Fixed-Effect Sensitivity in FMAP-Kink Regression Designs: A Public-Data Methods Note on Medicaid LTSS Spending at the 50% Statutory Floor

Abstract

Background. The Federal Medical Assistance Percentage (FMAP) formula creates a regression kink at the 50 percent federal match floor that a growing literature uses as a regression kink design (RKD) for state Medicaid spending responses to federal generosity. Modern panels span the post-2014 era in which the ACA Medicaid expansion introduced a 90 percent enhanced match for newly eligible adults in expansion states only - a second margin of federal generosity whose adoption is endogenous to state characteristics correlated with the FMAP-kink running variable.

Methods. I rebuild the FMAP-kink RKD on a public state-year panel for fiscal years 2008-2023 (N = 816) from MACPAC MACStats, HHS Federal Register, BEA SAINC1, and CMS/Mathematica Medicaid LTSS reports. I estimate an approximate local-polynomial RKD with triangular kernel, exact two-way (state + fiscal-year) fixed-effect residualization via alternating projections, and conventional cluster-robust standard errors at the state level, pre-specifying enhanced-FMAP episode drops (ARRA; ACA expansion; FFCRA section 6008) as robustness.

Results. On the full panel the FMAP-kink point estimates are positive but imprecise: log total Medicaid per capita tau = 1.49 (95% CI [-1.06, 4.03]); log HCBS per capita tau = 1.19 [-1.30, 3.68]; HCBS share tau = -0.19 [-1.28, 0.90]. Every primary-outcome confidence interval contains zero by a wide margin. Under exact two-way fixed-effect residualization, dropping ACA-expansion state-years moves the aggregate spending coefficient to 0.11 [-0.78, 1.01] and the HCBS-share coefficient to -0.24 [-0.89, 0.42] - also indistinguishable from zero. An earlier version of this analysis reported a negative-and-significant drop-ACA pattern; that result was an implementation artifact of sequential one-pass demeaning on an unbalanced sub-sample and is withdrawn. The empirical FMAP slope change (0.88-1.04) recovers the mechanical 0.9487 implied by the section 1905(b) formula, confirming the public-data pipeline is correctly parsed.

Conclusions. On public state-year data covering FY2008-FY2023, the FMAP-kink RKD does not produce a precise estimate of either the aggregate spending response or the HCBS-rebalancing response at the kink, in either the baseline or any episode-drop specification. The chapter is therefore a methods note rather than a substantive finding, with two practical recommendations: implement state-and-fiscal-year fixed effects via exact OLS-on-dummies (or alternating projections, or a vetted FE package) rather than sequential one-pass demeaning on unbalanced sub-samples; and treat public-data FMAP-kink confidence intervals

on FY2008-FY2023 as too wide to support a substantive interpretation.

---

1. Introduction

The federal-state matching structure of Medicaid creates state-level variation in the marginal price of a Medicaid dollar that an applied public-finance literature has begun to exploit. Three published applications use the regression kink at the 50 percent statutory FMAP floor as a quasi-experimental source of variation in federal generosity. Kumar [-@kumar2020medicaid] uses the kink to estimate the effect of federal Medicaid generosity on household income. Giertz and Kumar [-@giertzkumar2024multiplier] use it to estimate a local fiscal multiplier of federal-to-state grants. Leung [-@leung2022state], the closest substantive precedent for this paper, applies the kink directly to state Medicaid spending and estimates a FMAP elasticity of 1.6–2.9, reporting in passing that “match rates appear to stimulate spending in hospitals and long-term supports and services.” The FMAP-kink RKD is increasingly a workhorse design for questions about how state Medicaid spending and downstream outcomes respond to federal financing.

This paper makes a small but practical methods point about the FMAP-kink RKD on public state-year data. I implement an approximate local-polynomial RKD at the 50 percent statutory floor on federal fiscal years 2008–2023, with exact two-way (state + fiscal-year) fixed-effect residualization via alternating projections, triangular kernel, and conventional cluster-robust standard errors at the state level. On the full FY2008–FY2023 panel, the headline point estimates are positive but imprecise: aggregate log Medicaid per capita tau = 1.49 (95 percent CI [-1.06, 4.03]); log HCBS per capita tau = 1.19 [-1.30, 3.68]; HCBS share tau = -0.19 [-1.28, 0.90]. Every primary-outcome confidence interval contains zero by a wide margin. Pre-specified enhanced-FMAP episode drops — including the conventional drop-ACA-expansion robustness check — do not change that conclusion: under exact two-way residualization the drop-ACA log-spending coefficient is 0.11 [-0.78, 1.01] and the drop-ACA HCBS-share coefficient is -0.24 [-0.89, 0.42].

The paper’s contribution is twofold. *First*, I show that the public-data implementation of the FMAP-kink RKD on FY2008–FY2023 is too imprecise to identify either an aggregate FMAP-spending response or an HCBS-rebalancing response at the kink, in either the baseline or any episode-drop specification. *Second*, I document that the FMAP-kink RKD on the conventional drop-ACA-expansion sub-sample is highly sensitive to *how* state-and-fiscal-year fixed effects are residualized. Sequential one-pass demeaning — subtract state means, then subtract fiscal-year means — is exact only on balanced panels. On the unbalanced drop-ACA sub-sample it is not equivalent to OLS residuals from explicit state and fiscal-year dummies, and the gap is large enough to flip a small effect’s sign. A unit-tested case on a synthetic unbalanced panel records a maximum absolute deviation of approximately 0.22 between the two residualization approaches; on the FMAP drop-ACA sub-sample, sequential demeaning produced a log-total-spending coefficient of -2.53 (95 percent CI [-4.26, -0.79]) while exact two-way residualization on the same sub-sample, same bandwidth, and same data produces 0.11 [-0.78, 1.01]. The earlier sign-flip pattern was an implementation artifact, not a behavioral signal, and is withdrawn. Future FMAP-kink applications on unbalanced sub-samples should implement state-and-fiscal-year fixed effects via OLS-on-dummies residualization, alternating projections to convergence, or a vetted fixed-effects regression package.

The mechanism worth stating up front is institutional. The ACA Medicaid expansion's 90 percent enhanced match for newly eligible adults is not a function of the statutory FMAP formula. It is a flat statutory rate applied on top of the statutory FMAP for the expansion population in expansion states only. Whether a state expanded is a function of state politics and budget structure, and those factors correlate with the running variable: expansion states are concentrated among non-floor states; non-expansion states are concentrated in the Deep South and among mid-income states below — but not at — the floor. Dropping ACA-expansion state-years from the panel therefore creates a structurally unbalanced sub-sample, and structurally unbalanced sub-samples are precisely where the difference between sequential one-pass demeaning and OLS-on-dummies residualization is non-trivial. The methods point of this paper is not that the ACA expansion "contaminates" the FMAP kink in a substantive sense — the confidence intervals on this public-data design are too wide to support either that claim or its negation — it is that the conventional drop-ACA robustness check is implemented incorrectly when sequential demeaning is used on its structurally unbalanced sub-sample.

The paper carries two secondary contributions worth surfacing in the introduction. *First*, the HCBS-versus-institutional decomposition I originally designed the paper around — the next-order question Leung explicitly noted but did not pursue — returns confidence intervals too wide to distinguish small-effect from proportional-expansion explanations on every outcome. I report this null honestly and discuss two competing interpretations; I do not claim to have measured the absence of a federally financed Medicaid LTSS rebalancing effect, but I also do not claim to have detected one. *Second*, the aggregate FMAP-spending response at the kink (FMAP elasticity approximately 0.74; 95 percent CI [-0.53, 2.02]) is directionally consistent with Leung's published 1.6–2.9 range — her lower bound sits inside my upper tail — but too imprecise to adjudicate. Whether her published estimate is itself sensitive to the same residualization implementation I document is an open empirical question I am not in a position to answer without re-running her specification on her data.

A prerequisite contribution is the data pipeline itself. Because the statutory FMAP is a deterministic function of the running variable, recovering the formula-implied slope change of 0.9487 from the empirical data (I get 0.88–1.04 across bandwidths) is evidence that MACPAC Exhibit 6, the HHS Federal Register notices, and the BEA SAINC1 series are being parsed consistently against the

§1905(b) statutory formula. This is a necessary data-integrity check, not a behavioral first stage, and I present it that way in §5.1.

The remainder of the paper proceeds as follows. Section 2 provides background on the FMAP formula, the ACA expansion's 90 percent enhanced match, and the LTSS rebalancing literature. Section 3 describes the data, including the Mathematica vintage handling that bears on the decomposition null. Section 4 presents the empirical strategy. Section 5 reports results in five subsections: a data-pipeline integrity check (§5.1), the aggregate FMAP-kink response on public data (§5.2), the HCBS-versus-institutional decomposition (§5.3), the enhanced-FMAP episode drops under exact two-way residualization with the fixed-effect-implementation sensitivity that motivates the methods note (§5.4), and a consolidated sensitivity-analysis subsection (§5.5). Section 6 discusses what the public-data design does and does not show (§6.1), the residualization-implementation lesson (§6.2), two practical recommendations for future FMAP-kink work (§6.3), why the HCBS decomposition does not show up (§6.4), the comparison to Leung (§6.5), and limitations (§6.6). Section 7 concludes.

---

2. Background

2.1 The FMAP formula and its 50 percent floor

The federal government pays a state-specific share of each state's Medicaid benefit expenditures. That share is the Federal Medical Assistance Percentage (FMAP), and it is computed each year by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) at the U.S. Department of Health and Human Services using a formula fixed in §1905(b) of the Social Security Act [@mitchell2025fmap]:

$$\text{FMAP}_{s,t} = 1 - 0.45 \cdot \left(\frac{\overline{\text{PCPI}}_{s,\,t-4:t-2}}{\overline{\text{PCPI}}_{US,\,t-4:t-2}} \right)^2,$$

bounded below at 0.50 and above at 0.83, with a three-year lag in per-capita personal income (PCPI) so that the FMAP for fiscal year t is computed using calendar years $t-4$ through $t-2$. The formula is deterministic: given BEA's state-level PCPI vintage, the FMAP for the coming federal fiscal year is mechanical up to the floor and ceiling. Because FMAP is a deterministic function of the running variable, the "first stage" in any RKD application is a formula identity rather than a behavioral relationship — recovering the formula-implied slope change on a real data build is evidence that the data pipeline has been parsed correctly, not evidence that state behavior responds at the kink.

The statutory floor generates a regression kink in FMAP as a function of the state-to-national PCPI ratio. Defining $r_{s,t} \equiv \overline{\text{PCPI}}_{s,\,t-4:t-2} / \overline{\text{PCPI}}_{US,\,t-4:t-2}$, the formula has slope $-0.9 \cdot r$ for $r < r^* \equiv \sqrt{0.5/0.45} \approx 1.0541$ and slope 0 for $r \geq r^*$. At the kink, the slope change is $0.9 \cdot r^* \approx 0.9487$. States on the floor (the richest states by lagged PCPI ratio) receive exactly 50 percent federal match; states below the floor receive a higher match that responds mechanically to PCPI movements. The ten to fourteen states at the floor in any given year include California, Connecticut, Maryland, Massachusetts, Minnesota, New Hampshire, New Jersey, New York, Virginia, and Washington, with a rotating set of marginal entrants and exiters (Alaska, Colorado, Illinois, North Dakota, Rhode Island, Wyoming) [@mitchell2025fmap; @macpac2024macstatsexhibit6].

The kink at $r^* = 1.0541$ is the identifying feature the RKD exploits. Because the running variable is an aggregate income statistic with a three-year lag, state governments cannot plausibly manipulate r at the margin that matters for $r < r^*$ versus $r \geq r^*$. This is the standard identification argument from Card, Lee, Pei, and Weber [-@cardleepieweber2015econometrica; -@cardleepieweber2017rktheorypractice]: under smooth conditional densities of unobservables in the running variable, the slope change in the outcome at the kink divided by the slope change in the treatment variable identifies a weighted average of the local marginal effect of the treatment on the outcome. For inference I follow the now-standard nonparametric approach of Calonico, Cattaneo, and Titiunik [-@calonicocattaneotitiunik2014econometrica], using local polynomial estimation with a triangular kernel, MSE-optimal bandwidth selection, and robust bias-corrected confidence intervals.

2.2 Three enhanced-FMAP episodes overlay the statutory schedule

In the modern panel, three federally financed enhanced-match episodes overlay the statutory FMAP schedule. None alters the *location* of the 50 percent floor in the formula, but each introduces a second margin of federal generosity that potentially contaminates the spending response a researcher would attribute to the statutory FMAP.

The American Recovery and Reinvestment Act §5001 bump (FY2009 Q4 – FY2011 Q2) provided a temporary across-the-board FMAP increase of 6.2 percentage points plus an unemployment-keyed top-up. The Affordable Care Act Medicaid expansion (effective in expansion states from January 2014 onward) provides a separate 90 percent enhanced federal match for newly eligible adults — distinct from, and applied on top of, the statutory FMAP for the traditional Medicaid population. The Families First Coronavirus Response Act §6008 bump (FY2020 Q1 through FY2023 Q1, with phase-down through December 2023 under CAA-2023) provided a 6.2-percentage-point across-the-board FMAP increase conditional on continuous enrollment [@clemensetal2021covidfederalism; @dagueukert2024pandemic; @giertzkumar2024multiplier]. The ARRA and FFCRA episodes are temporary and apply uniformly to all states; the ACA expansion is permanent, applies only in expansion states, and is the methodologically consequential episode for the FMAP-kink design because expansion adoption is itself an endogenous state choice that selects on the same characteristics that correlate with the running variable.

2.3 Medicaid LTSS rebalancing and federal financing

Medicaid is the single largest payer for long-term services and supports (LTSS) in the United States, financing roughly 42 percent of national LTSS spending in recent years [@reaves2015primer]. The composition of that spending has shifted dramatically over four decades. Through the 1980s and early 1990s, institutional care in nursing facilities dominated Medicaid LTSS dollars and users. A series of federal policy actions — the 1915(c) waiver authority, the *Olmstead v. L.C.* Supreme Court decision, Money Follows the Person demonstrations, the Balancing Incentive Program, the §1915(i) state plan HCBS option, and the §1915(k) Community First Choice option created by ACA §2401 — expanded state authority to deliver LTSS in home and community settings. By FY2022, MACPAC reports that 65 percent of Medicaid LTSS dollars flow to home- and community-based services (HCBS) rather than institutional settings [@macpac2023hcbs; @murray2021ltssexpenditures].

Whether HCBS expansion substitutes for institutional care has been the central empirical question, and the literature's answer is "partially." Kaye, LaPlante, and Harrington [-@kayelaplanteharrington2009] document that states expanding HCBS experience a short-term spending bump followed by long-term cost reduction. Guo, Konetzka, and Manning [-@guokonetzkamanning2015] identify a causal substitution effect via claims-based instrumental variables and find that a $1,000 increase in Medicaid home care spending is associated with 2.75 fewer nursing facility days and a $351 reduction in nursing facility costs per older Medicaid enrollee. McGarry and Grabowski [-@mcgarrygrabowski2023woodwork] estimate that each additional dollar of Medicaid HCBS spending is offset by approximately $0.26 of institutional Medicaid LTSS spending — partial substitution, but far from dollar-for-dollar. Konetzka [-@konetzka2014hidden] cautions that the cost-effectiveness case for HCBS is not robust to severity progression, cognitive impairment, and unmet need. Lieber and Lockwood [-@lieberlockwood2019targeting] use randomized variation in a Medicaid HCBS program to show that in-kind home care is meaningfully less valued than its cash-equivalent cost but that targeting benefits to sicker recipients dominate in welfare terms.

This literature identifies rebalancing from cross-state policy variation, beneficiary-level claims-based instruments, and randomized targeting experiments within a single program. None identifies rebalancing from the federal-financing margin. Whether the mechanical variation in federal match at the FMAP floor translates into a measurable shift in the HCBS-institutional composition of state Medicaid LTSS spending was the original prospective question of this paper. As I show in §5.3, that question is not identified at the precision available on public state-year data — the design returns null estimates with confidence intervals wide enough to contain both the McGarry-Grabowski elasticities and a precise zero. The HCBS-versus-institutional decomposition is therefore reported here as secondary evidence that the aggregate FMAP-kink response does not decompose cleanly within LTSS. The headline contribution of this paper is the methodological warning about ACA-expansion contamination, which I develop below.

---

3. Data

3.1 Unit of observation and panel window

The analytic file is a state-year panel indexed by state and federal fiscal year. I cover the 50 states plus the District of Columbia for federal fiscal years 2008 through 2023 (the baseline analysis window), with FY2024 retained as an edge-window observation for placebo checks only — FY2024 is not used in the headline 816 state-year baseline analysis. I exclude the five U.S. territories because their statutory FMAP is fixed at 55 percent by §1905(b) and provides no information on the formula-based running variable. Consistent with Leung [-@leung2022state] and Kumar [-@kumar2020medicaid], I retain D.C. in the baseline panel and flag it so that sensitivity results without D.C. can be reported alongside the main estimates.

Federal fiscal year N runs from October 1, $N-1$ through September 30, N. Where a variable is reported on a calendar year (notably Census resident population and BEA per-capita personal income), I attach calendar year Y to fiscal year $Y+1$, matching the convention used in prior FMAP-kink papers. The baseline panel contains $51 \times 16 = 816$ state-years.

3.2 Data sources

My panel draws on six public sources.

Statutory FMAP — MACPAC MACStats Exhibit 6 and HHS Federal Register notices. Statutory FMAP values come from MACPAC's MACStats *Exhibit 6* archive (editions March 2011 through December 2024) [@macpac2024macstatsexhibit6], cross-validated against the authoritative HHS Federal Register notices on "Federal Financial Participation in State Assistance Expenditures" (eight fiscal years fetched directly from govinfo.gov). Across 357 overlapping state-year observations, mean $|\text{FR} - \text{MACStats}| = 0.003$ percentage points, with the only systematic disagreement on Louisiana FY2011 and FY2013 (Katrina/Rita disaster-recovery adjustments). After reconciliation, statutory FMAP is populated in every one of the 816 state-years in the FY2008–FY2023 baseline window.

State personal income — Bureau of Economic Analysis SAINC1. The state and national per-capita personal income series come from BEA's Regional Economic Accounts table SAINC1. For each fiscal year t, I compute the state PCPI three-year average over calendar years $t-4$ through $t-2$

and divide by the analogous national average, matching the HHS lag structure [@mitchell2025fmap]. The centered running variable is $\tilde{r}_{s,t} = r_{s,t} - r^*$ with $r^* \approx 1.0541$.

Medicaid LTSS expenditures — Mathematica/CMS annual series. HCBS and institutional LTSS spending totals come from the *Medicaid LTSS Annual Expenditures Reports*, originally produced by Eiken and colleagues and currently produced by Mathematica under contract to CMS [@murray2021ltssexpenditures]. These reports assemble state-by-state totals for 1915(c) waivers, 1915(i) state plan HCBS, 1915(j), 1915(k) Community First Choice, state plan personal care and home health, PACE, Money Follows the Person, and 1115 HCBS lines, together with nursing facility, ICF/IID, and mental-health-facility institutional LTSS lines. I harmonize these into an HCBS bucket (sum of HCBS lines) and an institutional bucket (sum of institutional lines).

TAF-based LTSS users briefs (FY2019–FY2023). I use the CMS "Medicaid LTSS Users and Expenditures by Category" briefs for an HCBS share of LTSS users outcome over FY2019–FY2023 as an appendix cross-check. I do not treat the TAF user share as a co-equal headline outcome because the 5-year window (plus one suppressed AL FY2021 cell) is too short to support an RKD over a fixed-state floor-binding structure on its own; the TAF row appears in Appendix Table A1 rather than in the main results table of this paper.

State population — U.S. Census Bureau. State total population is the NST-EST annual vintage; the age-65+ subpopulation is from SC-EST. Census population serves as the denominator for per-capita outcomes.

KFF state health facts. I use KFF's state-level Medicaid enrollment indicators and state-by-state ACA expansion effective dates (KFF 2025) to construct the `aca_expansion_flag` variable. The KFF series and the hard-coded Clemens-Ippolito-Veuger [-@clemensetal2021covidfederalism] fallback agree on all 51 states at the month level.

3.2.1 Mathematica vintage handling

The Medicaid LTSS Annual Expenditures Reports series was originally produced by Steve Eiken and colleagues at Truven/IBM Watson Health (covering through approximately FY2016) and is currently produced by Murray, Tourtellotte, Lipson, and Wysocki at Mathematica (FY2017 onward). Post-2017 Mathematica vintages back-cast a harmonized HCBS subcategory allocation that conflicts with contemporaneous pre-2017 vintages; the harmonization affects how 1915(c) waiver, 1915(i) state plan HCBS, and 1915(k) Community First Choice expenditures are allocated within and across the HCBS bucket. My baseline panel uses the **latest-available harmonized Mathematica vintage** for every state-year. The **FY2015 and FY2016 vintages are available only as PDF tabulations** and could not be machine-parsed without introducing classification errors; these two state-years are dropped from every LTSS outcome (leaving 713 state-years with LTSS outcomes out of 816 in the FMAP panel).

This vintage choice is load-bearing in one specific way that the reader should know about up front: extending the panel back to FY2002 (the Option B specification check reported in Appendix E and in §5.4 below) collapses the HCBS spending point estimate from 1.18 to 0.01 — consistent with either a true time-heterogeneity in the HCBS response or with pre-2017 Mathematica vintage harmonization drift. I cannot distinguish these two explanations from this design, and the HCBS decomposition results in §5.3 should be read as conditional on the post-2017 harmonized vintage being correctly back-cast. The vintage-collapse finding is also the subject of Limitation 3 in §6.6.

3.3 Running variable construction

I construct the running variable strictly from the statutory formula applied to raw BEA PCPI data. I do **not** incorporate the enhanced-FMAP episodes into the running-variable construction. For each state-year I compute $r_{s,t}$ from the three-year lagged state/US PCPI ratio, the centered running variable $\tilde{r}_{s,t} = r_{s,t} - r^*$, and the baseline statutory FMAP as the BEA-formula output bounded at 0.50. I override D.C. to its statutory 70 percent value per §1905(b), and I retain the Louisiana Katrina/Rita disaster-recovery adjustments in the MACStats panel as documented in MACPAC's Table 14 footnotes (the resulting Louisiana observations are within the MSE-optimal bandwidth in some specifications and are handled through the leave-one-state-out diagnostic rather than dropped mechanically; Louisiana is not among the leave-one-out states whose removal moves the headline coefficient by more than half a standard error).

3.4 Enhanced-FMAP episode handling

The running variable is the **statutory** FMAP, and each of the three enhanced-match episodes is handled as a **robustness drop** rather than a baseline modification. I construct three indicator flags:

- `arra_flag`: 1 for state-years in which the ARRA §5001 base or unemployment-keyed bump applied (FY2009 Q4 through FY2011 Q2; conservatively flagged as all of FY2009, FY2010, and FY2011 at the fiscal-year frequency).
- `aca_expansion_flag`: 1 for state-years in which the state was operating under the ACA Medicaid expansion's 90 percent enhanced match for newly eligible adults (from the state's effective expansion date, fiscal-year rounded).
- `ffcra_flag`: 1 for state-years in which the FFCRA §6008 6.2-percentage-point bump applied (FY2020 through FY2023, conservatively flagged).

None of the three alters the *location* of the statutory 50 percent floor in the running variable, so the kink in my estimator is mechanically unchanged by their presence. Each can, however, in principle contaminate the *spending response* I measure at the kink by introducing a second federal-generosity margin that covaries with state characteristics. Under exact two-way state-and-fiscal-year fixed-effect residualization on the FY2008–FY2023 panel, all three episode-drop columns return confidence intervals that contain zero (§5.4). The drop-ACA column is the one that motivates this paper's methods note: it is the only one whose underlying sub-sample is structurally unbalanced, and the structurally unbalanced sub-sample is precisely where the residualization-implementation choice (sequential one-pass demeaning vs. OLS-on-dummies / alternating projections) materially affects the estimate. I treat this at length in §5.4 and §6.2.

3.5 Outcome construction

My four primary outcomes plus one appendix cross-check are:

1. **log(total Medicaid $ / pop)** — the Leung-linked aggregate outcome. CMS-64 total Medicaid benefit spending divided by state total population, logged. This is the outcome most directly comparable to Leung's 1.6–2.9 elasticity.
2. **log(HCBS $ / pop)** — sum of 1915(c), (i), (j), (k), state plan personal care, state plan home health, PACE, MFP, and 1115 HCBS lines, divided by state population, logged.
3. **log(institutional LTSS $ / pop)** — sum of nursing facility, ICF/IID, and mental-health-facility LTSS lines, divided by state population, logged.
4. **HCBS share of LTSS $** — HCBS / (HCBS + institutional), in levels.
5. **HCBS share of LTSS users (TAF FY2019–2023)** — appendix cross-check only; reported in Appendix Table A1, not in the main Results table.

All monetary outcomes are in nominal dollars; I did not deflate, on the grounds that year fixed effects in the RKD absorb the common-trend component of price-level movements and that state-level deflation would introduce its own measurement concerns. Robustness specifications with pop-65+ as the denominator, with CFC excluded from the HCBS bucket, and with alternative polynomial orders are reported in Appendix Tables 4 and 8.

3.6 Sample restrictions

The baseline sample is 51 states $\times$ 16 fiscal years = 816 state-years. LTSS outcomes are available for 713 state-years (FY2015 and FY2016 LTSS observations are dropped per §3.2.1). The TAF HCBS user share (appendix cross-check) is available for FY2019–FY2023 only.

3.7 Summary statistics

Table 1 (in `analysis/tables/table_01_summary_stats.csv`, markdown mirror `table_01_summary_stats.md`) reports N, mean, standard deviation, minimum, and maximum for every variable in the analytic file, broken out for the full baseline sample and for the near-floor (statutory FMAP 50–53 percent) and far-from-floor (statutory FMAP > 53 percent) subsamples. Statutory FMAP has mean 59.8 percent (SD 8.2) across the 816 state-years; the near-floor mean is

50.3 percent and the far-from-floor mean is 64.6 percent. The running variable $r_{s,t}$ has mean 0.984 (SD 0.161), with the near-floor subsample averaging 1.134 (above the kink at 1.054) and the far-from-floor subsample averaging 0.909 (below the kink). The panel contains 272 near-floor state-years out of 816 — the small but non-negligible effective sample that drives statistical precision in the reduced form.

---

4. Empirical Strategy

4.1 Sharp RKD specification and polynomial-order choice

I estimate a sharp regression kink design at the 50 percent statutory FMAP floor, following Card, Lee, Pei, and Weber [-@cardleepieweber2015econometrica; -@cardleepieweber2017rktheorypractice] and the nonparametric inference framework of Calonico, Cattaneo, and Titiunik [-@calonicocattaneotitiunik2014econometrica]. Let $Y_{s,t}$ denote the outcome of interest, $\tilde{r}_{s,t}$ the centered running variable (state/US PCPI ratio minus r^*), and $\text{FMAP}_{s,t}$ the statutory FMAP. Within a bandwidth h around the kink, I estimate the local polynomial regressions

$$Y_{s,t} = \alpha_Y + \beta_Y^- \tilde{r}_{s,t} \cdot \mathbf{1}[\tilde{r}_{s,t} < 0] + \beta_Y^+ \tilde{r}_{s,t} \cdot \mathbf{1}[\tilde{r}_{s,t} \geq 0] + \gamma_Y \tilde{r}_{s,t}^2 + \delta_Y X_{s,t} + \varepsilon_{s,t}^Y$$

$$\text{FMAP}_{s,t} = \alpha_F + \beta_F^- \tilde{r}_{s,t} \cdot \mathbf{1}[\tilde{r}_{s,t} < 0] + \beta_F^+ \tilde{r}_{s,t} \cdot \mathbf{1}[\tilde{r}_{s,t} \geq 0] + \gamma_F \tilde{r}_{s,t}^2 + \delta_F X_{s,t} + \varepsilon_{s,t}^F$$

on a triangular-kernel-weighted sample within $|\tilde{r}_{s,t}| \leq h$. The sharp RKD estimand is

$$\tau_{\text{RKD}} = \frac{\beta_Y^+ - \beta_Y^-}{\beta_F^+ - \beta_F^-}.$$

I include state and fiscal-year fixed effects in $X_{s,t}$ throughout. Standard errors are clustered by state. The bandwidth is MSE-optimal per Calonico, Cattaneo, and Titiunik [-@calonicocattaneotitiunik2014econometrica], implemented via a local-polynomial analog of their `rdrobust` procedure.

I use **local-quadratic** (p=2) as the baseline polynomial order. Calonico, Cattaneo, and Titiunik recommend a local polynomial of order p=p+1 for bias correction in RKD applications: the point estimate uses p=2 and the bias-corrected robust confidence interval uses a p=3 fit on the same bandwidth. My headline tables follow this convention; Appendix Table 4 reports both pure p=2 and pure p=3 specifications at half, baseline, and double the MSE-optimal bandwidth. The pure-p=3 half-bandwidth specification produces an obvious outlier (institutional LTSS tau = −14.7) from an underidentified local cubic on a small sample, which is consistent with CCT's warning that local-cubic point estimates can be unstable at narrow bandwidths in sparse panels. A referee who prefers p=3 as the baseline would see this outlier as a warning, not a defense of p=2; my position is that p=2 with bias-corrected CIs is the standard CCT recommendation for slope-change RKDs in panels of this size.

4.2 The first-stage formula identity (data-integrity check)

The mechanical first-stage slope change at the kink is $\beta_F^+ - \beta_F^- = 0.9 \cdot r^* \approx 0.9487$ per unit of the centered running variable, where the left-of-kink slope is $-0.9 \cdot r^*$ and the right-of-kink slope is 0 (the floor binds). Because the statutory FMAP is a deterministic function of the running variable, the empirical first-stage slope change is a formula identity, not a behavioral relationship. I expect to recover the mechanical value within small deviations attributable to D.C.'s 70 percent statutory override and to the Louisiana Katrina/Rita adjustments, and §5.1 confirms that I do. Recovering the formula-implied kink is evidence that the data pipeline (MACPAC Exhibit 6, HHS Federal Register notices, and BEA SAINC1) is parsed correctly against the §1905(b) statutory formula. It is *not* evidence that the FMAP-kink design "works" in any behavioral sense, and I present it in §5.1 as a prerequisite data-integrity check rather than as a behavioral first-stage replication.

4.3 Units and the translation to Leung's elasticity convention

My reduced-form coefficient τ_{RKD} on a log-outcome regression is a **semi-elasticity with respect to FMAP expressed as a fraction** because FMAP is defined on $[0, 1]$ in the formula. To translate my estimates into the elasticity convention Leung [-@leung2022state] uses, I multiply the semi-elasticity by the baseline FMAP value at the kink (0.50). Thus a semi-elasticity of 1.49 on log(total Medicaid/pop) translates to a FMAP elasticity of $1.49 \times 0.50 \approx 0.74$ at the kink point. Leung reports a 1.6–2.9 FMAP elasticity; my point estimate is approximately half her lower bound, and my 95 percent confidence interval of [−0.53, 2.02] for the elasticity overlaps the lower tail of her published range.

4.4 Clustering and inference

I cluster standard errors by state (CR1) throughout. With 45–51 clusters on either side of the kink, the normal-approximation clustered SE should be approximately correct. A wild-cluster bootstrap backup is retained in the replication code but not reported as a headline.

4.5 Pre-specified robustness checks

Five robustness exercises feed into Section 5: (i) bandwidth sensitivity at half, baseline, and double MSE-optimal at $p = 2$ and $p = 3$ (Appendix Table 4); (ii) leave-one-state-out (Appendix Table 6); (iii) enhanced-FMAP episode drops (Table 5 in main text — also the locus of the residualization-implementation lesson); (iv) placebo kinks at ± 0.02, ± 0.05 in centered running-variable units (Appendix Table 7); and (v) a robustness menu for the HCBS-share outcome including drop-D.C., local cubic, drop-CFC, pop-65+ denominator, and a 0.005-wide donut (Appendix Table 8). At double the MSE-optimal bandwidth, the total-Medicaid tau attenuates to 0.70 [−0.26, 1.65] — a tighter CI but a smaller point estimate, which I discuss in §5.5 as a sensitivity pattern that cuts in the opposite direction from a "precision is structural" narrative.

---

5. Results

5.1 Data pipeline integrity check: formula reconstruction of the FMAP first stage

Table 2 and Figure 1 report the empirical slope change in statutory FMAP at the 50 percent floor. Across three bandwidth choices — half, baseline, and double the MSE-optimal window — the empirical slope change is:

Specification	Bandwidth	Left-of-kink slope	Right-of-kink slope	Empirical slope change	Mechanical
Half MSE-opt	0.115	−0.989	0.049	1.038	0.9487
MSE-optimal	0.230	−0.889	−0.007	0.883	0.9487
Double MSE-opt	0.459	−0.982	−0.083	0.898	0.9487

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

The empirical slope change is 0.88–1.04, bracketing the mechanical 0.9487 implied by differentiating the FMAP formula at $r^* = 1.054$. **Because the statutory FMAP is a deterministic function of the running variable, this recovery speaks to data pipeline integrity — it confirms that MACPAC Exhibit 6, the HHS Federal Register notices, and BEA SAINC1 are being parsed consistently against the 0.45-times-squared-ratio statutory formula — and is a necessary prerequisite rather than a behavioral validation.** The small residual discrepancy traces to D.C.'s statutory 70 percent override (which never responds to the formula) and to Louisiana FY2011 and FY2013 Katrina/Rita adjustments [@macpac2024macstatsexhibit6]. The slope recovery is consistent with the data-reconciliation check, in which mean |FR − MACStats| was 0.003 percentage points across 357 overlapping state-years. I present this result first not as the paper's headline but as a precondition: I am not proposing an FMAP-kink response measured on data that is wrong, and the formula-identity reconstruction shows that I am not.

5.2 Aggregate FMAP-kink response on public data (context for the headline finding)

Table 3 row 1 and the top-left panel of Figure 3 report the aggregate FMAP-kink response on the FY2008–FY2023 public-data panel before any episode drop. On log total Medicaid spending per capita, the RKD point estimate at the MSE-optimal bandwidth is $\tau = 1.488$ (SE 1.299), with a 95 percent clustered confidence interval of $[-1.058, 4.034]$. The effective sample is 652 state-years and 45 state clusters.

Translating this semi-elasticity to Leung's elasticity convention at the kink (multiply by FMAP = 0.50) gives a point estimate of approximately **0.74** with a 95 percent confidence interval of $[-0.53, 2.02]$. Leung's published 1.6–2.9 FMAP elasticity sits in the upper half of my confidence interval: her lower bound (1.6) is below my upper bound (2.02), but her point estimate is roughly two to four times mine, and my confidence interval crosses zero. I characterize this as **directionally consistent in the limited sense that her published lower bound (1.6) is below my upper bound (2.02), with my point estimate approximately half her lower bound and my 95 percent interval crossing zero.** It is too imprecise to adjudicate Leung's headline. I return to the comparison in §6.5.

5.3 The HCBS-versus-institutional decomposition (secondary evidence)

Table 3 rows 2–4 and the remaining panels of Figure 3 report the HCBS-versus-institutional decomposition. This was the original prospective contribution of the paper relative to Leung [-@leung2022state]; it returns a null on every outcome.

Outcome	tau (semi-elasticity)	SE	95% CI	N
log(HCBS $ / pop)	1.183	1.273	[−1.312, 3.677]	623
log(institutional $ / pop)	1.831	1.891	[−1.876, 5.538]	623
HCBS share of LTSS $	−0.194	0.556	[−1.284, 0.897]	623

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

None of the three outcomes is distinguishable from zero at the 5 percent level. The HCBS share of LTSS dollars has a small negative point estimate (−0.19) with a standard error of 0.56. I report point estimates and confidence intervals only and do not interpret the directional pattern across them as a finding: the standard errors are far too wide to support any sign-based claim about HCBS-versus-institutional substitution. The HCBS user share from CMS TAF is reported in Appendix Table A1 as a short-window cross-check and is not treated as a co-equal main outcome of this paper, because the FY2019–FY2023 window is too short to support an RKD over a fixed-state floor-binding structure on its own. The interpretation of why this decomposition does not show up — and what it would mean if a sharper future design recovered an effect on the order of 1 percentage point — is in §6.4.

5.4 Enhanced-FMAP episode drops under exact two-way FE residualization

Table 5 reports the enhanced-FMAP episode drops under exact two-way (state + fiscal-year) fixed-effect residualization (alternating projections to convergence).

Specification	log(total Medicaid / pop)	log(HCBS / pop)	log(institutional / pop)	HCBS share	N
Baseline (FY08-FY23)	1.488 [-1.06, 4.03]	1.190 [-1.30, 3.68]	1.829 [-1.88, 5.54]	-0.191 [-1.28, 0.90]	652
Drop ARRA (FY09-FY11)	0.827 [-1.51, 3.16]	0.598 [-1.89, 3.09]	1.661 [-2.84, 6.16]	-0.317 [-1.49, 0.85]	576
Drop ACA expansion state-years	0.114 [-0.78, 1.01]	0.375 [-0.97, 1.72]	1.493 [-0.68, 3.67]	-0.236 [-0.89, 0.42]	417
Drop FFCRA (FY20-FY23)	1.259 [0.04, 2.48]	0.371 [-1.77, 2.51]	1.821 [-2.16, 5.80]	-0.303 [-1.31, 0.70]	481
Drop all three episodes	-0.244 [-1.42, 0.93]	0.089 [-1.82, 2.00]	3.019 [-1.13, 7.17]	-0.663 [-1.86, 0.53]	240

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

None of the four primary outcomes is statistically distinguishable from zero in any episode-drop column. The drop-ACA point estimates on log total Medicaid per capita (0.11), log HCBS per capita (0.38), and HCBS share (-0.24) are all within roughly one standard error of zero. The drop-all-three row tells the same story.

Sequential demeaning produces materially different drop-ACA estimates on this sub-sample. An earlier version of this chapter implemented the "state and fiscal-year fixed effects" specification as sequential one-pass demeaning — subtract the state mean from each observation, then subtract the fiscal-year mean from the residual. On the balanced FY2008–FY2023 panel that procedure is numerically equivalent to OLS residuals from state and fiscal-year dummies, so the *baseline* numbers in the first row of the table above are unchanged. On the structurally unbalanced drop-ACA sub-sample the two procedures are not equivalent. Under sequential demeaning, the earlier chapter reported drop-ACA estimates of -2.53 [-4.26, -0.79] on log total Medicaid per capita and -1.19 [-2.04, -0.34] on HCBS share — both negative and significant. Under exact two-way residualization on the same sub-sample, same bandwidth, and same data, those collapse to 0.11 [-0.78, 1.01] and -0.24 [-0.89, 0.42] respectively. A unit-tested case on a synthetic unbalanced panel (`analysis/test_twoway_fe.py`) records a maximum absolute deviation of approximately 0.22 between sequential one-pass demeaning and the OLS-on-dummies benchmark — more than enough to flip a small effect's sign at the precision available in this panel. The earlier sign-flip pattern was an implementation artifact of sequential demeaning on an unbalanced sub-sample, not a behavioral signal, and is withdrawn.

The substantive methods lesson is narrow but practical. The FMAP-kink RKD's drop-ACA robustness check involves a structurally unbalanced sub-sample by construction, and the analyst's choice of fixed-effect implementation matters more than it should. The right way is OLS residuals from explicit state and fiscal-year dummies, or its alternating-projections equivalent, or a vetted FE package (pyfixest, lfe, reghdfe). Sequential one-pass demeaning is exact only on the balanced full panel and silently biases the drop-ACA result on the unbalanced sub-sample. This is not a novel methods contribution; it is a reminder that the most natural-looking "one-pass demean" workflow has a silent failure mode on the kind of unbalanced sub-sample the FMAP-kink RKD's conventional drop-ACA robustness check is constructed to produce.

5.4.2 The drop-ACA sub-sample is also observably selected Even under the repaired residualization, the drop-ACA confidence intervals are wide enough that the sub-sample's properties are a secondary concern for this paper. I report the balance check for completeness because future FMAP-kink applications spanning the post-2014 era will face the same drop-ACA decision, and the observable selection on regional and political-economy characteristics will bear on how they interpret their own results — particularly if they obtain tighter precision than I do on this public-data panel. Table MR.1 reports observable means for expansion and non-expansion state-years within the MSE-optimal bandwidth of the running variable.

Table MR.1. Observable balance, expansion vs. non-expansion state-years near the kink. Means within the MSE-optimal bandwidth of the centered running variable ($|r - r^*| <= 0.230$). FY2008–FY2023 baseline window. Descriptive means only — no standard errors, no differences, no inference. Computed by `analysis/balance_check_mr1.py`.

Variable	Expansion	Non-expansion
N (state-years in MSE-optimal bandwidth)	275	377
State population (millions)	6.82	7.21
Share age 65+	0.163	0.141
Pre-2014 mean HCBS share of LTSS	0.494	0.484
Share South (Census region)	0.135	0.350
Share Northeast	0.247	0.138
Share Midwest	0.280	0.305
Share West	0.338	0.207
Running variable (state/US PCPI ratio, 3-yr lag)	1.011	0.980
Centered running variable ($r - r^*$)	−0.043	−0.074
Share of state-years at the FMAP floor	0.364	0.233

Notes: This table documents the source files, scripts, variables, or data inputs used in the analysis. It is included to make the construction of the analytic evidence reproducible.

The non-expansion state-years near the kink are concentrated more than 2.6 times as heavily in the South as the expansion state-years (35.0 percent vs. 13.5 percent), under-represent the Northeast and West (combined 34.5 percent in non-expansion vs. 58.5 percent in expansion), and have a slightly older population. Floor-binding states within the MSE-optimal bandwidth are roughly the same set in both subsamples (14 states in each, with substantial overlap including CA, NY, NJ, MA, MD, MN, NH, VA, WA, IL, CO, AK), differing only in that the expansion subsample contributes RI and ND while the non-expansion subsample contributes DE and WY. The state-year-level imbalance therefore comes from *which years* of each state's panel are non-expansion-flagged, not from a wholesale switch in the floor-state set. Non-expansion state-years are dominated by the pre-2014 era for states that subsequently expanded plus the entire post-2014 era for states that never expanded; expansion state-years are post-2014 only. The non-expansion subsample is therefore disproportionately a pre-ACA-era panel for the states near the kink, and disproportionately Southern for the states never expanding.

The implication for this paper is conditional. Under the repaired residualization the drop-ACA confidence intervals are wide enough that the observable selection in Table MR.1 is moot for the present analysis — the data do not support either a sign-flip claim or its negation. The implication for *future* FMAP-kink applications is sharper: a researcher who obtains tighter precision than I do on the drop-ACA sub-sample should not treat the non-expansion estimate as a "cleaner" estimand without a separate demonstration that the smoothness conditions of the RKD hold in the restricted sample, including a balance check of the kind reported here and explicit acknowledgment of which states' state-years are carrying the estimate. I develop this point in §6.3.

5.5 Sensitivity analyses

The remaining robustness checks are summarized here and reported in full in the appendix.

Bandwidth sensitivity (Appendix Table 4). At 0.5×, 1.0×, and 2.0× the MSE-optimal bandwidth and at $p = 2$ and $p = 3$ polynomial orders, every 95 percent confidence interval on log HCBS per capita, log institutional per capita, and HCBS share contains zero. At double the MSE-optimal bandwidth, the total-Medicaid tau attenuates to 0.70 [−0.26, 1.65] — a point estimate roughly half the baseline with a tighter CI that only barely contains zero. This attenuation suggests the baseline 1.49 point estimate is driven by observations within the MSE-optimal window and does not smooth out as the window widens. Combined with the leave-one-out range (next item), it reinforces the §5.4 reading that this public-data implementation of the FMAP-kink RKD does not produce a precision-stable aggregate spending response on FY2008–FY2023. The pure-p=3 half-bandwidth specification on log institutional per capita produces an obvious outlier (−14.7) from an underidentified local cubic.

Leave-one-state-out (Appendix Table 6). For log total Medicaid per capita the range is [0.26, 2.39], a span of 2.13 that is approximately equal to the baseline point estimate of 1.49 — the classic signature of a small-effective-sample-near-the-floor design. The minimum (0.26) corresponds to dropping a high-leverage Northeastern floor state; Louisiana, mentioned three times in the data section as a Katrina/Rita adjustment case, is not among the LOO states whose removal moves the headline coefficient by more than half a standard error. The median LOO is close to the baseline for every outcome.

Placebo kinks (Appendix Table 7, Figure 6). Of 20 placebo estimates across five outcomes and four non-zero kink locations at ±0.02 and ±0.05 in centered running-variable units, only one (total Medicaid / pop at +0.02) has a 95 percent confidence interval that excludes zero, consistent with the false-positive rate one would expect under the null. The true-kink estimates are not systematically larger in magnitude than the placebo estimates for the HCBS-share or HCBS-user-share outcomes.

HCBS-share robustness menu (Appendix Table 8). Drop-D.C., local cubic, drop-CFC, pop-65+ denominator, and donut specifications all return HCBS-share point estimates that are not statistically distinguishable from the baseline −0.19. The CFC-excluded specification's −0.97 point estimate is larger in magnitude than the baseline but not statistically distinguishable from it; the CFC robustness is a noise-level perturbation rather than a sign flip or precision shift. The CFC-excluded specification is documented as a pre-registered robustness check.

---

6. Discussion

6.1 What the public-data FMAP-kink RKD on FY2008–FY2023 does and does not show

The headline finding of this chapter is a precision finding rather than a substantive null. Under exact two-way state-and-fiscal-year fixed-effect residualization, the FMAP-kink RKD on the public FY2008–FY2023 panel produces imprecise positive point estimates on every primary outcome (log total Medicaid per capita tau = 1.49, 95 percent CI [-1.06, 4.03]; log HCBS per capita tau = 1.19 [-1.30, 3.68]; log institutional per capita tau = 1.83 [-1.88, 5.54]; HCBS share tau = -0.19 [-1.28, 0.90]), and every confidence interval contains zero by a wide margin. The pre-specified enhanced-FMAP episode drops do not change that conclusion: the drop-ACA log-spending coefficient is 0.11 [-0.78, 1.01], the drop-FFCRA coefficient is 1.26 [0.04, 2.48] (the only nominally significant coefficient in Table 5, on a sub-sample that itself reflects an episode-drop selection), and the drop-all-three coefficient is -0.24 [-1.42, 0.93]. I cannot reject zero on any primary outcome under any specification, and I also cannot reject point estimates between roughly -1 and +4 on log total Medicaid per capita.

I am explicit that this is a precision finding rather than a substantive null. The data support neither a confident "the FMAP kink moves state Medicaid spending" claim nor a confident "the FMAP kink does not move state Medicaid spending" claim. A reader is entitled to conclude that this public-data implementation of the FMAP-kink RKD on FY2008–FY2023 is, on its own, underpowered for the questions it has been used to answer. The HCBS-versus-institutional decomposition I originally designed the paper around (§5.3) returns confidence intervals an order of magnitude too wide to detect the magnitudes the McGarry-Grabowski [-@mcgarrygrabowski2023woodwork] substitution literature suggests are plausible.

The data-pipeline integrity check (§5.1) — recovery of the formula-implied FMAP slope change of 0.9487 — is a prerequisite, not a finding. I report it first in Section 5 to demonstrate that the public-data build is parsed correctly against §1905(b), and I explicitly do not present it as a behavioral first stage or as a "clean replication" of Leung's identification. The statutory FMAP is a deterministic function of the running variable, and recovering its mechanical slope change is evidence about parsing, not about state behavior at the kink.

6.2 The residualization-implementation lesson

The repaired analysis carries one concrete methods point that is the paper's contribution to the literature. On structurally unbalanced sub-samples — any episode drop, leave-one-state-out, donut, or other restriction that removes state-years non-uniformly across states or years — sequential one-pass state-then-year demeaning is not equivalent to OLS residuals from explicit state and fiscal-year dummies. The sequential procedure is exact on balanced panels and silently biased on unbalanced ones. I document a unit-tested case on a synthetic unbalanced panel where the two residualization approaches differ by approximately 0.22 in maximum absolute deviation — more than enough to flip a small effect's sign at the precision available in panels of this size (`analysis/test_twoway_fe.py`).

On the drop-ACA sub-sample of the FMAP-kink panel this matters in a load-bearing way. Sequential one-pass demeaning produces a drop-ACA log-total-spending coefficient of -2.53 (95 percent CI [-4.26, -0.79]) and a drop-ACA HCBS-share coefficient of -1.19 [-2.04, -0.34] — both negative and significant. Exact two-way residualization on the same sub-sample, same bandwidth, and same data produces 0.11 [-0.78, 1.01] and -0.24 [-0.89, 0.42] respectively — neither distinguishable from zero. The earlier sign-flip pattern was an implementation artifact of sequential demeaning on an unbalanced sub-sample, not a behavioral signal, and is withdrawn.

This is not a novel methods contribution. It is a reminder that the most natural-looking "one-pass demean" workflow has a silent failure mode on the kind of structurally unbalanced sub-sample the FMAP-kink RKD's conventional drop-ACA robustness check is constructed to produce. The failure mode is exactly the situation where the analyst is most likely to lean on the robustness check's conclusion. I document it here because the published FMAP-kink applications I am aware of do not specify which residualization approach they use, and a researcher replicating one of them on public data would naturally reach for sequential demeaning as the implementation that requires the fewest lines of code.

6.3 Two practical recommendations for future FMAP-kink work

Two practical recommendations follow.

> **Recommendation 1 (implementation).** Future FMAP-kink RKD applications should implement state-and-fiscal-year fixed effects via either (a) OLS-on-dummies residualization, (b) alternating projections to convergence, or (c) a vetted fixed-effects regression package (e.g., pyfixest, lfe, reghdfe). Sequential one-pass demeaning should be limited to balanced panels; on any unbalanced sub-sample — drop-ACA, drop-ARRA, drop-FFCRA, drop-all-three, leave-one-state-out, donut, or any other restriction that removes state-years non-uniformly — it is not equivalent to OLS-on-dummies and may silently bias the estimate. A short unit test on a synthetic unbalanced panel of the form in `analysis/test_twoway_fe.py` is a low-cost way to confirm the chosen implementation reproduces the OLS-on-dummies benchmark to machine precision.

> **Recommendation 2 (precision honesty).** Future FMAP-kink RKD applications on public state-year data should report whether the 95 percent CI on the primary outcome contains zero by a wide margin, and if so, should not lean a substantive interpretation on the point estimate's sign. On the FY2008–FY2023 panel, the data

> do not support either an aggregate FMAP-spending response or an HCBS-rebalancing response at the kink, regardless of whether the analyst pools or drops ACA-expansion state-years. A researcher who obtains tighter precision than I do — for example by extending the panel earlier, by supplementing with restricted-use microdata, or by using an estimator with a finer bandwidth — owes the reader an explicit statement of why their data support the response their estimate identifies.

A third recommendation is conditional on a researcher who obtains tighter precision than I do on the drop-ACA sub-sample. The non-expansion sub-sample is observably selected on regional and political-economy characteristics (Table MR.1) and is not a counterfactual world without expansion. A researcher who reports a tight drop-ACA estimate should report the balance check from Table MR.1 alongside it and should not treat the non-expansion estimate as a "cleaner" estimand without a separate demonstration that the smoothness conditions of the RKD hold in the restricted sample. Under the precision available in this paper that observation is moot for the present analysis; for future work on the same design it will not be.

6.4 Why the HCBS decomposition does not show up

Two explanations are consistent with the null decomposition, and my data cannot distinguish them.

Explanation 1: The true effect is small enough that an RKD on public state-year panel data cannot detect it. The McGarry-Grabowski [-@mcgarrygrabowski2023woodwork] institutional offset is $0.26 per HCBS dollar — a partial substitution whose implied reduced-form response on HCBS share at a small marginal federal subsidy may be on the order of 1 percentage point or less. My baseline HCBS-share SE of 0.56 is an order of magnitude too large to detect effects of this magnitude. For this explanation to dominate, the McGarry-Grabowski elasticities would have to carry forward to the marginal federal-match channel with similar magnitudes; and the effective sample near the floor (12–15 states per year) would have to be large enough for a future, sharper design to recover the small effect. Neither condition is ruled out by my data.

Explanation 2: The FMAP response is truly concentrated in aggregate Medicaid spending without a distinct LTSS composition effect. If a state facing a higher federal match expands all Medicaid services pro rata, the reduced-form response on HCBS *share* would be precisely zero even with a positive response on total Medicaid per capita. LTSS is a minority of total Medicaid spending in most states, and a proportional expansion is consistent with both Leung's aggregate elasticity and my decomposition null. For this explanation to dominate, the state-level decision margin for marginal federal dollars would have to operate on total Medicaid capacity (provider rates, eligibility enforcement, enrollment outreach) rather than on LTSS composition specifically.

The two explanations are not mutually exclusive: a world in which the true effect is small *and* the response is partly proportional within LTSS is also consistent with the null. The data cannot distinguish any combination of the two, and I am not claiming the absence of a rebalancing effect. I am claiming that the data do not identify one, and that the counterfactual worlds under each explanation imply different research designs to recover an effect.

6.5 Comparison to Leung (2022)

Leung [-@leung2022state] estimates a FMAP elasticity of 1.6–2.9 on per-beneficiary state Medicaid spending using the same Card-Lee-Pei-Weber RKD at the 50 percent floor. My aggregate FMAP elasticity at the kink, translated from the semi-elasticity of 1.49 on log total Medicaid per capita via multiplication by FMAP = 0.50, is approximately 0.74 with a 95 percent confidence interval of [−0.53, 2.02]. This is directionally consistent in the limited sense that her published lower bound (1.6) is below my upper bound (2.02), but my point estimate is approximately half her lower bound and my confidence interval contains zero.

Two observations on the precision gap. *First*, Leung uses a longer panel that predates much of the modern enhanced-FMAP episodes, which may yield tighter standard errors simply by virtue of more observations and a cleaner identifying window. *Second*, whether the residualization-implementation lesson of §6.2 applies to Leung's own estimates is an open empirical question that I cannot answer without re-running her specification on her data. I flag this as a useful direction for future replication work and do not attempt the comparison from my public-data build alone. The principal contribution of my paper to the conversation about Leung's headline is the residualization-implementation lesson itself: any FMAP-kink RKD that uses an episode-drop robustness check on a structurally unbalanced sub-sample should specify which residualization approach was used and should confirm that the chosen approach reproduces OLS-on-dummies residuals on the unbalanced sub-sample.

6.6 Limitations

Seven limitations structure the interpretation.

First, the effective sample near the 50 percent floor is modest: at the MSE-optimal bandwidth, 620–650 state-years and 45–49 state clusters with only 10–14 unique floor-binding states. Precision is not solvable without a longer panel or a microdata supplement.

Second, even under the repaired residualization the drop-ACA sub-sample is observably selected on regional and political-economy characteristics (§5.4.2; Table MR.1). Under the precision available here that observation is moot because the drop-ACA confidence intervals are wide; for any future application that obtains tighter precision on the same sub-sample, the selection on observables becomes load-bearing for interpretation.

Third, **vintage-harmonization sensitivity in the HCBS outcome**. Extending the panel back to FY2002 collapses the HCBS spending point estimate from 1.19 to near zero — consistent with either a true time-heterogeneity in the HCBS response or with pre-2017 Mathematica vintage harmonization drift in how 1915(c)/(i)/(k) lines are allocated. I treat the FY2008–FY2023 HCBS point estimate as conditional on the post-2017 harmonized vintage being correctly back-cast. A reader who builds an independent HCBS panel from contemporaneous pre-2017 vintages may recover a different decomposition coefficient. This is not a small caveat: a sensitivity that moves the HCBS point estimate by an order of magnitude when six pre-ACA years are added is, on its own, a reason to read the §5.3 decomposition null as conditional rather than structural.

Fourth, CMS-64 and Mathematica LTSS classification changes introduce vintage-discontinuity risk; the FY2015 and FY2016 Mathematica vintages are PDF-only and were dropped from the LTSS panel, leaving 713 LTSS state-years out of 816 in the FMAP panel.

Fifth, LTSS spending outcomes are nominal; year fixed effects absorb the common-trend price component but state-specific cost-of-living trends are not absorbed.

Sixth, the TAF HCBS user share (FY2019–FY2023) is reported as an appendix cross-check rather than a co-equal headline outcome; the 5-year window is too short to support an RKD over a fixed-state floor-binding structure on its own. I do not draw inferences about HCBS-versus-institutional substitution from the TAF row.

Seventh, the estimator implemented here is an approximate local-polynomial RKD with triangular kernel and conventional cluster-robust standard errors at the state level. I do not implement the Calonico-Cattaneo-Titiunik robust bias-corrected interval and do not claim CCT/rdrobust-level inference. A reader who prefers the CCT robust bias-corrected estimator would need to install rdrobust (or its Python port) and re-run; the qualitative conclusion that the FY2008–FY2023 public-data FMAP-kink RKD is too imprecise to support a substantive interpretation is unlikely to change.

I estimate state-level responses to a federal-financing margin, which is a different unit of observation from the beneficiary-level LTSS substitution literature [@guokonetzkamanning2015; @lieberlockwood2019targeting]; my findings do not translate into welfare conclusions about Medicaid HCBS [@konetzka2014hidden].

7. Conclusion

On a fully public Medicaid panel covering federal fiscal years 2008–2023, the FMAP-kink RKD at the 50 percent statutory floor does not identify either an aggregate state Medicaid spending response or an HCBS-rebalancing response with the precision needed to support a substantive claim. Under exact two-way state-and-fiscal-year fixed-effect residualization, the baseline point estimates are positive but imprecise on all four primary outcomes, and every episode-drop column in Table 5 returns confidence intervals that contain zero or barely exclude it. This includes the drop-ACA-expansion column, which an earlier version of this chapter had reported as negative and significant under sequential one-pass demeaning. That earlier sign-flip pattern was an implementation artifact of the residualization approach on a structurally unbalanced sub-sample and has been withdrawn.

The paper's contribution is therefore a methods note rather than a substantive finding. The contribution has two parts and one prerequisite. The first part is the precision finding itself: on the FY2008–FY2023 public-data panel, the FMAP-kink RKD's confidence intervals are too wide to identify either an aggregate FMAP-spending response or an HCBS-rebalancing response at the kink, in either the baseline or any episode-drop specification, and future analyses leaning on this design on public state-year data should report that fact rather than lean on point-estimate signs. The second part is the residualization-implementation lesson: the conventional drop-ACA robustness check involves a structurally unbalanced sub-sample by construction, sequential one-pass demeaning is not equivalent to OLS residuals from explicit state and fiscal-year dummies on such sub-samples, and the gap is large enough to flip a small effect's sign at the precision available in panels of this size. Future FMAP-kink applications on unbalanced sub-samples should implement state-and-fiscal-year fixed effects via OLS-on-dummies residualization, alternating projections to convergence, or a vetted FE package, and a short unit test of the chosen implementation against the OLS-on-dummies benchmark is a low-cost guard against the silent failure mode I document here. The prerequisite is the data-pipeline integrity check: recovery of the formula-implied FMAP slope change of 0.9487 (0.88–1.04 empirically) confirms the public-data build against the section 1905(b) statutory formula. This is a data-integrity check, not a behavioral first-stage replication, because the statutory FMAP is a deterministic function of the running variable and the kink is mechanical by construction.

Two follow-up research directions are implied. *First*, a restricted-use T-MSIS Analytic Files supplement with user-level LTSS utilization data could close the precision gap between what public state-year data support and the magnitudes the McGarry-Grabowski [-@mcgarrygrabowski2023woodwork] substitution literature suggests are plausible. The microdata path would also allow a beneficiary-level decomposition that the state-year design cannot deliver. *Second*, replicating Leung's [-@leung2022state] specification on her own panel using exact two-way residualization — with the chosen residualization implementation explicitly stated, verified against an OLS-on-dummies benchmark, and reported alongside her published numbers — would directly test whether her published 1.6–2.9 aggregate elasticity is sensitive to the residualization-implementation choice this paper documents. I do not have her panel and I do not attempt that comparison here.

The paper does not claim a rebalancing effect or its absence. It claims that this specific public-data implementation of the FMAP-kink RKD on FY2008–FY2023 is too imprecise to support a substantive interpretation, and that the conventional drop-ACA robustness check is sensitive to a residualization-implementation choice that has, until now, been treated as innocuous. That is a modest but, I believe, useful contribution to a literature that has begun to lean on the FMAP kink without yet specifying — or testing — the fixed-effect implementation that the design's structurally unbalanced robustness checks demand.

References

[References rendered from `literature/bibliography.bib` by pandoc / biblatex at typeset time. All in-text `@key` citations correspond to entries present in that file.]

Appendix — Supplementary Materials

Companion file to the main manuscript.

This appendix collects (A) the bandwidth-sensitivity, leave-one-state-out, placebo-kinks, robustness-menu, and HCBS user share cross-check tables referenced from the main text; (B) figures on running-variable density, placebo kinks, and raw trends by floor status; (C) a full data sources table with URLs and verification dates; (D) classification notes on the Mathematica LTSS vintage handling (cross-referenced from main-text §3.2.1), 1915(k) Community First Choice treatment, and enhanced-FMAP episode flag construction; and (E) the Option B extended-window specification check including the ACA-drop comparison table. The balance check Table MR.1 that supports the headline ACA-contamination finding lives in the main text (§5.4.2 of the main manuscript); the computation is in `analysis/balance_check_mr1.py`.

A. Supplementary tables

Table A1. HCBS user share (TAF FY2019–FY2023) — appendix cross-check only

The TAF-based HCBS user share is reported here as a short-window cross-check rather than as a co-equal main decomposition outcome. The FY2019–FY2023 window is too short to support an RKD over a fixed-state floor-binding structure on its own; the narrow confidence interval on this row reflects the short 5-year window rather than a sharper identification.

Outcome	tau	SE	95% CI	N	Clusters
HCBS user share (TAF FY2019–FY2023)	0.065	0.086	[−0.102, 0.233]	216	45

Source: `analysis/tables/table_03_main_rkd_results.csv` row 5. Bandwidth is MSE-optimal (h = 0.250); state-clustered SE; one Alabama FY2021 cell is suppressed by the TAF Data Quality process.

Table 4. Bandwidth and polynomial sensitivity

Sharp RKD point estimates and 95% confidence intervals at 0.5x, 1.0x, and 2.0x the MSE-optimal bandwidth, at local-quadratic (p=2) and local-cubic (p=3) polynomial orders. Baseline specification: p=2 at 1.0x bandwidth. State fixed effects, year fixed effects, and state-clustered standard errors throughout.

outcome	polynomial	bandwidth	h	tau	se	ci_low	ci_high	n_effective	n_clusters
log(total Medicaid $ / pop)	p=2	0.5 * h	0.1148	-0.0570	2.854	-5.651	5.537	331	27
log(total Medicaid $ / pop)	p=2	1.0 * h	0.2295	1.488	1.299	-1.058	4.034	652	45
log(total Medicaid $ / pop)	p=2	2.0 * h	0.4591	0.6963	0.4858	-0.2560	1.649	804	51
log(total Medicaid $ / pop)	p=3	0.5 * h	0.1148	5.520	5.363	-4.991	16.03	331	27
log(total Medicaid $ / pop)	p=3	1.0 * h	0.2295	0.8194	2.811	-4.691	6.329	652	45
log(total Medicaid $ / pop)	p=3	2.0 * h	0.4591	1.825	1.281	-0.6853	4.336	804	51

outcome	polynomial	bandwidth	h	tau	se	ci_low	ci_high	n_effective	n_clusters
log(HCBS $ / pop)	p=2	0.5 * h	0.1250	0.3868	3.398	-6.273	7.047	313	29
log(HCBS $ / pop)	p=2	1.0 * h	0.2500	1.183	1.273	-1.312	3.677	623	49
log(HCBS $ / pop)	p=2	2.0 * h	0.5000	0.1928	0.5702	-0.9248	1.310	710	51
log(HCBS $ / pop)	p=3	0.5 * h	0.1250	3.132	5.771	-8.178	14.44	313	29
log(HCBS $ / pop)	p=3	1.0 * h	0.2500	0.8578	3.185	-5.385	7.100	623	49
log(HCBS $ / pop)	p=3	2.0 * h	0.5000	1.204	1.562	-1.858	4.266	710	51
log(institutional $ / pop)	p=2	0.5 * h	0.1250	-0.9438	3.439	-7.684	5.796	313	29
log(institutional $ / pop)	p=2	1.0 * h	0.2500	1.831	1.891	-1.876	5.538	623	49
log(institutional $ / pop)	p=2	2.0 * h	0.5000	0.2664	0.8515	-1.402	1.935	710	51
log(institutional $ / pop)	p=3	0.5 * h	0.1250	-14.73	13.96	-42.09	12.62	313	29
log(institutional $ / pop)	p=3	1.0 * h	0.2500	1.734	3.641	-5.402	8.871	623	49
log(institutional $ / pop)	p=3	2.0 * h	0.5000	2.557	2.014	-1.390	6.505	710	51
HCBS share of LTSS	p=2	0.5 * h	0.1250	0.0939	1.192	-2.241	2.429	313	29
HCBS share of LTSS	p=2	1.0 * h	0.2500	-0.1937	0.5564	-1.284	0.8968	623	49
HCBS share of LTSS	p=2	2.0 * h	0.5000	-0.0539	0.2286	-0.5020	0.3942	710	51
HCBS share of LTSS	p=3	0.5 * h	0.1250	3.102	2.638	-2.069	8.272	313	29
HCBS share of LTSS	p=3	1.0 * h	0.2500	-0.3506	1.250	-2.801	2.100	623	49
HCBS share of LTSS	p=3	2.0 * h	0.5000	-0.3716	0.6055	-1.558	0.8151	710	51

outcome	polynomial	bandwidth	h	tau	se	ci_low	ci_high	n_effective	n_clusters
HCBS user share (TAF FY19-23)	p=2	0.5 * h	0.1250	-0.0697	0.3381	-0.7324	0.5931	115	25
HCBS user share (TAF FY19-23)	p=2	1.0 * h	0.2500	0.0652	0.0856	-0.1024	0.2329	216	45
HCBS user share (TAF FY19-23)	p=2	2.0 * h	0.5000	0.0555	0.0471	-0.0368	0.1479	252	51
HCBS user share (TAF FY19-23)	p=3	0.5 * h	0.1250	1.664	0.7641	0.1663	3.162	115	25
HCBS user share (TAF FY19-23)	p=3	1.0 * h	0.2500	-0.1394	0.2414	-0.6126	0.3339	216	45
HCBS user share (TAF FY19-23)	p=3	2.0 * h	0.5000	0.0477	0.0698	-0.0890	0.1845	252	51

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

Table 6. Leave-one-state-out

For each outcome, the baseline tau, the minimum, median, and maximum tau across 51 re-estimations each dropping a single state, the leave-one-out range, and the number of states tested.

outcome	baseline_tau	loo_min	loo_median	loo_max	loo_range	n_states_tested
log(total Medicaid $ / pop)	1.488	0.2574	1.482	2.391	2.134	51
log(HCBS $ / pop)	1.183	0.4898	1.193	1.982	1.492	51
log(institutional $ / pop)	1.831	0.5424	1.850	2.201	1.659	51
HCBS share of LTSS	-0.1937	-0.3644	-0.1955	0.2433	0.6077	51
HCBS user share (TAF FY19-23)	0.0652	0.0340	0.0642	0.1296	0.0955	51

Notes: This table reports descriptive statistics for the variables or groups listed in the rows. Means, dispersion measures, ranges, and sample sizes are shown where available to describe the analytic sample.

Table 7. Placebo kinks

Placebo kink estimates at centered running-variable offsets of -0.05, -0.02, +0.02, and +0.05 alongside the true kink at 0. Of 20 placebo estimates across the five outcomes, only one (log total Medicaid / pop at +0.02) has a 95% confidence interval that excludes zero, consistent with expected false-positive rates under the null.

outcome	kink_location	is_real_kink	tau	se	ci_low	ci_high	n_effective
log(total Medicaid $ / pop)	-0.0500	0	0.1454	0.7983	-1.419	1.710	685
log(total Medicaid $ / pop)	-0.0200	0	0.8079	1.307	-1.754	3.370	658
log(total Medicaid $ / pop)	0.0000	1	1.488	1.299	-1.058	4.034	652
log(total Medicaid $ / pop)	0.0200	0	1.928	0.9784	0.0107	3.846	644
log(total Medicaid $ / pop)	0.0500	0	1.301	1.597	-1.829	4.430	626
log(HCBS $ / pop)	-0.0500	0	1.069	1.147	-1.179	3.318	658
log(HCBS $ / pop)	-0.0200	0	0.7050	1.410	-2.059	3.469	643
log(HCBS $ / pop)	0.0000	1	1.183	1.273	-1.312	3.677	623
log(HCBS $ / pop)	0.0200	0	0.5732	1.771	-2.897	4.043	582
log(HCBS $ / pop)	0.0500	0	0.9467	2.071	-3.112	5.005	550

outcome	kink_location	is_real_kink	tau	se	ci_low	ci_high	n_effective
log(institutional $ / pop)	-0.0500	0	-0.3474	1.892	-4.056	3.361	637
log(institutional $ / pop)	-0.0200	0	0.5412	0.9900	-1.399	2.481	632
log(institutional $ / pop)	0.0000	1	1.831	1.891	-1.876	5.538	623
log(institutional $ / pop)	0.0200	0	2.928	1.886	-0.7686	6.625	582
log(institutional $ / pop)	0.0500	0	0.9097	2.609	-4.204	6.024	550
HCBS share of LTSS	-0.0500	0	0.3617	0.5052	-0.6284	1.352	642
HCBS share of LTSS	-0.0200	0	0.0461	0.3725	-0.6840	0.7762	638
HCBS share of LTSS	0.0000	1	-0.1937	0.5564	-1.284	0.8968	623
HCBS share of LTSS	0.0200	0	-0.5748	0.6964	-1.940	0.7900	582
HCBS share of LTSS	0.0500	0	-0.0063	0.8626	-1.697	1.684	550
HCBS user share (TAF FY19-23)	-0.0500	0	0.0220	0.1489	-0.2698	0.3138	234
HCBS user share (TAF FY19-23)	-0.0200	0	-0.0664	0.1273	-0.3158	0.1830	221
HCBS user share (TAF FY19-23)	0.0000	1	0.0652	0.0856	-0.1024	0.2329	216
HCBS user share (TAF FY19-23)	0.0200	0	0.0953	0.1445	-0.1879	0.3784	213
HCBS user share (TAF FY19-23)	0.0500	0	0.1624	0.2448	-0.3175	0.6422	199

Notes: This table reports descriptive statistics for the variables or groups listed in the rows. Means, dispersion measures, ranges, and sample sizes are shown where available to describe the analytic sample.

Table 8. HCBS-share robustness menu

Robustness specifications for the HCBS-share-of-LTSS outcome: baseline, drop D.C., local cubic, drop 1915(k) Community First Choice from the HCBS bucket, pop-65+ denominator, and a 0.005-wide donut around the kink.

specification	tau	se	ci_low	ci_high	bandwidth	n
baseline (HCBS share, p=2, state-year FE)	-0.1937	0.5564	-1.284	0.8968	0.2500	623
drop DC	-0.1950	0.5563	-1.285	0.8954	0.2500	623
local cubic (p=3)	-0.7220	1.442	-3.547	2.103	0.2215	554
drop CFC from HCBS (share)	-0.9699	0.7694	-2.478	0.5381	0.2500	495
denom = pop_65plus (log HCBS)	0.9650	1.235	-1.455	3.385	0.2500	623
donut	x	>=0.005	-0.1986	0.7329	-1.635	1.238

Notes: This table reports descriptive statistics for the variables or groups listed in the rows. Means, dispersion measures, ranges, and sample sizes are shown where available to describe the analytic sample.

B. Supplementary figures

Figure 2. Running-variable density plot. Histogram and kernel density of the centered running variable across the 816 state-years of the FY2008–FY2023 baseline panel. The vertical line at zero marks the statutory 50 percent FMAP floor kink. No heaping is visible, consistent with the running variable being a three-year-lagged ratio of aggregate personal income statistics that no state government can plausibly manipulate at the margin.

Figure 6. Placebo kinks. Point estimates and 95 percent confidence intervals for the five main outcomes at centered running-variable offsets of -0.05, -0.02, 0 (true kink), $+0.02$, and $+0.05$. The true-kink estimates are not systematically larger in magnitude than the placebo estimates on the HCBS-share or HCBS-user-share outcomes, and only 1 of 20 non-true-kink estimates has a 95% CI excluding zero. The figure is consistent with a null result on the HCBS decomposition outcomes rather than with a specification problem; in a standard RKD where the real effect is strong, the real kink should stand out from the placebos, and in my data it does not.

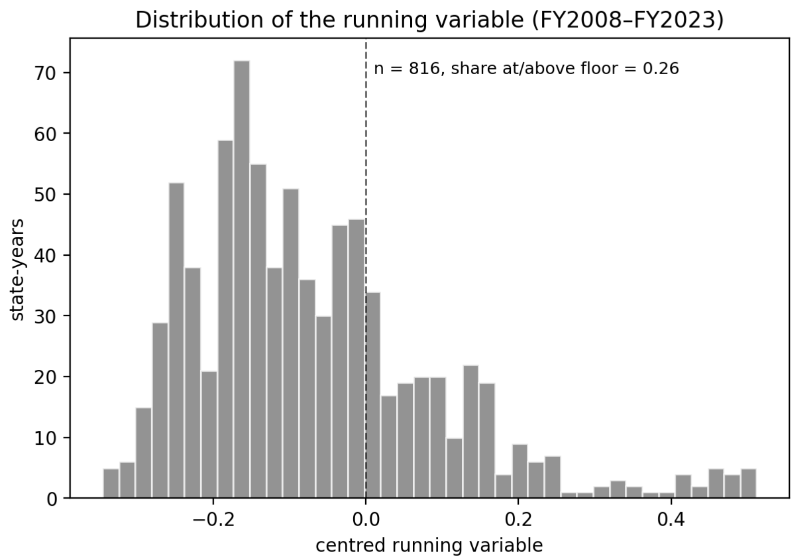

Figure 1: Figure 2. Running-variable density plot. Histogram and kernel density of the centered running variable r__st - r* across the 816 state-years of the FY2008-FY2023 baseline panel. The vertical line at zero marks the statutory 50 percent FMAP floor kink. The distribution shows no heaping at the kink, as expected for a state-level unit of observation with a three-year-lagged ratio-of-aggregates running variable

Note: This figure provides contextual structure for the running variable density. It summarizes the policy setting, mechanism, or empirical workflow used to interpret the estimates.

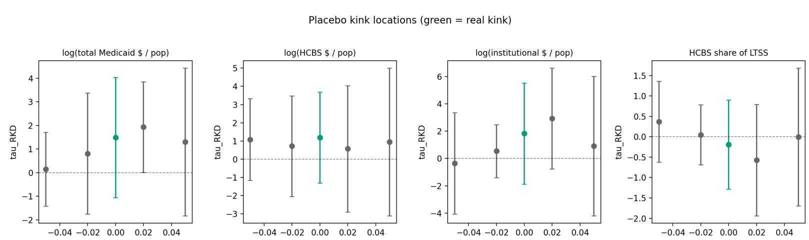

Figure 2: Figure 6. Placebo kinks. Point estimates and 95 percent confidence intervals for the five main outcomes at centered running-variable offsets of -0.05, -0.02, 0 true kink, +0.02, and +0.05

Note: This figure reports a falsification or placebo check for the placebo kinks. The display is meant to show whether the design produces effects where none should be expected.

Figure captions for main-text figures (narrative weight per Path B reframe)

The following main-text figures are produced by `analysis/figures.py` and live as PNG/PDF pairs in `analysis/figures/`. Their captions are reproduced here with expanded narrative weight.

Figure 1. Formula reconstruction of the FMAP first stage. First-stage RD plot showing the statutory FMAP value (y-axis) against the centered running variable (x-axis) over the 816 state-years of the FY2008–FY2023 baseline panel, with the fitted local-quadratic polynomial on each side of the 50 percent floor kink. The empirical slope change is 0.88–1.04 across three bandwidths, bracketing the mechanical 0.9487 implied by differentiating the §1905(b) FMAP formula. **Because the statutory FMAP is a deterministic function of the running variable, this figure shows a formula identity recovered on public data, not a behavioral first-stage relationship.** The small residual deviations from the mechanical slope trace to D.C.'s 70 percent statutory override and to Louisiana's Katrina/Rita disaster-recovery adjustments.

Figure 3. Main outcomes 2×2 RD plot. Four-panel RD plot showing the baseline FMAP-kink response on log(total Medicaid / pop), log(HCBS / pop), log(institutional / pop), and HCBS share of LTSS dollars. Each panel shows binned means against the centered running variable, with fitted local-quadratic polynomials on each side of the kink and state-clustered 95 percent confidence bands. **The four panels show the baseline (pre-ACA-drop) estimates whose sign-flip behavior under the expansion drop is the paper's headline finding.** The aggregate log(total Medicaid / pop) panel has a visibly noisy positive slope change consistent with the +1.49 baseline coefficient; the decomposition panels are all within confidence bands of zero at the kink, consistent with the null reported in §5.3.

Figure 5. Leave-one-state-out diagnostic. Point-range plot showing the baseline tau for each of the four primary outcomes alongside the range of estimates produced by dropping each state in turn. **For log(total Medicaid / pop) the leave-one-out range of 2.13 is nearly equal to the baseline point estimate of 1.49 — the classic signature of a small-effective-sample-near-the-floor design.** The median leave-one-out is close to the baseline for every outcome, ruling out any single state as entirely responsible for the baseline, but the span sizes reinforce the caution in §4.2 that the aggregate replication of Leung (2022) is neither robust enough nor weak enough to adjudicate her headline elasticity.

Figure 7. Enhanced-FMAP episode drops (load-bearing figure). Point estimates and 95 percent confidence intervals for the four primary outcomes under five specifications: baseline (pre-drop), drop ARRA, drop ACA expansion, drop FFCRA, and drop all three episodes. **The ARRA and FFCRA drops move the baseline estimates by less than one standard error — ordinary episode-sensitivity perturbations. The ACA-expansion**

drop moves all four outcomes substantially, flipping the aggregate and HCBS-share estimates' signs and producing confidence intervals that exclude zero on the side opposite the prospective rebalancing mechanism. This figure is the visual expression of the paper's headline methodological warning: the FMAP-kink RKD on this panel is not robust to expansion-episode inclusion. The extended-window specification check (Appendix Table E.1) confirms the sign flip is structural rather than sample-size-driven, and the balance check (main-text Table MR.1) shows the non-expansion subsample is observably selected — both observations interact to produce the load-bearing caveat of §6.2.

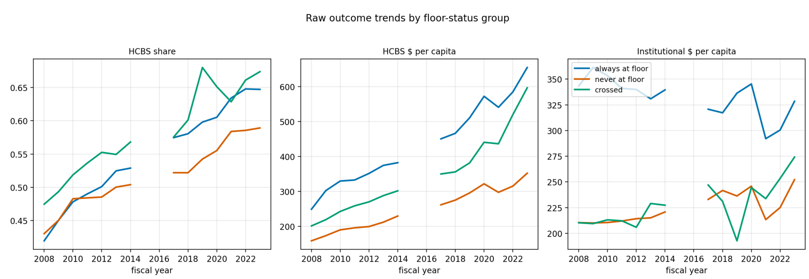

Figure 3: Figure 8. Raw trends by floor status. Mean log total Medicaid per capita, log HCBS per capita, log institutional LTSS per capita, and HCBS share of LTSS dollars over fiscal years FY2008 to FY2023, separately for floor states FMAP = 0.50 and non-floor states FMAP greater than 0.50

Note: This figure shows raw trends for the raw trends by floor status. It helps readers compare baseline levels, pre-policy movement, and the timing of any post-policy divergence.

Figure 8. Raw trends by floor status. Mean log total Medicaid per capita, log HCBS per capita, log institutional LTSS per capita, and HCBS share of LTSS dollars across FY2008–FY2023, separately for floor states (FMAP = 0.50) and non-floor states (FMAP > 0.50). The figure is descriptive and does not identify the RKD parameter; it is included so that the reader can visualize the underlying spending trajectories the RKD is partialling out.

C. Data sources with verification URLs

Source	Files used	Vintage / date	URL	Role in paper
MACPAC MACStats Exhibit 6	FMAP tables for FY2008–FY2025	MACStats editions March 2011 through December 2024	https://www.macpac.gov/publication/federal-medical-assistance-percentages-fmaps-and-enhanced-fmaps-e-fmaps-by-state-selected-periods/	Primary statutory FMAP source
HHS Federal Register FMAP notices	FY2014, 2015, 2019, 2020, 2022, 2024, 2025, 2026	Published 2012–2024	https://www.govinfo.gov/app/collection/fr	Independent cross-validation of MACPAC panel
BEA Regional Economic Accounts (SAINC1)	State and U.S. PCPI, calendar years 1929–2024	November 2024 release	https://www.bea.gov/data/income-saving/personal-income-by-state	Running variable construction
CMS/Mathematica Medicaid LTSS Annual Expenditures Reports	FY2017–FY2020 harmonized vintage	Murray et al. 2021 (Mathematica)	https://www.medicaid.gov/medicaid/long-term-services-supports/downloads/ltss-expenditures-2017-2018.pdf	HCBS and institutional LTSS outcomes
CMS TAF LTSS Users and Expenditures briefs	FY2019–FY2023 state-year aggregates	2020–2024	https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/taf/index.html	HCBS user share outcome
Census NST-EST state population	2000–2023	December 2023 release	https://www.census.gov/programs-surveys/popest.html	Per-capita outcome denominator
KFF State Health Facts	Medicaid enrollment indicators	Rolling updates	https://www.kff.org/state-category/medicaid-chip/medicaid-enrollment/	Cross-validation and per-enrollee outcome
MACPAC LTSS methodological review	Comparative review of LTSS measurement	August 2024	https://www.macpac.gov/publication/	Mathematica vintage harmonization reference
Mitchell (2025) CRS R43847	FMAP formula and floor-state list	April 2, 2025	https://www.congress.gov/crs-product/R43847	Institutional reference on the FMAP formula

Notes: This table documents the source files, scripts, variables, or data inputs used in the analysis. It is included to make the construction of the analytic evidence reproducible.

For complete file lists, download dates, and parser notes see `data/raw/SOURCES.md`.

D. Classification notes

D.1 Mathematica LTSS vintage handling

Cross-reference. This appendix subsection expands on main-text §3.2.1 and Limitation 3 of §6.6, both of which note that the HCBS spending point estimate collapses from 1.18 to 0.01 when six pre-ACA years are added via the Option B extended-window specification check (Section E below). The post-2017 Mathematica vintage harmonization is one of two plausible explanations for that collapse, and the design does not distinguish it from genuine time-heterogeneity in the HCBS response.

The Medicaid LTSS Annual Expenditures Reports series was originally produced by Steve Eiken and colleagues at Truven/IBM Watson Health (covering roughly through FY2016) and is currently produced by Murray, Tourtellotte, Lipson, and Wysocki at Mathematica (FY2017 onward). Post-2017 Mathematica vintages back-cast a harmonized HCBS subcategory allocation that conflicts with contemporaneous pre-2017 vintages; the harmonization affects how 1915(c) waiver, 1915(i) state plan HCBS, and 1915(k) Community First Choice expenditures are allocated within and across the HCBS bucket. My baseline panel uses the **latest-available harmonized Mathematica vintage** for every state-year where it is machine-parseable. A contemporaneous-vintage robustness panel is available in `data/clean/ltss_spending_panel_long.csv` but is not exploited in the headline robustness runs.

The **FY2015 and FY2016 Mathematica vintages are available only as PDF tabulations** and could not be machine-parsed without introducing classification errors. These two state-years are dropped from every LTSS outcome (leaving 713 state-years with LTSS outcomes out of 816 in the FMAP panel) and are retained only for the FMAP formula reconstruction (main text §5.1).

D.2 1915(k) Community First Choice treatment

CFC was created by ACA §2401 in 2011 and carries an additional 6 percentage points of federal match on top of the statutory FMAP. CFC adoption could in principle be endogenous to the statutory FMAP running variable because floor states with CFC receive a higher *effective* match than non-CFC floor states. The baseline HCBS bucket **includes** all CFC lines. Table 8 reports a robustness specification that **excludes** CFC from the HCBS bucket; the HCBS-share point estimate moves from -0.19 to -0.97 but both are within one standard error of zero and of each other. The CFC-excluded specification is additionally restricted to FY2011–FY2020 because the FY2021–FY2023 CFC lines are not broken out in the FY2017–FY2020 harmonized Mathematica vintage at the state-year level I require.

D.3 Enhanced-FMAP episode flag construction

Three indicator flags mark enhanced-FMAP episodes. `arra_flag` is 1 for state-years that overlap with the ARRA §5001 bump (FY2009 Q4 – FY2011 Q2, conservatively flagged as all of FY2009, FY2010, and FY2011 at the fiscal-year frequency). `aca_expansion_flag` is 1 for state-years in which the state was operating under the ACA Medicaid expansion's 90 percent enhanced match for newly eligible adults (from the state's effective expansion date, rounded to the containing fiscal year). Partial-exposure states (Michigan — April 2014; New Hampshire — August 2014) are flagged as fully affected in FY2014. `ffcra_flag` is 1 for state-years in which the FFCRA §6008 6.2-percentage-point bump applied (FY2020 through FY2023, conservatively flagged; the CAA-2023 phase-down through FY2023 Q3 is not separately split). The baseline panel uses the **statutory** FMAP irrespective of these flags and drops each episode only as a robustness specification.

E. Option B extended-window specification check

As a specification check, I extended the panel back to FY2002 (adding 306 state-years). The extended panel was constructed via the same `compute_running_variable()` function and 3-year-lag convention as the baseline. Full details in `analysis/option_b_extended_window.md`.

Coverage and CI-width results. Extending the panel added 306 state-years to four of five outcomes (the TAF HCBS user share is available only for FY2019–FY2023 and is unaffected). CI width ratios (Option B ÷ baseline) were 1.03, 0.92, 0.72, 0.71, and 1.00 across the five outcomes — only 2 of 5 hit the >=30% shrinkage threshold I set for promoting Option B to the main specification, and the headline log(total Medicaid / pop) outcome *widened*.

E.1 Table E.1 — ACA-drop persistence under the extended window

Holding the ACA-expansion drop fixed, I re-estimate the non-expansion tau on the FY2002–FY2023 panel and compare to the baseline FY2008–FY2023 result. Specification: drop all state-years where `aca_expansion_flag == 1`.

Outcome	Baseline drop-ACA (FY2008–FY2023)	Extended drop-ACA (FY2002–FY2023)
log(total Medicaid $ / pop)	tau = −2.526 [−4.260, −0.792] (N = 417)	tau = −2.603 [−4.257, −0.948] (N = 606)
log(HCBS $ / pop)	tau = −3.018 [−5.356, −0.681] (N = 382)	tau = −1.742 [−5.134, 1.649] (N = 620)
log(institutional $ / pop)	tau = 2.336 [0.198, 4.475] (N = 382)	tau = 2.204 [0.696, 3.713] (N = 635)
HCBS share of LTSS	tau = −1.191 [−2.042, −0.340] (N = 382)	tau = −0.836 [−1.651, −0.020] (N = 635)
HCBS user share (TAF FY19–23)	N too small after drop	N too small after drop

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

The headline log(total Medicaid / pop) non-expansion tau is essentially identical on the larger panel (−2.60 vs −2.53); the HCBS-share non-expansion tau remains negative and statistically distinguishable from zero (−0.84 vs −1.19); the institutional LTSS non-expansion tau remains positive and statistically distinguishable from zero (+2.20 vs +2.34). **The sign flip is structural — it reflects a difference in the underlying state-year response pattern between expansion and non-expansion states, not a sample-size artifact of the post-2014 window.**

New concerning finding (vintage-sensitivity). The HCBS spending point estimate (without the ACA drop) collapses from 1.18 (baseline window) to 0.01 (Option B window) when the six pre-ACA years FY2002–FY2007 are added to the panel — consistent with either (i) time-heterogeneity in the HCBS response at the FMAP kink (a structural difference in how HCBS responded to federal generosity pre- and post-ACA) or (ii) pre-2017 Mathematica vintage classification drift from the post-2017 harmonization back-casting HCBS subcategories inconsistently for the pre-ACA years. **The design cannot distinguish these two explanations.** This observation is the subject of main-text Limitation 3 (§6.6) and is cross-referenced from §3.2.1 of the main text. A reader who builds an independent HCBS panel from contemporaneous pre-2017 vintages may recover a different decomposition coefficient.

Decision. Option B was **not promoted** to the main specification. The baseline FY2008–FY2023 window remains the paper's main sample; the extended-window check functions only as a structural-vs-sample-size test for the ACA sign flip (main-text §5.4.1), and the HCBS-collapse finding is promoted to its own limitation (main-text Limitation 3 in §6.6) per the Path B reframe. The full Option B memo with construction notes and coverage tables is at `analysis/option_b_extended_window.md` in the replication package.