), who use the synthetic control method to estimate the effects of Arizona's Legal Arizona Workers Act (LAWA) on the state's likely unauthorized immigrant population. Our narrow replication confirms their main finding: LAWA led to a 1.5-2 percentage point decline in the noncitizen Hispanic share. In a wide replication, we apply the augmented synthetic control method and synthetic difference-in-differences and extend the sample through 2015. The core finding is robust to these alternatives, though the effect attenuates after 2011. Migration data suggest responses along both outflow and inflow margins.

Keywords: LAWA, immigration enforcement, undocumented immigrants, E-Verify, synthetic control

JEL Codes: C23, J15, J61, K37

---

*Hugo Sant'Anna (hsantanna@uab.edu), Collat School of Business, University of Alabama at Birmingham. Samyam Shrestha (samyam@uga.edu), Department of Agricultural and Applied Economics, University of Georgia. *Data Availability Statement*: The data used in this paper are drawn from publicly available sources. Current Population Survey (CPS) microdata are available from IPUMS-CPS. American Community Survey (ACS) microdata are available from IPUMS-USA. Replication code and constructed datasets will be deposited in the JAE Data Archive upon acceptance. *Funding Statement*: The authors declare that this research received no external funding. *Conflict of Interest Disclosure*: The authors declare no conflict of interest.

1 Introduction

Over the past several decades, interior immigration enforcement in the United States has intensified substantially. These policies generally fall into a few broad categories. One approach is police-based enforcement, in which local or federal law enforcement agencies identify and apprehend undocumented immigrants. Another is government-service-based enforcement, where access to public services or benefits, such as driver's licenses, public housing, healthcare, or school enrollment, is conditioned on proof of legal status. A distinct third approach is labor-market-based enforcement, in which employers are required to verify the work authorization of new hires. Because this approach operates indirectly through firms' hiring decisions rather than through direct interaction with immigrants themselves, it may generate fundamentally different economic responses than other forms of immigration enforcement.

One of the clearest examples of this labor-market-based approach is the 2007 Legal Arizona Workers Act (LAWA), which was arguably among the strictest state-level immigration enforcement laws at the time. The law required all employers in Arizona to use E-Verify, a federal electronic system for verifying the work authorization of newly hired employees, and imposed sanctions on firms found to knowingly hire unauthorized workers. By increasing the expected costs of employing undocumented immigrants, LAWA was intended to reduce employer demand for unauthorized labor through the labor market itself rather than through direct apprehension or removal. Implemented in January 2008, the law substantially expanded employer participation in E-Verify in Arizona and became a central case in the literature on immigration enforcement and labor markets.

This paper replicates and extends Bohn, Lofstrom, and Raphael (2014), which examines the effects of LAWA on the likely undocumented population in Arizona. Using the Synthetic Control Method (SCM), Bohn, Lofstrom, and Raphael (2014) find that LAWA led to a significant reduction in the share of Arizona residents most likely to be unauthorized. We begin with a narrow replication, confirming the paper's main findings. While our estimates differ marginally, they are consistent in direction, magnitude, and significance. We also assess implementation robustness by re-estimating the model across seven SCM implementations in R and Python following Becker and Klößner (2017). The estimated effects remain highly consistent across implementations.

We then conduct replication in a wide sense by extending the analysis along several dimensions. First, we apply recently developed alternatives to SCM, including the augmented synthetic control method (Ben-Michael, Feller, and Rothstein, 2021) and synthetic difference-in-differences (Arkhangelsky et al., 2021), both of which offer improved bias correction and more reliable inference than standard SCM. Second, we extend the sample through 2015, beyond the original paper's window of up to 2009, to assess how the effect evolves over a longer horizon. The original finding holds across all alternative estimators, though the extended sample reveals substantial attenuation after 2011.

Third, we decompose the reduction in the share of likely undocumented immigrants into inflow and outflow margins using the American Community Survey (ACS). Specifically, we use the ACS to construct bilateral state-level migration flows for noncitizen Hispanic individuals with a high school diploma or less, and estimate a Poisson Pseudo-Maximum Likelihood gravity difference-in-differences model to quantify the extent to which the decline in Arizona's immigrant population reflects redirected interstate migration versus deterred inflows. We find that LAWA nearly doubled out-migration of this population from Arizona to neighboring states and reduced inflows into Arizona by 36-58 percent, with outflows responding at signing and inflows responding at implementation.

The evidence strengthens the contribution of Bohn, Lofstrom, and Raphael (2014) along three dimensions: the population effect is robust to implementation choices, operates through both outflow displacement and inflow deterrence with distinct timing, and while sustained through 2015, attenuates after 2011, pointing to partial reversion over the longer run.

2 Replication in a Narrow Sense

In this section, we provide a replication in the narrow sense of Bohn, Lofstrom, and Raphael (2014), reproducing the paper's main results using the original methods and samples. Following Bohn, Lofstrom, and Raphael (2014), we use monthly Current Population Survey (CPS) data from 1998-2009 and treat Arizona beginning in 2008. We construct a synthetic control as a weighted average of other U.S. states chosen to match Arizona's pre-treatment characteristics and outcome trajectory over the 1998-2006 period. States with broad E-Verify mandates (Mississippi, Rhode Island, South Carolina, and Utah) are excluded from the donor pool. We treat 2007 as a transition year because LAWA was signed in mid-2007 but implemented in January 2008.

Table A.1 reports trends in population groups of our interest over the 1998-2015 period. The values perfectly match those reported in Table 1 of Bohn, Lofstrom, and Raphael (2014), with only discrepancies in 2001-2003, likely reflecting updates to CPS extracts and population weights.

Formally, let Y_{st} denote the outcome for state s in year t. SCM chooses weights $w_j \geq 0$ (summing to one) on donor states j to minimize the pre-treatment discrepancy:

$$\min_{w} \sum_{t \leq T_0} \left(Y_{\mathrm{AZ},t} - \sum_{j \in \mathcal{D}} w_j Y_{jt} \right)^2, \tag{1}$$

where T_0 is the last pre-LAWA year. The estimated treatment effect at time $t > T_0$ is the gap between Arizona and its synthetic control:

$$\widehat{\tau}_t = Y_{\mathrm{AZ},t} - \sum_{j \in \mathcal{D}} w_j Y_{jt}. \tag{2}$$

The matching procedure uses pre-treatment outcome values along with state-level labor market and demographic characteristics, including industry composition, educational attainment shares, and unemployment rates. These predictors help ensure that Arizona is compared to states with similar economic conditions and exposure to the Great Recession prior to LAWA implementation. We assess statistical significance using permutation (placebo) tests that iteratively assign treatment to donor states and compare Arizona's post-treatment gap to the placebo distribution.

Table A.2 reports the weights assigned to each donor state for the different outcome variables. While the donor states and weights differ somewhat from those reported in Bohn, Lofstrom, and Raphael (2014), such differences are common in SCM applications because multiple donor combinations can generate very similar pre-treatment fit. Small differences in data revisions, optimization routines, or predictor weighting can therefore lead to alternative donor compositions without materially changing the estimated treatment effect.

Figure 1 plots Arizona's noncitizen Hispanic share (ages 15-45, high school diploma or less) alongside its synthetic control from 1998 to 2009. The two series track closely throughout the pre-treatment period, diverge sharply after 2007, and remain separated through 2009. The small pre-treatment RMSPE indicates close pre-LAWA fit.

Figure 1(b) displays the permutation test. Arizona's post-treatment gap is the most negative in the distribution, lying well below the placebo gaps for all 46 donor states. This suggests that the estimated decline is unlikely to be driven by chance or broader contemporaneous shocks affecting other states.

Table 1 reports the difference-in-differences (DD) estimates across several outcome definitions and pre-treatment windows. For the main outcome, the share of noncitizen Hispanics aged 15-45 with a high school diploma or less, the DD estimate using the full 1998-2006 pre-treatment period is -0.022, with Arizona ranked first out of 47 states in the permutation distribution ($p = 0.022$). The results are consistent across broader population definitions: the estimated effect for all noncitizen Hispanics aged 15 and older is -0.017 ($p = 0.023$), while the effect for noncitizen Hispanics in the full resident population is -0.015 ($p = 0.021$). These magnitudes closely align with those reported in Bohn, Lofstrom, and Raphael (2014). Appendix Table D.2 and Figure D.1 show that the SCM estimate remains negative when the highest-weighted donor states are excluded. Table A.3 shows that the results are robust to alternative pre-treatment windows and to restricting the donor pool to states that do not border Arizona, consistent with the findings reported in Table 5 of Bohn, Lofstrom, and Raphael (2014).

In addition, Table A.4 reports falsification tests using naturalized Hispanic citizens, a group not directly affected by employer verification mandates. Across all estimators, outcome definitions, and sample periods, the estimated effects are small and statistically insignificant. Figure A.1(a) plots Arizona and its synthetic control for this group. In contrast to the main results, the two series remain closely aligned throughout the post-treatment period, showing no evidence of a treatment effect and supporting the interpretation that LAWA primarily affected noncitizen populations.

Becker and Klößner (2017) document that the same synthetic control specification can yield economically different estimates across software, because the original Abadie, Diamond, and Hainmueller (2011) optimizer solves a non-convex outer problem in the predictor weight matrix. We re-estimate our main outcome through seven independent implementations on identical inputs: `Synth`, `tidysynth`, `scpi` (Cattaneo, Feng, and Titiunik, 2021), `synthdid`, and `gsynth` in R, and `pysyncon`'s `Synth` and `AugSynth` in Python. See Appendix B for details.

All seven implementations return a negative effect, with the five direct synthetic control estimates concentrated in $[-0.022, -0.017]$. The residual variation is interpretable: implementations solving only the simplex-constrained problem on outcome lags (`tidysynth`, `scpi`, `Synth` restricted to outcome lags) return an ATE of -0.020, while the original `Synth` with the Ben-Michael, Feller, and Rothstein (2021) predictor list returns -0.022. This within-`Synth` specification choice fully accounts for the cross-package gap, sharpening Kaul et al. (2022): predictor selection, not solver or language, drives implementation-level variation. The `scpi` prediction interval is $[-0.052, -0.007]$, bounded away from zero, and the headline finding survives substitution of solver, language, and software team.

3 Replication in a Wide Sense

3.1 Augmented Synthetic Control Method

To assess the robustness of the baseline SCM estimates, we implement the Augmented Synthetic Control Method (ASCM) proposed by Ben-Michael, Feller, and Rothstein (2021). ASCM augments the standard synthetic control with a ridge-regularized outcome model estimated on the donor pool, correcting for bias when pre-treatment fit is imperfect and allowing limited extrapolation beyond the convex hull. The ridge penalty is selected via cross-validation. This bias correction directly addresses the concern that standard SCM estimates may be sensitive to the convex-hull restriction or finite-sample imbalance. See Appendix C.1 for more details on this method.

Table 2, Panels D-F, reports the ASCM estimates. For the main outcome, noncitizen Hispanics aged 15-45 with a high school diploma or less, the ASCM DD estimate using the full pre-period is -0.021 ($p = 0.021$), virtually identical to the SCM estimate of -0.022. The broader noncitizen Hispanic population aged 15-45 shows a DD of -0.027 ($p = 0.021$), and the all-ages noncitizen Hispanic share declines by 0.016 ($p = 0.021$). The ridge-based bias correction shifts the estimates by at most 0.001 in absolute value relative to SCM, indicating that the original synthetic control achieved adequate pre-treatment fit and that the estimated LAWA effect is not driven by extrapolation beyond the convex hull of the donor pool or finite-sample bias in the standard synthetic control estimator.

3.2 Synthetic Difference-in-Differences

As an additional robustness exercise, we implement the synthetic difference-in-differences (SDID) estimator of Arkhangelsky et al. (2021). SDID combines the reweighting logic of synthetic control with the double-differencing structure of DID. The estimator constructs non-negative unit weights and time weights that balance pre-treatment outcomes, then estimates the treatment effect via a weighted two-way fixed effects regression. This hybrid design is more tolerant of imperfect pre-treatment fit than either method alone. See Appendix C.2 for more details on this method.

Table 2, Panels G-I, reports the SDID estimates. For the main outcome, the SDID DD using the full pre-period is -0.027 ($p = 0.021$), and the broader noncitizen Hispanic 15-45 group shows -0.030 ($p = 0.021$). All outcomes produce negative estimates with Arizona ranked 1st in the permutation distribution. The SDID point estimates are slightly larger in magnitude than SCM and ASCM because SDID's unit and time weights produce a different counterfactual that more aggressively down-weights pre-treatment periods where Arizona's gap was already slightly negative (see Appendix D.3). Appendix Figure D.2 shows no significant pre-treatment trend and the SDID time weights.

3.3 Extended Time Period

Bohn, Lofstrom, and Raphael (2014) end their analysis in 2009, largely because Arizona's SB1070, a broad omnibus immigration enforcement law requiring local law enforcement to verify immigration status during lawful stops when reasonable suspicion existed, was enacted in 2010. This makes it difficult to separately identify the effects of LAWA from those of the subsequent immigration policy. However, Amuedo-Dorantes and Lozano (2015) find that SB1070 had minimal independent effects on the noncitizen Hispanic population in Arizona. We thus extend the sample through 2015 to assess the longer-run effects of LAWA.

The right half of Table 2 reports the extended (1998-2015) estimates. The SCM DD moves from -0.022 to -0.020 and the ASCM estimate is stable at -0.021, both negligible changes. The SDID estimate grows from -0.027 to -0.038 as additional post-treatment periods receive weight.

Figure 2 plots Arizona against its synthetic control over the full window. The series diverge sharply after 2007, reaching a trough of approximately -0.03 around 2009-2011, before narrowing steadily to roughly -0.005 by 2015, suggesting partial reversion over the longer horizon. Whether this reflects fading enforcement salience, compositional changes in the CPS, or unrelated differential trends remains an open question, though the average post-treatment effect remains negative across all three estimators (Figure A.2).

Because the extended sample overlaps with SB 1070 (key provisions effective 2011), we estimate effects separately for the LAWA-only period (2008-2010) and the post-SB 1070 period (2011-2015; Table D.1). All three estimators confirm that the decline is concentrated in the immediate post-LAWA window (estimates from -0.015 to -0.030), while post-SB 1070 marginal effects are small, mixed in sign, and statistically insignificant. The post-2011 attenuation thus reflects partial reversion of LAWA's impact rather than a confounding SB 1070 effect.

3.4 Population Outflow versus Inflow

The SCM estimates confirm that LAWA reduced Arizona's likely unauthorized immigrant population, but Bohn, Lofstrom, and Raphael (2014) do not examine the adjustment margin underlying this decline. As a further check on the mechanism, we ask whether the population response they document reflects displacement of existing residents, deterrence of prospective migrants, or both. This decomposition helps interpret their main finding: a policy that primarily induces out-migration is geographic redistribution, while one that deters inflows represents a more durable reduction in unauthorized presence.

To do so, we use ACS microdata 2005-2010 to construct population-weighted bilateral interstate migration flows for working-age Hispanic noncitizens with a high-school diploma or less, the same demographic group as the Bohn, Lofstrom, and Raphael (2014) headline outcome. We estimate gravity-style Poisson difference-in-differences models with saturated origin-destination and state-by-year fixed effects to separately identify post-LAWA changes in outflows from and inflows to Arizona. Appendix E provides the full specification, identifying variation, and event-study framework.

Three findings emerge. First, AZ-to-neighbor flows rose by approximately 94 % post-

LAWA-signing ($\hat{\beta} = 0.661$, $p = 0.010$), robust to less saturated fixed effects and log-OLS, with pre-trend Wald tests passing under both specifications. Second, inflows to Arizona fell by 36 % ($\hat{\beta} = -0.448$, $p = 0.061$) under implementation-only timing, and the neighbor-state mirror specification yields -58 % ($p < 0.001$). Third, the timing differs: outflow jumps in the LAWA-signing window (ACS 2008, capturing calendar-2007 flows), while inflow drops one year later at enforcement onset (ACS 2009 onward). Together, these results indicate that the population decline documented by Bohn, Lofstrom, and Raphael (2014) reflects both geographic redistribution and ex-ante deterrence, with the two channels operating through distinct behavioral responses to signing versus implementation.

4 Conclusion

In this paper, we replicate Bohn, Lofstrom, and Raphael (2014) and confirm that LAWA reduced Arizona's noncitizen Hispanic population share by approximately 1.5-2 percentage points, a result robust to three estimators: the original SCM, the augmented synthetic control method of Ben-Michael, Feller, and Rothstein (2021), and the synthetic difference-in-differences estimator of Arkhangelsky et al. (2021), as well as across seven independent software implementations of the SCM framework. Extending the sample through 2015 reveals substantial attenuation after 2011, with the gap narrowing from roughly 3 percentage points to less than 1 percentage point by 2015; falsification tests on naturalized Hispanic citizens detect no effect. Using ACS bilateral migration data, a Poisson gravity difference-in-differences model estimates a 94 % post-LAWA increase in outflows to neighboring states and a 36-58% decline in inflows, with outflows responding at signing and inflows responding at implementation, indicating that LAWA operated through both geographic redistribution and ex-ante deterrence of prospective migrants. Overall, the evidence suggests that the original LAWA finding is not only reproducible, but also robust to alternative estimators, software implementations, and longer-run evaluation.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program”. In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2011). “Synth: An R Package for Synthetic Control Methods in Comparative Case Studies”. In: *Journal of Statistical Software* 42.13, pp. 1–17. DOI: 10.18637/jss.v042.i13.

Amuedo-Dorantes, C. and F. Lozano (2015). “On the Effectiveness of SB1070 in Arizona”. In: *Economic Inquiry* 53.1, pp. 335–351.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021). “Synthetic Difference-in-Differences”. In: *American Economic Review* 111.12, pp. 4088–4118.

Becker, M. and S. Klößner (2017). “Estimating the Economic Costs of Organized Crime by Synthetic Control Methods”. In: *Journal of Applied Econometrics* 32.7, pp. 1367–1369. DOI: 10.1002/jae.2572.

Ben-Michael, E., A. Feller, and J. Rothstein (2021). “The Augmented Synthetic Control Method”. In: *Journal of the American Statistical Association* 116.536, pp. 1789–1803.

Bohn, S., M. Lofstrom, and S. Raphael (2014). “Did the 2007 Legal Arizona Workers Act Reduce the State’s Unauthorized Immigrant Population?” In: *Review of Economics and Statistics* 96.2, pp. 258–269.

Cattaneo, M. D., Y. Feng, and R. Titiunik (2021). “Prediction Intervals for Synthetic Control Methods”. In: *Journal of the American Statistical Association* 116.536, pp. 1865–1880. DOI: 10.1080/01621459.2021.1979561.

Hoekstra, M. and S. Orozco-Aleman (2017). “Illegal Immigration, State Law, and Deterrence”. In: *American Economic Journal: Economic Policy* 9.2, pp. 228–252.

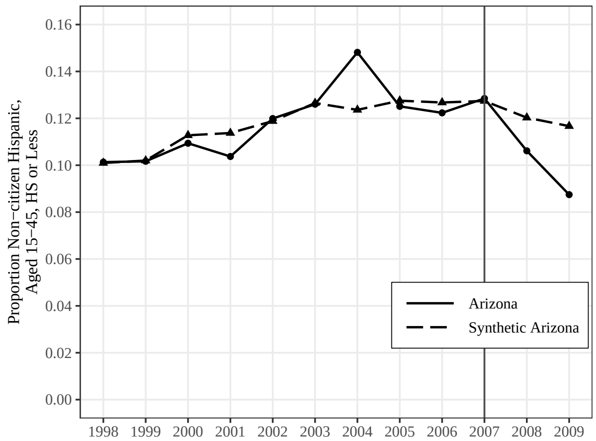
Kaul, A., S. Klößner, G. Pfeifer, and M. Schieler (2022). “Standard Synthetic Control Methods: The Case of Using All Preintervention Outcomes Together with Covariates”. In: *Journal of Business & Economic Statistics* 40.3, pp. 1362–1376. DOI: 10.1080/07350015.2021.1930012.

Klößner, S., A. Kaul, G. Pfeifer, and M. Schieler (2018). “Comparative Politics and the Synthetic Control Method Revisited: A Note on Abadie et al. (2015)”. In: *Swiss Journal of Economics and Statistics* 154.1, p. 11. DOI: 10.1186/s41937-017-0004-9.

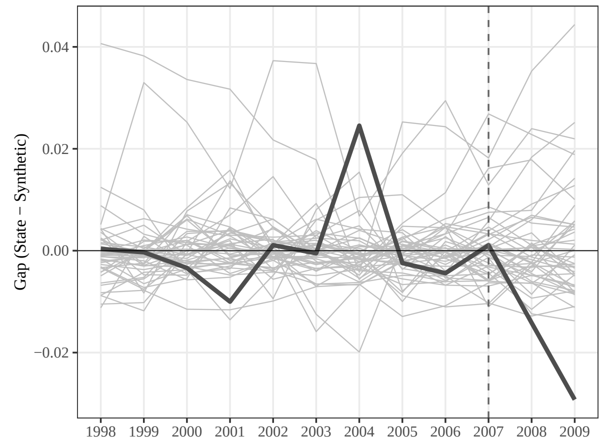
Santos Silva, J. M. C. and S. Tenreyro (2006). “The Log of Gravity”. In: *Review of Economics and Statistics* 88.4, pp. 641–658. DOI: 10.1162/rest.88.4.641.

Xu, Y. (2017). “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models”. In: *Political Analysis* 25.1, pp. 57–76. DOI: 10.1017/pan.2016.2.

Figures and Tables



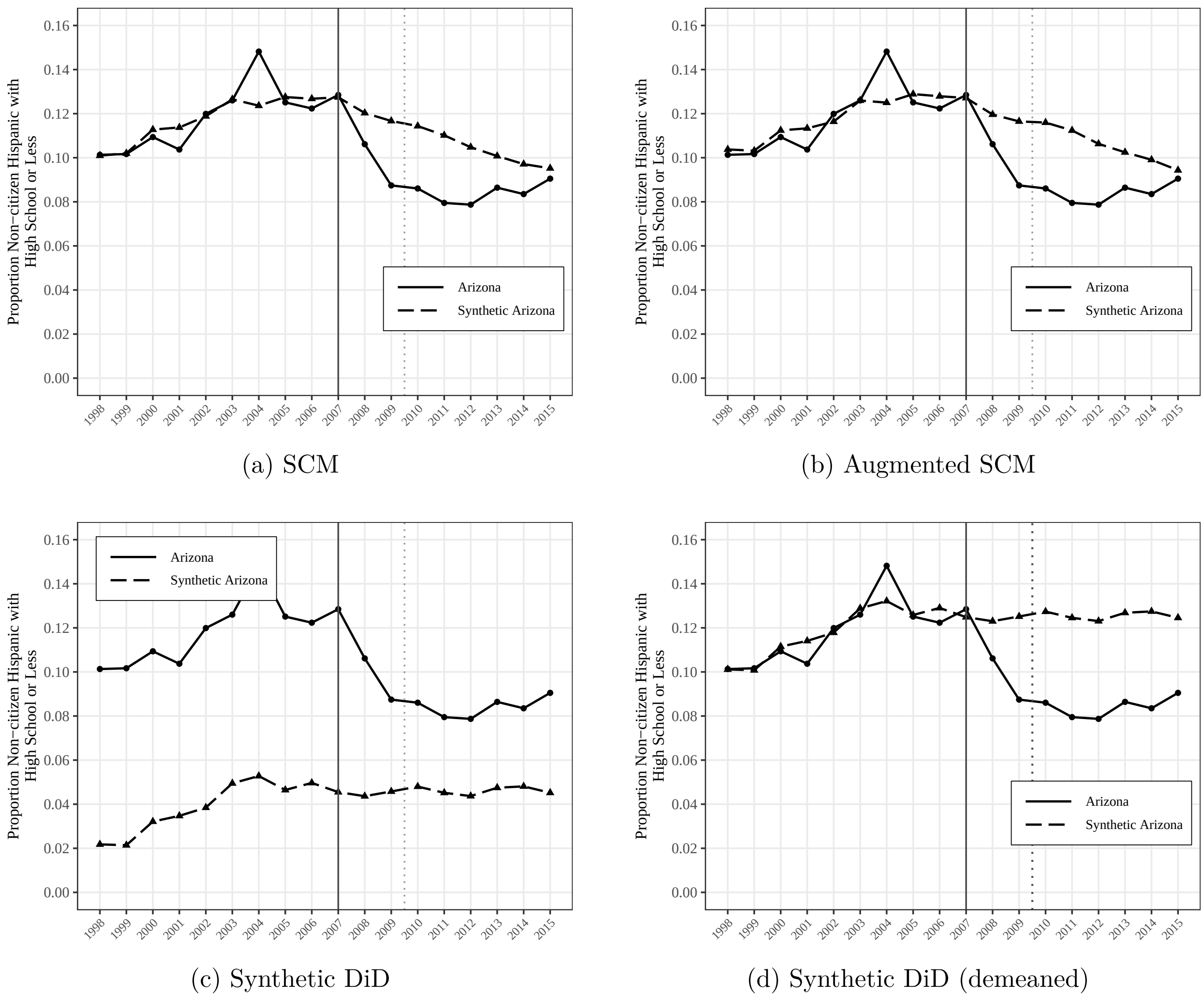
(a) Arizona vs. Synthetic Arizona, 1998-2009



(b) Permutation test: Arizona (thick line) vs. 46 placebos

Figure 1: Synthetic control estimates of the effect of LAWA on the proportion of noncitizen Hispanics with a high school diploma or less among prime working-age persons, 1998-2009

Notes: Panel (a) plots the outcome for Arizona and its synthetic control constructed from the donor pool of 46 states, excluding states with broad E-Verify mandates (MS, RI, SC, UT). Panel (b) plots the gap (state minus synthetic) for all states; Arizona is shown with the thick black line. The vertical dashed line marks 2007, the year LAWA was signed.



(a) SCM

(b) Augmented SCM

(c) Synthetic DiD

(d) Synthetic DiD (demeaned)

Figure 2: Arizona vs. Synthetic Arizona over the extended period, 1998-2015

Notes: Panels (a)-(c) plot Arizona and its synthetic control for the main outcome (proportion of noncitizen Hispanics aged 15-45 with a high school diploma or less) over the extended 1998-2015 window under three estimators: (a) the original SCM of Abadie, Diamond, and Hainmueller, 2010, (b) the augmented SCM of Ben-Michael, Feller, and Rothstein (2021), and (c) the synthetic difference-in-differences of Arkhangelsky et al. (2021). Panel (d) replots the SDID counterfactual after adding the mean pre-treatment gap to the synthetic line, so that pre-treatment levels align and the post-treatment divergence is visible on the same scale as panels (a)-(b). The solid vertical line marks 2007 (LAWA signed); the dotted vertical line marks 2009, the endpoint of the original Bohn, Lofstrom, and Raphael, 2014 analysis.

Table 1: Narrow Replication: Difference-in-Differences Estimates of the Effect of LAWA on Population Shares, 1998-2009

	Mean difference: AZ − Synthetic AZ			DD using pre 1998-2006			DD using pre 2005-2006		
Outcome	1998-2006	2005-2006	2008-2009	DD	Rank	p	DD	Rank	p
Panel A: Prime working-age population (15-45)									
Noncitizen Hispanic	0.000	−0.005	−0.027	−0.027	1/47	0.022	−0.022	2/47	0.043
Noncitizen Hispanic, HS or less	0.000	−0.004	−0.022	−0.022	1/47	0.022	−0.018	1/47	0.022
Panel B: Population age 15 and over									
Noncitizen Hispanic	0.000	−0.001	−0.017	−0.017	1/47	0.023	−0.016	1/47	0.023
Noncitizen Hispanic, HS or less	−0.001	0.000	−0.013	−0.013	1/47	0.022	−0.013	1/47	0.022
Panel C: Entire resident population									
Noncitizen Hispanic	0.000	−0.002	−0.015	−0.015	1/47	0.021	−0.013	1/47	0.021

Notes: Average differences between Arizona and its synthetic control in the pre- and post-treatment periods. DD denotes the difference-in-differences estimate. The one-tailed p-value uses the empirical distribution of placebo DD estimates from 46 donor states. States with broad E-Verify mandates (MS, RI, SC, UT) are excluded from the donor pool. Data: Monthly CPS, 1998-2009.

Table 2: Comprehensive Difference-in-Differences Estimates: SCM, SDID, and ASCM across Population Groups and Time Periods

	Mean difference: AZ − Synth. AZ				Original period (post: 2008–2009)						Extended period (post: 2008–2015)					
					DD (pre 1998–2006)			DD (pre 2005–2006)			DD (pre 1998–2006)			DD (pre 2005–2006)		
Outcome	1998–2006	2005–2006	2008–2009	2008–2015	DD	Rank	p	DD	Rank	p	DD	Rank	p	DD	Rank	p
Synthetic Control Method (SCM)																
Panel A: Prime working-age (15–45)																
Noncitizen Hispanic	0.000	−0.005	−0.027	−0.026	−0.027	1/46	0.022	−0.022	2/46	0.043	−0.027	1/46	0.022	−0.022	1/46	0.022
Noncitizen Hispanic, HS or less	0.000	−0.004	−0.022	−0.020	−0.022	1/46	0.022	−0.018	1/46	0.022	−0.020	1/46	0.022	−0.016	1/46	0.022
Panel B: Population 15 and over																
Noncitizen Hispanic	0.000	−0.001	−0.017	−0.015	−0.017	1/45	0.022	−0.017	1/45	0.022	−0.015	1/45	0.022	−0.014	1/45	0.022
Noncitizen Hispanic, HS or less	−0.001	0.000	−0.013	−0.011	−0.013	1/45	0.022	−0.013	1/45	0.022	−0.010	1/45	0.022	−0.010	1/45	0.022
Panel C: All residents																
Noncitizen Hispanic	0.000	−0.002	−0.015	−0.014	−0.015	1/47	0.021	−0.013	1/47	0.021	−0.013	1/47	0.021	−0.012	1/47	0.021
Augmented Synthetic Control (ASCM)																
Panel D: Prime working-age (15–45)																
Noncitizen Hispanic	0.000	−0.005	−0.027	−0.027	−0.027	1/47	0.021	−0.022	2/47	0.043	−0.027	1/47	0.021	−0.022	1/47	0.021
Noncitizen Hispanic, HS or less	0.000	−0.005	−0.021	−0.021	−0.021	1/47	0.021	−0.017	1/47	0.021	−0.021	1/47	0.021	−0.016	1/47	0.021
Panel E: Population 15 and over																
Noncitizen Hispanic	0.000	−0.002	−0.015	−0.012	−0.015	1/47	0.021	−0.013	1/47	0.021	−0.012	1/47	0.021	−0.010	1/47	0.021
Noncitizen Hispanic, HS or less	0.000	−0.001	−0.012	−0.009	−0.012	1/47	0.021	−0.011	1/47	0.021	−0.009	1/47	0.021	−0.008	2/47	0.043
Panel F: All residents																
Noncitizen Hispanic	0.000	−0.001	−0.016	−0.012	−0.016	1/47	0.021	−0.015	1/47	0.021	−0.012	1/47	0.021	−0.011	1/47	0.021
Synthetic Difference-in-Differences (SDID)																
Panel G: Prime working-age (15–45)																
Noncitizen Hispanic	0.000	−0.004	−0.031	−0.042	−0.030	1/47	0.021	−0.026	1/47	0.021	−0.042	1/47	0.021	−0.038	1/47	0.021
Noncitizen Hispanic, HS or less	0.000	−0.003	−0.027	−0.038	−0.027	1/47	0.021	−0.023	1/47	0.021	−0.038	1/47	0.021	−0.034	1/47	0.021
Panel H: Population 15 and over																
Noncitizen Hispanic	0.000	−0.001	−0.020	−0.024	−0.019	1/47	0.021	−0.018	1/47	0.021	−0.024	1/47	0.021	−0.023	1/47	0.021
Noncitizen Hispanic, HS or less	0.000	−0.001	−0.017	−0.021	−0.017	1/47	0.021	−0.016	1/47	0.021	−0.021	1/47	0.021	−0.021	1/47	0.021
Panel I: All residents																
Noncitizen Hispanic	0.000	−0.001	−0.019	−0.024	−0.019	1/47	0.021	−0.018	1/47	0.021	−0.024	1/47	0.021	−0.023	1/47	0.021

Notes: Each cell reports the difference-in-differences estimate computed as the average post-treatment gap (Arizona minus its counterfactual) minus the average pre-treatment gap. Panels A-C use the standard synthetic control method (Abadie, Diamond, and Hainmueller, 2010); Panels D-F use the augmented synthetic control (Ben-Michael, Feller, and Rothstein, 2021); Panels G-I use synthetic difference-in-differences (Arkhangelsky et al., 2021). The counterfactual gap is constructed from each method's fitted weights. Rank and p-values are from permutation (placebo) tests: each donor state is iteratively assigned as treated, the same DD is computed, and Arizona is ranked (most negative = rank 1). One-tailed $p = \text{rank}/N$. States with broad E-Verify mandates (MS, RI, SC, UT) are excluded from the donor pool. 2007 is excluded as a transition year. Data: Monthly CPS, 1998-2015.

ONLINE SUPPLEMENTARY APPENDICES

A Additional Figures and Tables

Table A.1: Trends in the Proportion of Arizona Residents Who Are Hispanic Noncitizens, by Age and Education, 1998-2015

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Population proportions																		
Hisp. noncitizen (all ages)	0.082	0.083	0.085	0.084	0.085	0.085	0.100	0.089	0.092	0.093	0.078	0.066	0.065	0.061	0.061	0.065	0.068	0.077
Hisp. noncitizen, 15+	0.093	0.093	0.096	0.093	0.099	0.099	0.115	0.104	0.106	0.109	0.092	0.080	0.079	0.074	0.075	0.078	0.083	0.094
By education, 15+																		
Less than high school	0.067	0.069	0.065	0.060	0.062	0.062	0.076	0.064	0.064	0.069	0.060	0.047	0.045	0.044	0.049	0.050	0.053	0.058
High school graduate	0.013	0.013	0.019	0.017	0.021	0.024	0.023	0.027	0.028	0.026	0.019	0.020	0.022	0.019	0.015	0.018	0.019	0.024
Some college or more	0.013	0.011	0.012	0.016	0.016	0.014	0.016	0.013	0.014	0.014	0.013	0.012	0.013	0.011	0.010	0.010	0.011	0.012
By education, 15–45																		
Hisp. noncitizen, 15–45	0.121	0.115	0.126	0.125	0.140	0.146	0.171	0.144	0.143	0.148	0.124	0.104	0.102	0.096	0.093	0.100	0.098	0.107
Less than high school	0.084	0.085	0.083	0.079	0.089	0.090	0.111	0.086	0.080	0.089	0.078	0.062	0.054	0.053	0.056	0.061	0.058	0.056
High school graduate	0.018	0.017	0.027	0.024	0.031	0.036	0.037	0.039	0.042	0.039	0.028	0.025	0.032	0.026	0.023	0.026	0.026	0.034
Some college or more	0.019	0.014	0.016	0.022	0.021	0.020	0.023	0.019	0.020	0.020	0.018	0.017	0.016	0.016	0.014	0.014	0.015	0.016

Notes: Tabulated using all monthly Current Population Survey files between 1998 and 2015. Each cell reports the proportion of Arizona's population in the indicated demographic group. Education categories sum to the group total (minor rounding discrepancies possible). Values for 2003-2009 match Bohn, Lofstrom, and Raphael, 2014 Table 1 exactly; values for 2000-2002 differ by 0.005-0.007 due to post-2010-census population weight revisions in later IPUMS CPS extracts. The key patterns, rising shares through 2006, sharp decline in 2008-2009, are identical across extracts.

Table A.2: SCM Donor State Weights by Outcome

	Panel A: 15–45		**Panel B: 15+**		**Panel C: All**
State	Noncit. Hisp.	HS or less	Noncit. Hisp.	HS or less	Noncit. Hisp.
TX	0.519	0.478	0.295	0.088	0.211
CA	0.401	0.392	0.547	0.657	0.494
CO	0.080	0.084	—	0.017	—
FL	—	—	0.002	—	0.278
NV	—	0.046	—	—	—
DE	—	—	0.086	0.096	0.011
ID	—	—	—	—	0.003
NC	—	—	—	0.142	—
AL	—	—	0.071	—	—

Notes: Synthetic control weights assigned to each donor state for five outcome variables. Only states receiving weight > 0.001 are shown; "–" indicates weight < 0.001. Panel A outcomes restrict the population to ages 15-45; Panel B to ages 15 and over; Panel C to the entire resident population. Donor pool: 46 states after excluding AZ, MS, RI, SC, and UT (states with broad E-Verify mandates). Pre-treatment period: 1998-2006. Data: Monthly CPS.

Table A.3: Robustness Checks: Alternative Difference-in-Differences Estimates

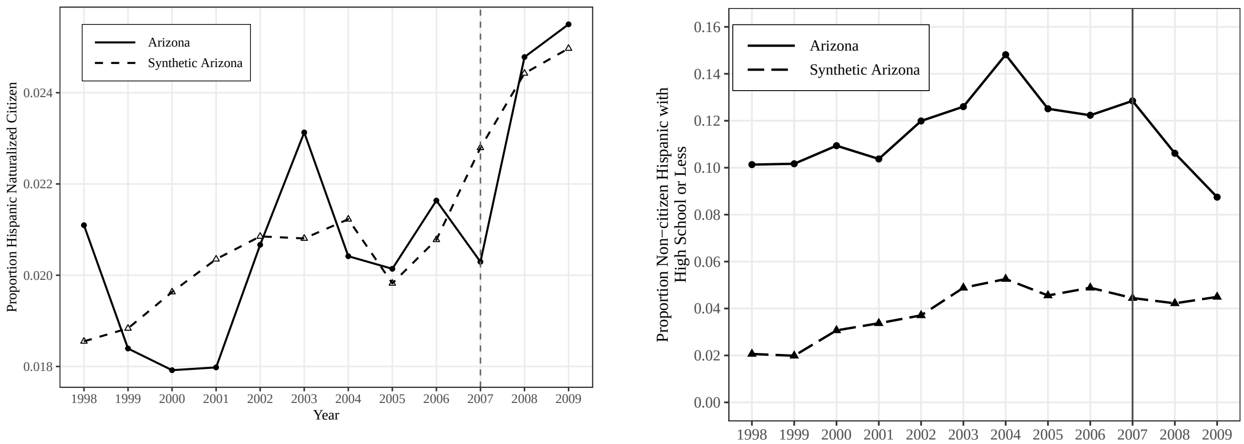
	Avg. Diff.	Avg. Diff.	Avg. Diff.	DiD: Nine Pre-intervention			DiD: 2005/2006 Base		
	Rel. to Comp., Nine Pre-int.	Rel. to Comp., 2005–2006	Rel. to Comp., Post-int.[a]	Change	Rank, Low–High	p-value $P(\Delta < \Delta_{AZ})$	Change	Rank, Low–High	p-value $P(\Delta < \Delta_{AZ})$
Panel A. Including 2007 as a posttreatment year									
Noncitizen Hispanic, high school or less, ages 15–45	0.001	-0.003	-0.014	-0.015	1/47	0.021	-0.011	2/47	0.043
Noncitizen Hispanic, high school or less, age 15 and over	0.000	-0.000	-0.007	-0.007	1/47	0.021	-0.007	2/47	0.043
Noncitizen Hispanic	-0.000	-0.002	-0.010	-0.010	1/46	0.022	-0.008	1/46	0.022
Panel B. Dropping states that border Arizona from the donor pool									
Noncitizen Hispanic, high school or less, ages 15–45	0.012	0.004	-0.016	-0.028	1/43	0.023	-0.020	1/43	0.023
Noncitizen Hispanic, high school or less, age 15 and over	0.004	0.001	-0.016	-0.020	1/43	0.023	-0.017	1/43	0.023
Noncitizen Hispanic	0.008	0.003	-0.014	-0.022	1/43	0.023	-0.017	1/43	0.023

Notes: Average differences pre- and post-intervention are estimates of the difference in the proportion of the Arizona population in a given category relative to the matched synthetic comparison group. The one-tailed test of the significance of the difference-in-differences estimates employs the empirical distribution of the placebo effect estimates of LAWA for 46 additional states. [a]Postintervention period includes 2007 in panel A but excludes 2007 in panel B.

Table A.4: Falsification Test: Naturalized Hispanic Citizens

	1998–2009			1998–2015		
Outcome	SCM	ASCM	SDID	SCM	ASCM	SDID
All ages	0.0004	0.0009	0.0010	0.0027	0.0028	0.0044*
			(0.0019)			(0.0024)
Age 15–45	−0.0004	−0.0003	−0.0016	0.0014	−0.0001	0.0023
			(0.0017)			(0.0027)
Age 15–45, HS or less	0.0010	0.0009	−0.0012	0.0021	0.0025	0.0010
			(0.0013)			(0.0017)

Notes: Estimates comparing Arizona to its synthetic control for naturalized Hispanic citizens, a population not subject to employer verification mandates. SCM reports the difference-in-differences estimate (no standard inference available). ASCM reports the average ATT from augmented synthetic control with ridge bias correction; conformal p-values in brackets. SDID reports the ATT from synthetic difference-in-differences; standard errors in parentheses computed via the placebo method. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The absence of significant effects across all three estimators and both time windows confirms that the main results are specific to noncitizen populations. U.S.-born Hispanics are omitted because their population share mechanically increases when noncitizen Hispanics leave the state, making them unsuitable as a falsification group.

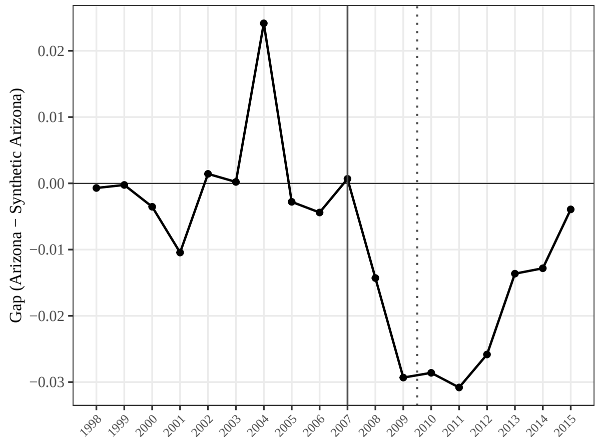


(a) Naturalized Hispanic citizens, 1998-2009

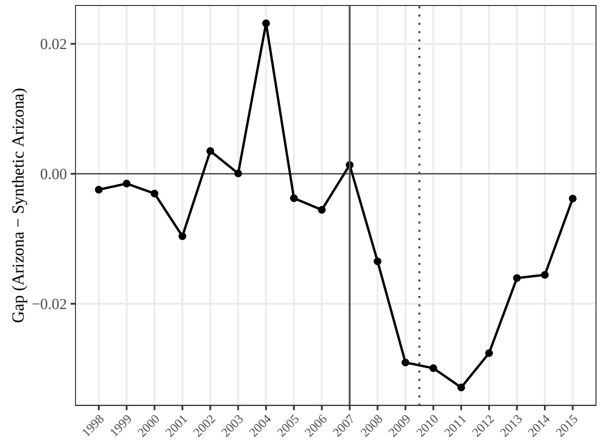
(b) SDID estimate for noncitizen Hispanics 15-45, HS or less

Figure A.1: Placebo and alternative estimator results

Notes: Panel (a) plots Arizona and its synthetic control for the naturalized Hispanic citizen population (placebo outcome). Panel (b) shows the SDID treatment effect estimate with placebo confidence band.



(a) SCM



(b) Augmented SCM

Figure A.2: Gap between Arizona and Synthetic Arizona, 1998-2015

Notes: Each panel plots the difference between Arizona's outcome and its synthetic control (Arizona minus Synthetic) over the extended 1998-2015 period under (a) SCM and (b) ASCM. SDID is omitted because its unit weights do not produce a level-matched counterfactual; the SDID treatment effect is identified through a DiD intercept adjustment rather than the level gap. The solid vertical line marks 2007 (LAWA signed); the dotted vertical line marks 2009, the endpoint of the original Bohn, Lofstrom, and Raphael, 2014 analysis. Outcome: proportion of noncitizen Hispanics with a high school diploma or less among persons aged 15-45.

B Cross-Implementation Robustness

A recurring concern in synthetic control applications is that the same nominal specification can yield different estimates across software implementations (Becker and Klößner, 2017; Klößner et al., 2018). To assess whether our headline result depends on the particular software pipeline, we re-estimate the main outcome (the proportion of noncitizen Hispanics aged 15-45 with a high school diploma or less) using six R implementations and two Python implementations on the exact same data, donor pool, predictors, and pre-treatment window.

The implementations differ in several respects. The original `Synth` package (Abadie, Diamond, and Hainmueller, 2011) jointly optimizes donor and predictor weights, whereas `tidysynth`, `scpi` (Cattaneo, Feng, and Titiunik, 2021), and `pysyncon`'s `Synth` class solve the simplex-constrained donor-weight problem using pre-treatment outcome lags only. `augsynth` (Ben-Michael, Feller, and Rothstein, 2021) adds ridge-based bias correction, while `synthdid` (Arkhangelsky et al., 2021) incorporates time weights. Finally, `gsynth` (Xu, 2017) estimates counterfactuals using interactive fixed effects rather than donor weights.

Figure B.1 visualizes donor-weight allocations across implementations, while Figure B.2 plots the resulting synthetic Arizona series. Despite modest differences in donor weights and optimization routines, all implementations closely track Arizona during the pre-treatment period and produce a similar post-2007 divergence.

Table B.1 reports the average treatment effect on the treated, pre-treatment RMSPE, post/pre RMSPE ratios, and donor weights across implementations. All eight implementations produce negative estimates. The six direct synthetic-control implementations place the average treatment effect in a narrow interval between -0.022 and -0.017, while the methodologically distinct estimators, `synthdid` and `gsynth`, produce somewhat larger negative estimates of -0.029 and -0.032, respectively. The `scpi` package additionally produces a Cattaneo-Feng-Titiunik prediction interval of $[-0.052, -0.007]$, bounded away from zero on the negative side.

The residual variation across implementations is small and largely attributable to specification choices rather than software differences. Implementations using only outcome lags as predictors (`tidysynth`, `scpi`, and `Synth` without covariates) return nearly identical estimates and donor weights, whereas the original `Synth` specification reallocates some weight toward Colorado once the Bohn-Lofstrom-Raphael covariates enter through the predictor-weight matrix. Consistent with Kaul et al. (2022), this suggests that predictor specification rather than solver choice is the main source of implementation-level variation in our setting.

Overall, the cross-implementation exercise shows that the estimated LAWA effect is highly robust to software implementation, optimization routine, and programming language. No implementation produces a sign reversal or economically negligible effect.

Table B.1: Cross-Implementation Robustness for the Main Outcome

Implementation	Lang	ATE	RMSPE_{pre}	Ratio	Runtime (s)	$N_{w>0.01}$	Top donor weights
Synth (R, ADH original)	R	−0.022	0.0090	2.55	14.40	3	TX (0.46), CA (0.43), CO (0.11)
Synth (R, outcome lags only)	R	−0.020	0.0089	2.45	1.75	3	CA (0.50), TX (0.38), NC (0.13)
tidysynth (R)	R	−0.020	0.0089	2.45	1.31	3	CA (0.50), TX (0.38), NC (0.13)
augsynth — vanilla SCM (R)	R	−0.021	0.0089	2.51	0.51	3	CA (0.48), TX (0.41), NC (0.11)
augsynth — Ridge ASCM (R)	R	−0.021	0.0088	2.56	0.02	3	CA (0.48), TX (0.42), NC (0.11)
scpi (R)	R	−0.020	0.0089	2.45	0.88	3	CA (0.50), TX (0.38), NC (0.13)
synthdid (R)	R	−0.029	0.0804	0.67	0.01	10	NH (0.22), KS (0.17), TX (0.14)

Notes: Each row reports the same synthetic control problem solved by a different software package on identical inputs, treated unit Arizona, pre-treatment window 1998-2006, donor pool of 46 states excluding MS, RI, SC, UT. ATE is the mean of the post-period gap between observed and synthetic Arizona, except for `synthdid`, which reports its canonical lambda-time-weighted ATT. $\text{RMSPE}_{\text{pre}}$ is the root mean squared prediction error over 1998-2006; Ratio is post-period RMSPE divided by pre-period RMSPE. $N_{w>0.01}$ is the number of donor states receiving weight above 0.01. "Top donor weights" lists the three largest donors with two-letter state abbreviation. `gsynth` produces an interactive-fixed-effects imputation rather than donor weights and is reported here for completeness.

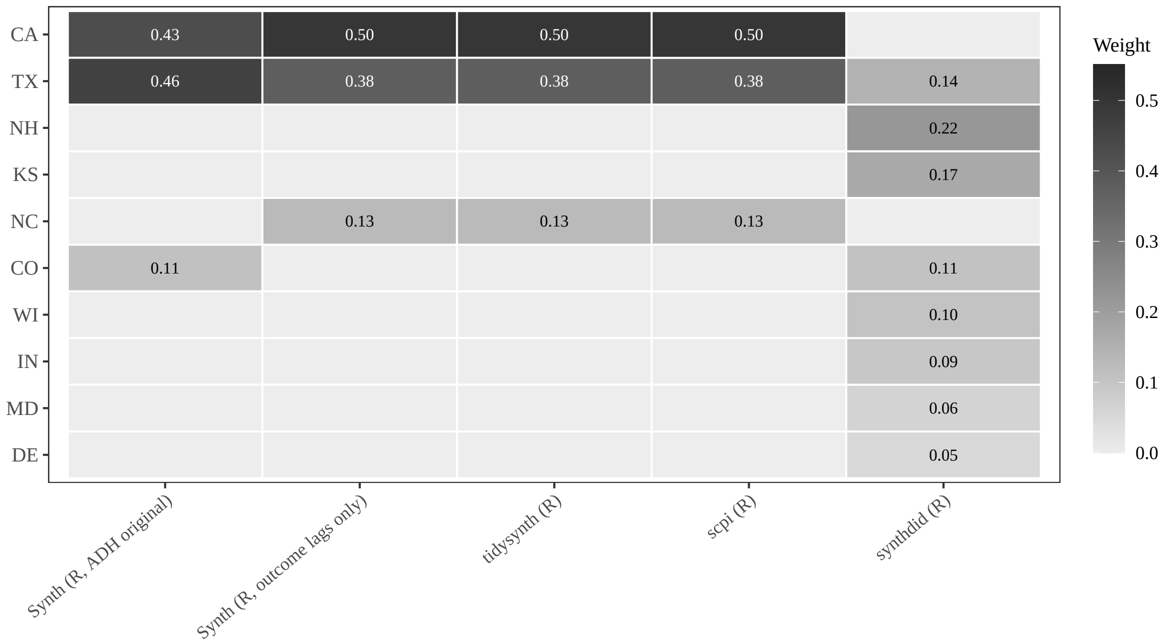


Figure B.1: Donor weights across SCM implementations

Notes: Heat-map of donor weights across implementations. Each cell reports the weight assigned to a donor state by a given package; cells with weight below 0.02 are blank. `tidysynth`, `scpi`, and `pysyncon Synth` produce visually similar weight patterns concentrated on California, Texas, and either North Carolina or Nevada. The original `Synth` (ADH) re-allocates a portion of the California weight to Colorado because the 39 industry-share, education, and unemployment covariates enter through the predictor weight matrix V rather than only through outcome lags. `synthdid` spreads weight more uniformly because it pairs unit weights with time weights. `gsynth` produces no donor weights (interactive fixed-effects imputation).

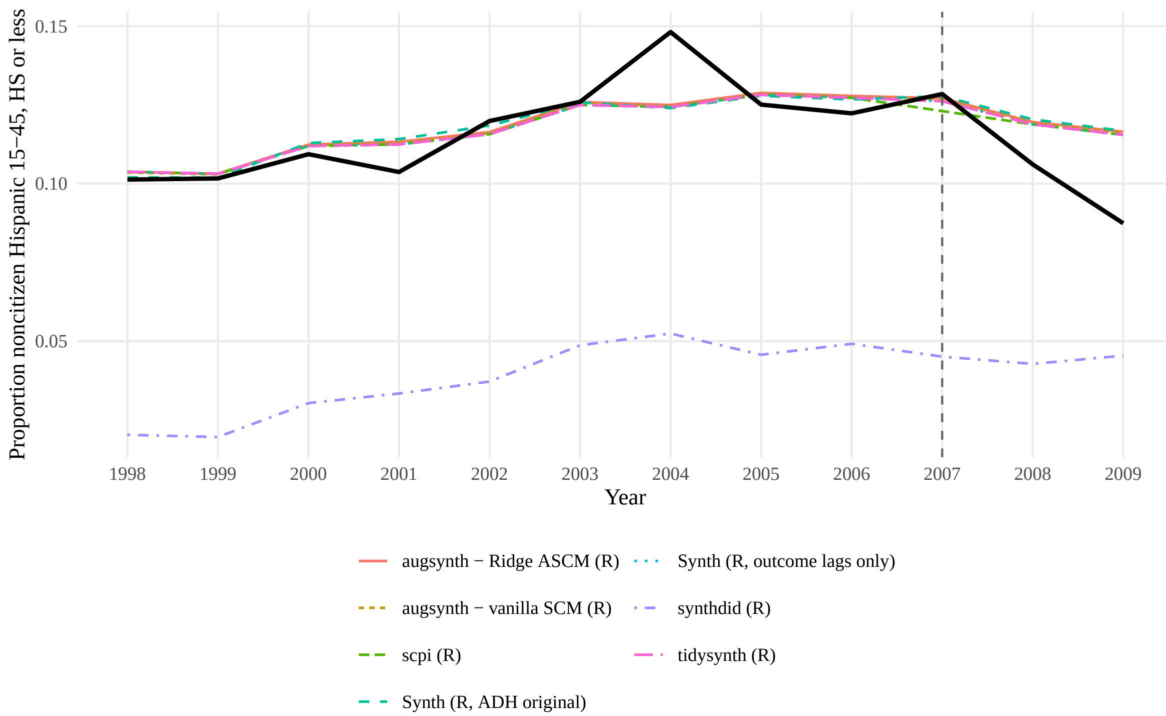


Figure B.2: Synthetic Arizona series across SCM implementations

Notes: Observed Arizona in solid black; the synthetic Arizona constructed by each implementation is plotted with its own color and linetype. The vertical dashed line marks 2007 (LAWA signed into law). All implementations track Arizona's pre-treatment trajectory closely, and all produce a visible post-2007 divergence consistent with a 1.5-2.5 percentage-point decline in the noncitizen-Hispanic share among adults aged 15-45 with a high school diploma or less.

C Extended Estimation Methods

The main text reports treatment effects estimated by the augmented synthetic control method (ASCM) and synthetic difference-in-differences (SDID) alongside the original SCM. This section provides the formal definitions, estimating equations, and key identifying assumptions for each method.

C.1 Augmented Synthetic Control Method

The augmented synthetic control method of Ben-Michael, Feller, and Rothstein (2021) addresses a fundamental limitation of standard SCM: when the convex hull of the donor pool does not contain the treated unit's pre-treatment trajectory, the SCM estimator may be biased. ASCM augments the standard synthetic control weights with a bias-correction term from a ridge-regularized outcome model.

Let Y_{1t} denote the treated unit's outcome and Y_{jt} the outcome for donor unit $j \in \{2, \ldots, J+1\}$. The standard SCM counterfactual is $\hat{Y}_{1t}^{N,\text{scm}} = \sum_{j=2}^{J+1} \hat{w}_j Y_{jt}$, where $\hat{w}_j \geq 0$ and $\sum_j \hat{w}_j = 1$. The ASCM counterfactual augments this with a bias-correction function $\hat{m}(\cdot)$:

$$\hat{Y}_{1t}^{N,\text{aug}} = \sum_{j=2}^{J+1} \hat{w}_j Y_{jt} + \left[\hat{m}(\mathbf{X}_1) - \sum_{j=2}^{J+1} \hat{w}_j \hat{m}(\mathbf{X}_j) \right], \tag{3}$$

where $\mathbf{X}_i = (Y_{i1}, \ldots, Y_{iT_0})'$ is the vector of pre-treatment outcomes for unit i, and $\hat{m}(\cdot)$ is a ridge regression of post-treatment outcomes on pre-treatment outcomes estimated on the donor pool:

$$\hat{m} = \arg\min_{m \in \mathcal{M}} \sum_{j=2}^{J+1} \sum_{t > T_0} (Y_{jt} - m(\mathbf{X}_j))^2 + \lambda \|m\|^2. \tag{4}$$

The ridge penalty $\lambda \geq 0$ is selected via cross-validation and controls the degree of extrapolation beyond the convex hull. When $\lambda \to \infty$, the bias correction vanishes and ASCM reduces to standard SCM; when $\lambda = 0$, the correction is an unconstrained regression adjustment. The treatment effect estimate at time $t > T_0$ is:

$$\hat{\tau}_t^{\text{ascm}} = Y_{1t} - \hat{Y}_{1t}^{N,\text{aug}}. \tag{5}$$

The key assumptions of ASCM are given below.

(i) *Linear factor model.* Potential outcomes under no treatment follow $Y_{it}^N = \boldsymbol{\mu}_t' \boldsymbol{\phi}_i + \varepsilon_{it}$, where $\boldsymbol{\mu}_t$ is a vector of common factors, $\boldsymbol{\phi}_i$ is a vector of unit-specific factor loadings, and ε_{it} is a mean-zero transitory shock. This nests parallel trends as a special case with a single factor μ_t.

(ii) *Approximate balancing.* The SCM weights approximately balance the factor loadings: $\boldsymbol{\phi}_1 \approx \sum_j \hat{w}_j \boldsymbol{\phi}_j$. When this balance is imperfect, the ridge correction absorbs the residual imbalance.

(iii) *Bounded extrapolation.* The ridge penalty bounds how far the augmented estimator can extrapolate beyond the convex hull, trading off bias and variance. The cross-validated λ selects the penalty that minimizes out-of-sample prediction error in the pre-treatment period.

C.2 Synthetic Difference-in-Differences

The synthetic difference-in-differences estimator of Arkhangelsky et al. (2021) combines the reweighting logic of SCM with the double-differencing structure of standard DID. Where SCM matches levels and DID assumes parallel trends, SDID uses unit weights to improve pre-treatment fit *and* time weights to identify the most informative pre-treatment periods for constructing the counterfactual trend.

Define unit weights $\hat{\omega}_i \geq 0$ for control units i (with $\sum_i \hat{\omega}_i = 1$ and $\hat{\omega}_i = 1/N_{\text{tr}}$ for treated units) and time weights $\hat{\lambda}_t \geq 0$ for pre-treatment periods (with $\sum_{t \leq T_0} \hat{\lambda}_t = 1$ and $\hat{\lambda}_t = 1/(T - T_0)$ for post-treatment periods). The SDID estimator solves:

$$(\hat{\tau}^{\text{sdid}},\, \hat{\mu},\, \hat{\alpha}_i,\, \hat{\beta}_t) = \underset{\tau,\, \mu,\, \alpha_i,\, \beta_t}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\omega}_i \hat{\lambda}_t \big(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau\big)^2, \tag{6}$$

where $W_{it} = 1$ if unit i is treated and $t > T_0$, and zero otherwise. The unit weights $\hat{\omega}$ are chosen to balance pre-treatment outcomes across treated and control groups:

$$\hat{\omega} = \underset{\omega \geq 0,\, \sum \omega_i = 1}{\arg\min} \sum_{t=1}^{T_0} \left(\sum_{i:\, W_i = 0} \omega_i Y_{it} - \frac{1}{N_{\text{tr}}} \sum_{i:\, W_i = 1} Y_{it} \right)^2 + \zeta^2 T_0 \|\omega\|_2^2, \tag{7}$$

and the time weights $\hat{\lambda}$ are chosen to balance pre- and post-treatment periods for the control units:

$$\hat{\lambda} = \underset{\lambda \geq 0,\, \sum \lambda_t = 1}{\arg\min} \sum_{i:\, W_i = 0} \left(\sum_{t=1}^{T_0} \lambda_t Y_{it} - \frac{1}{T - T_0} \sum_{t > T_0} Y_{it} \right)^2 + \zeta^2 (T - T_0) \|\lambda\|_2^2. \tag{8}$$

The regularization parameter ζ prevents the weights from concentrating on too few units or periods. The time weights $\hat{\lambda}_t$ typically concentrate on the pre-treatment periods closest to the treatment date, which are most predictive of the post-treatment counterfactual.

The key assumptions of SDID are given below.

(i) *Latent factor model.* As in ASCM, potential outcomes follow $Y_{it}^N = \boldsymbol{\mu}_t' \boldsymbol{\phi}_i + \varepsilon_{it}$. This allows for heterogeneous trends across units, which the unit weights correct for.

(ii) *Parallel trends after reweighting.* The reweighted control group must satisfy parallel trends with the treated unit. The unit weights achieve this by matching the pre-treatment trajectory, while the time weights identify the pre-treatment periods most informative about the counterfactual trend.

(iii) *No anticipation.* Treatment has no effect before $T_0 + 1$. In our application, we treat 2007 as a transition year and exclude it from both the pre- and post-treatment windows, which accommodates anticipatory behavioral responses to LAWA's July 2007 signing.

SDID nests both methods as special cases. When the time weights are uniform ($\hat{\lambda}_t = 1/T_0$ for all t), SDID reduces to a unit-weighted DID estimator. When the unit intercepts α_i are dropped from Equation 6, SDID reduces to SCM. The combination of unit-specific intercepts (the DID element) with outcome-balancing unit weights (the SCM element) makes SDID more tolerant of imperfect pre-treatment fit than SCM and more robust to violations of strict parallel trends than standard DID.

D Additional Robustness

D.1 LAWA vs. SB 1070 Period Split

Table D.1: Period Split: LAWA Only (2008-2010) vs. Post-SB 1070 (2011-2015)

	LAWA only (2008–2010)			Post-SB1070 (2011–2015)		
Outcome	SCM	SDID	ASCM	SCM	SDID	ASCM
Noncit. Hisp. 15–45, HS or less	-0.0244	-0.0204	-0.0241	-0.0101	0.0025	-0.0050
Noncit. Hisp. 15–45	-0.0298	-0.0241	-0.0295	-0.0126	0.0013	0.0034
Noncit. Hisp. (all ages)	-0.0161	-0.0148	-0.0170	-0.0085	-0.0015	-0.0023

Notes: Each cell reports the estimated treatment effect (DD for SCM, ATT for SDID and ASCM). The left panel restricts the post-treatment window to 2008-2010, before SB 1070 takes effect. The right panel estimates the marginal effect of SB 1070 over 2011-2015, using 1998-2010 as the pre-treatment period. Pre-treatment period for LAWA: 1998-2006. Donor pool: 46 states (excluding MS, RI, SC, UT).

D.2 Leave-One-Out Donor Robustness

To assess whether the synthetic control estimate depends on any single donor state, we iteratively remove each of the three highest-weighted donors and re-estimate the SCM treatment effect for the main outcome (noncitizen Hispanics aged 15-45, high school diploma or less). Table D.2 reports the results.

The baseline DD estimate of -0.022 is stable across all configurations. The largest perturbation occurs when dropping the highest-weighted donor, which shifts the DD to -0.018, still a meaningful negative effect. Dropping all three top donors simultaneously yields -0.018, with the synthetic control relying entirely on the remaining 43 states. Figure D.1 plots Arizona against each leave-one-out synthetic control.

Table D.2: Leave-One-Out Donor Robustness

Donor pool	DD	Top weights
Baseline (46 states)	-0.0222	TX 0.48, CA 0.39, CO 0.08
Drop TX (45 states)	-0.0176	CA 0.67, NC 0.25, NV 0.08
Drop CA (45 states)	-0.0236	TX 0.55, NV 0.45
Drop CO (45 states)	-0.0233	TX 0.65, CA 0.32, NV 0.03
Drop top 3 (TX+CA+CO) (43 states)	-0.0181	NV 1.00

Notes: DD is the difference-in-differences estimate (mean post-treatment gap minus mean pre-treatment gap) for the main outcome: proportion of noncitizen Hispanics aged 15-45 with a high school diploma or less. Pre-treatment: 1998-2006. Post-treatment: 2008-2009. Top weights are the three largest donor weights in each configuration.

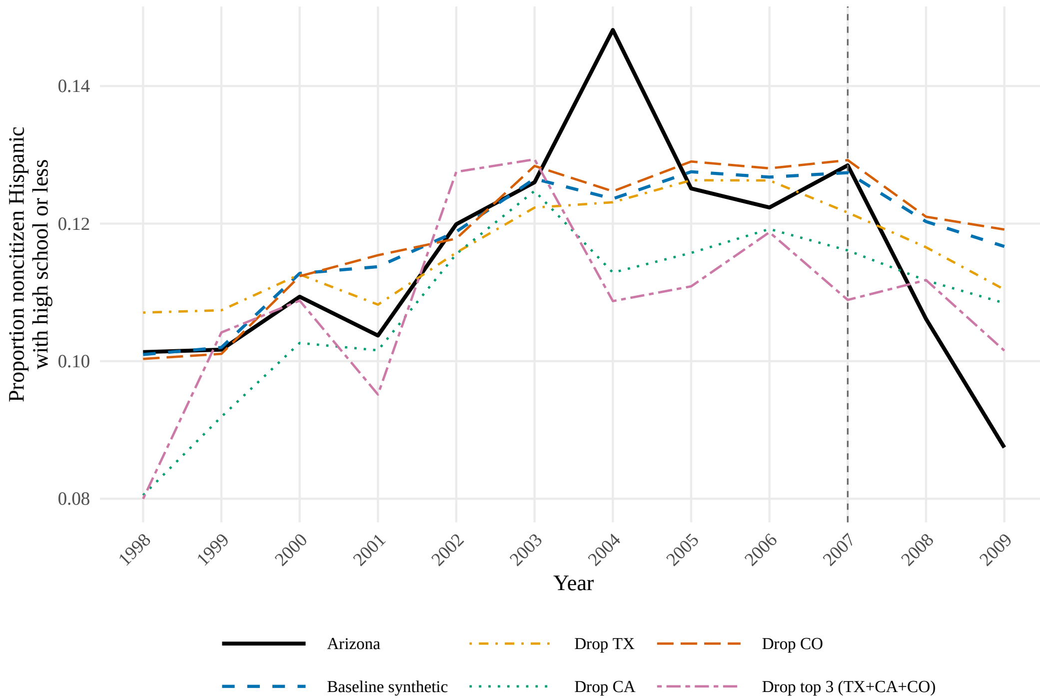


Figure D.1: Leave-one-out donor robustness: Arizona vs. synthetic controls

Notes: Each line shows a synthetic Arizona constructed after removing one or more top-weighted donors from the donor pool. The solid black line is actual Arizona. The vertical dashed line marks 2007 (LAWA signed). Outcome: proportion of noncitizen Hispanics with a high school diploma or less among persons aged 15-45.

D.3 SDID Time Weights and Pre-Trend Diagnostics

The SDID estimator produces attenuated point estimates relative to SCM (-0.012 vs. -0.022 for the main outcome). This section examines whether the attenuation reflects a pre-existing trend or is a mechanical consequence of SDID's time-weighting structure.

Figure D.2(a) plots the pre-treatment SCM gaps (Arizona minus synthetic Arizona) for 1998-2006 with a fitted linear trend. The slope is 0.0007 (SE $= 0.0013$, $p = 0.59$), positive, not negative, and statistically indistinguishable from zero. There is no evidence of a pre-LAWA downward trend that could bias the post-treatment estimates.

Figure D.2(b) displays the SDID time weights $\hat{\lambda}_t$ for each pre-treatment year. The weights concentrate heavily on 2005-2006, which together receive 74% of the total weight (0.32 and 0.42 respectively). Because Arizona's gap was already slightly negative in 2005-2006 (-0.003 and -0.004), the SDID baseline incorporates this small pre-existing deviation, mechanically shrinking the estimated post-treatment effect. The attenuation is a feature of SDID's design, it aggressively down-weights early pre-treatment years, not evidence of a confounding pre-trend.

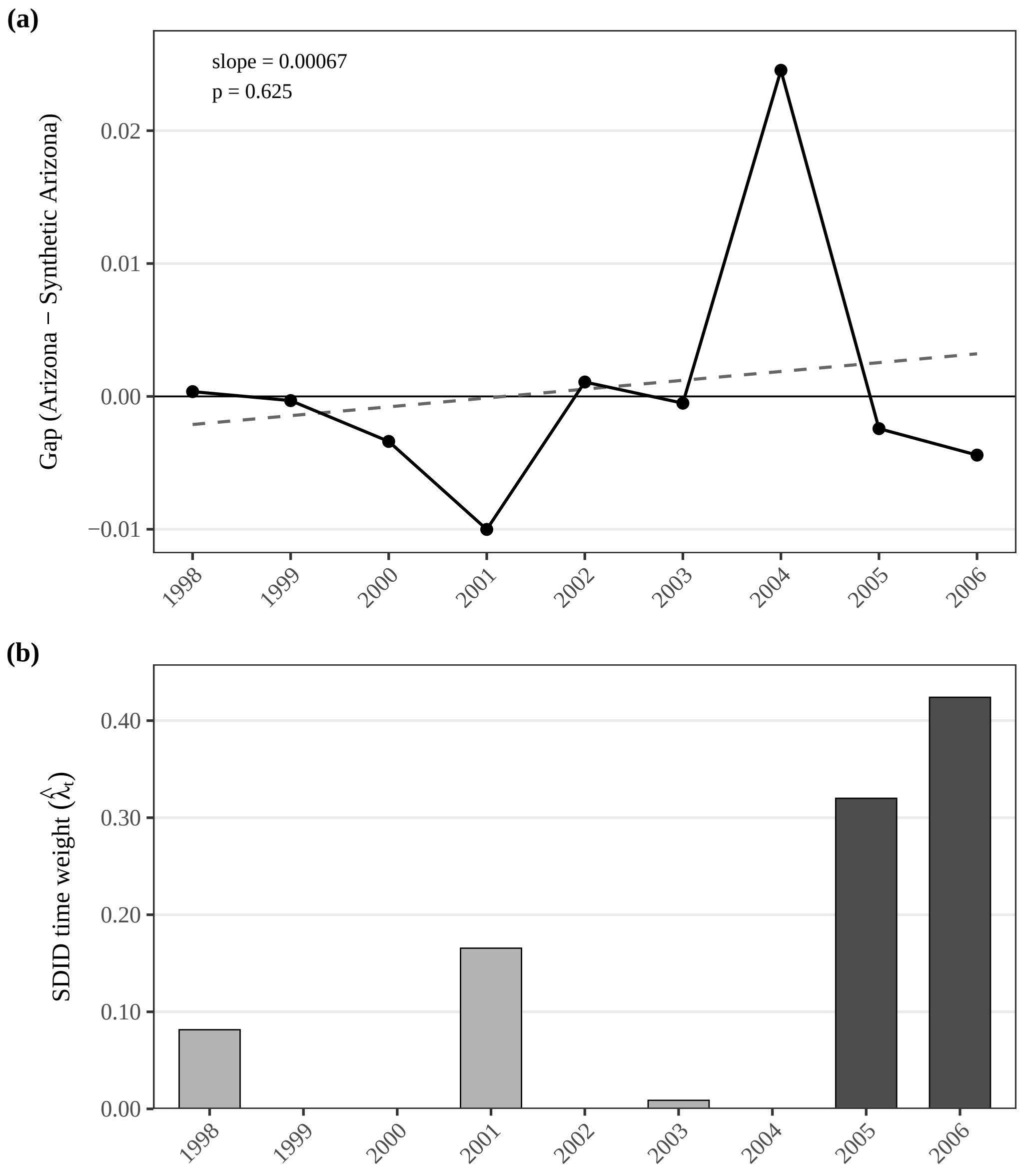


Figure D.2: Pre-trend diagnostics and SDID time weights

Notes: Panel (a) plots the SCM gap (Arizona minus synthetic Arizona) for each pre-treatment year with a linear trend line. The trend slope is 0.0007 ($p = 0.59$). Panel (b) shows the SDID time weights $\hat{\lambda}_t$ for each pre-treatment year. The darker bars represent 2005-2006, which together receive 74% of the total weight.

E Decomposing the Population Effect: Outflow and Inflow

The CPS-based synthetic control analysis cannot distinguish whether the decline in Arizona's noncitizen Hispanic population share reflects out-migration from Arizona or reduced inflows into Arizona. To examine these channels directly, we use American Community Survey (ACS) bilateral migration data and estimate gravity difference-in-differences models for the same population studied in the main analysis: working-age Hispanic noncitizens with a high school diploma or less.

Using ACS microdata from 2005-2010 and the `migplac1` variable, which records state of residence one year earlier, we construct bilateral interstate migration flows by origin state, destination state, and ACS year. Because ACS year t measures migration during calendar year $t-1$, ACS 2008 captures migration during 2007, the LAWA-signing year. We estimate Poisson Pseudo-Maximum Likelihood (PPML; Santos Silva and Tenreyro (2006)) gravity difference-in-differences models of the form

$$\text{flow}_{ijt} = \exp(\beta\, T_{ijt} + \alpha_{ij} + \gamma_{i,t} + \delta_{j,t}) \cdot \varepsilon_{ijt}, \tag{9}$$

where α_{ij} denotes origin-destination fixed effects, $\gamma_{i,t}$ origin-by-year fixed effects, and $\delta_{j,t}$ destination-by-year fixed effects. Standard errors are clustered at the origin-destination pair level.

For outflows, we define the treatment indicator as $T^{\text{out}}_{ijt} = \mathbf{1}\{i = \text{AZ}\} \times \mathbf{1}\{j \in \mathcal{N}\} \times \mathbf{1}\{t \geq t^*\}$, where $\mathcal{N} = \{\text{CA, NV, NM, UT}\}$ denotes Arizona's neighboring states. For inflows, we define $T^{\text{in}}_{ijt} = \mathbf{1}\{j = \text{AZ}\} \times \mathbf{1}\{t \geq t^*\}$. We report specifications using both $t^* = 2008$, which captures responses beginning in the signing year, and $t^* = 2009$, which isolates post-implementation effects.

Table E.1 reports outflow estimates. The baseline gravity PPML specification with $t^* = 2008$ yields $\hat{\beta}^{\text{out}} = 0.661$ ($p = 0.010$), implying that AZ→neighbor migration flows increased by approximately 94 % following LAWA. The estimates remain positive across alternative specifications and timing definitions, though less precisely estimated when only post-implementation years are used. Figure E.1(a) shows flat pre-treatment coefficients and an immediate increase in outflows beginning in 2008, consistent with anticipatory migration responses following LAWA's signing.

Table E.2 reports inflow estimates. The implementation-only specification ($t^* = 2009$) yields $\hat{\beta}^{\text{in}} = -0.448$ ($p = 0.061$), implying a 36 % decline in inflows to Arizona. Restricting origins to neighboring states produces a larger estimate of -57.8 % ($p < 0.001$). Figure E.1(b) shows little response during the announcement period but a sharp decline beginning in 2009, suggesting that inflow deterrence responded primarily to implementation rather than signing.

The decomposition reveals what the CPS-based main result cannot. Outflow displacement to neighboring states is a meaningful component of the population effect, suggesting that part of LAWA's local impact reflected geographic redistribution rather than an aggregate reduction in unauthorized residence, consistent with the broader sub-federal enforcement literature (Hoekstra and Orozco-Aleman, 2017; Amuedo-Dorantes and Lozano, 2015). Inflow deterrence is the second component, and the estimated declines in inflows (-36 % to -58 %) are economically meaningful and consistent with prior evidence that immigration enforcement policies can deter prospective migrants from relocating to affected destinations (Hoekstra and Orozco-Aleman, 2017; Bohn, Lofstrom, and Raphael, 2014). One limitation is that ACS `migplac1` captures interstate migration only and therefore cannot measure international migration responses from Mexico or other countries. The estimates should therefore be interpreted as identifying interstate redistribution and deterrence channels within the United States.

Table E.1: ACS Bilateral Flows: Gravity DiD on Outflow from Arizona

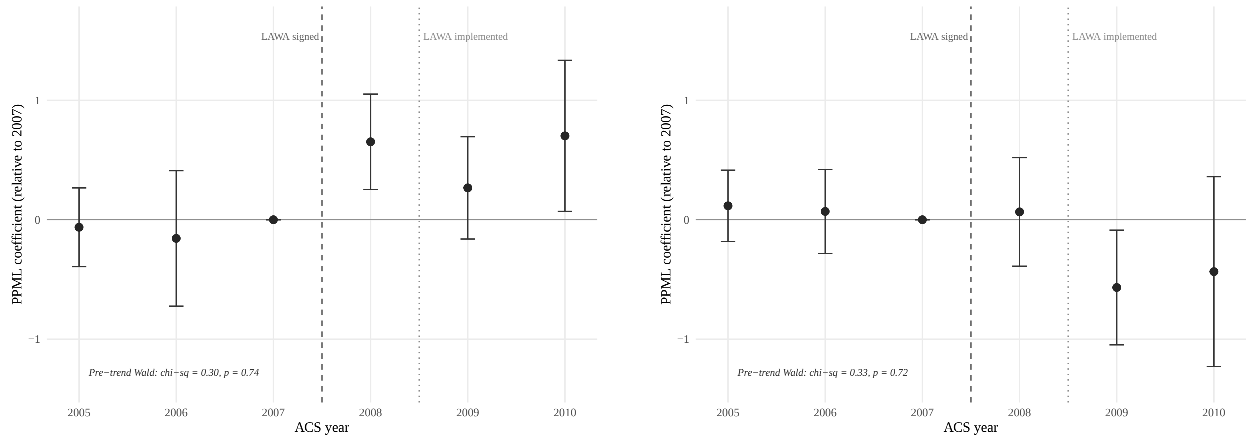
	PPML gravity FE post $\geq$ 2008 (1) headline	PPML gravity FE post $\geq$ 2009 (2) impl. only	log(1+flow) gravity FE post $\geq$ 2008 (3) OLS check	PPML pair+year FE post $\geq$ 2008 (4) less saturated
AZ $\rightarrow$ Neighbors	0.661***	0.316	0.778*	0.629***
	(0.255)	(0.367)	(0.413)	(0.174)
% change in flow	93.7%	37.1%	117.6%	87.5%
Pair FE	Yes	Yes	Yes	Yes
Origin $\times$ year FE	Yes	Yes	Yes	No
Destination $\times$ year FE	Yes	Yes	Yes	No
Year FE	–	–	–	Yes
Cluster (pair)	Yes	Yes	Yes	Yes
Pre-trend Wald p-value (simple FE)			0.740	
Pre-trend Wald p-value (gravity FE)			0.438	
Observations	5,478	5,478	15,000	5,682

Notes: Bilateral migration flows from origin state i to destination state j in ACS year t, restricted to working-age (16-64) Hispanic noncitizens with high-school diploma or less. Inflow rows (destination = AZ) are dropped so the regression isolates outflow from Arizona. Treatment indicator is $T_{ijt}^{\text{out}} = \mathbf{1}\{i = \text{AZ}\} \cdot \mathbf{1}\{j \in \{\text{CA, NV, NM, UT}\}\} \cdot \mathbf{1}\{t \geq t^{*}\}$. Columns (1)-(3) saturate with pair, origin-by-year, and destination-by-year fixed effects (gravity DiD). Column (4) replaces origin-by-year and destination-by-year with a common year fixed effect (less saturated). PPML estimated by `fixest::fepois`; column (3) is OLS on $\log(1 + \text{flow})$ for functional-form robustness. Standard errors clustered at the pair level in parentheses. Pre-trend Wald p-values are joint tests that the 2005 and 2006 event-study coefficients on AZ$\rightarrow$neighbor pairs equal zero (reference year 2007). Stars: $^{*}\, p < 0.10$, $^{**}\, p < 0.05$, $^{***}\, p < 0.01$.

Table E.2: ACS Bilateral Flows: PPML on Inflow to Arizona

	PPML broad inflow post $\geq$ 2008 (1) headline	PPML broad inflow post $\geq$ 2009 (2) impl. only	PPML gravity mirror post $\geq$ 2008 (3) neigh. orig.
Inflow to AZ (any origin)	-0.169	-0.448*	—
	(0.180)	(0.239)	
Neighbor origin $\rightarrow$ AZ	—	—	-0.862***
			(0.252)
% change in flow	-15.6%	-36.1%	-57.8%
Pair FE	Yes	Yes	Yes
Origin $\times$ year FE	Yes	Yes	Yes
Destination $\times$ year FE	No	No	Yes
Cluster (pair)	Yes	Yes	Yes
Pre-trend Wald p-value (simple FE)		0.716	
Pre-trend Wald p-value (gravity FE)		0.740	
Observations	5,528	5,528	5,437

Notes: Bilateral migration flows for the same demographic group as Table E.1, restricted to flows into Arizona by dropping rows where origin = AZ. Columns (1)-(2) report the broad inflow specification: $T_{ijt}^{\text{in}} = \mathbf{1}\{j = \text{AZ}\} \cdot \mathbf{1}\{t \geq t^{*}\}$, identifying the change in inflow to AZ from any origin under pair and origin-by-year fixed effects (destination-by-year FE cannot be included because they are perfectly collinear with the treatment). Column (3) reports the gravity mirror $T_{ijt}^{\text{in,N}} = \mathbf{1}\{i \in \{\text{CA, NV, NM, UT}\}\} \cdot \mathbf{1}\{j = \text{AZ}\} \cdot \mathbf{1}\{t \geq t^{*}\}$, identifying the change in flows from neighbor states to AZ specifically, which permits the full gravity FE structure. PPML estimated by `fixest::fepois`, standard errors clustered at the pair level. Pre-trend Wald p-values are joint tests that 2005 and 2006 event-study coefficients on AZ-destination pairs equal zero (reference year 2007). Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



(a) Outflow: AZ → neighbor states

(b) Inflow: any origin → AZ

Figure E.1: Event studies: AZ outflow to neighbors and inflow to AZ from any origin

Notes: Year-by-year PPML coefficients estimated under pair and year fixed effects with standard errors clustered at the pair level. Each coefficient is the log relative flow in that ACS year compared to 2007 (the omitted reference, capturing calendar 2006 flows). Vertical lines mark LAWA's signing date (July 2, 2007; dashed) and effective date (January 1, 2008; dotted). Panel (a): outflow coefficients on AZ-origin × neighbor-destination interactions, pre-trend Wald $\chi^2 = 0.30$, $p = 0.74$. Panel (b): inflow coefficients on AZ-destination interactions across all origins, pre-trend Wald $\chi^2 = 0.33$, $p = 0.72$. The dynamic patterns differ: outflow jumps in 2008 ACS (calendar 2007, the LAWA-signing window), whereas inflow drops in 2009 ACS (calendar 2008, the first full implementation year), consistent with existing residents responding to the announcement and prospective migrants responding to enforcement onset.