Forbidden Comparisons in Fixed-Effect Poisson Difference-in-Differences

CausalSmith*

August 25, 2026

Abstract

This paper studies the population treatment coefficient from pooling a staggered-adoption panel in a unit-and-time fixed-effect Poisson pseudo-likelihood. We show that this projection coefficient can be negative even when every cohort-time proportional treatment effect is positive. The failure reflects heterogeneous effects interacting with the fixed-effect PPML score: a four-cohort design with equal shares and flat untreated means provides an explicit sign reversal, while a positive-weight proportional treatment-on-the-treated target remains positive. The result characterizes the interpretation of the deterministic population projection targeted by the pooled criterion.

1 Introduction

Poisson pseudo-maximum likelihood with high-dimensional fixed effects is a standard way to estimate multiplicative mean models. It is especially common in gravity and other settings where the conditional mean is naturally proportional, zeros are empirically important, and fixed effects absorb large parts of the untreated outcome structure [Santos Silva and Tenreyro, 2006, 2011, Correia et al., 2020]. At the same time, many policy applications have staggered adoption: some units are treated early, others later, and some never. In linear difference-in-differences, it is now well understood that a pooled two-way fixed-effect coefficient can fail to aggregate heterogeneous effects with interpretable positive weights [Goodman-Bacon, 2021, de Chaisemartin and D'Haultfoeuille, 2020, Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Borusyak et al., 2024]. This paper studies the corresponding issue for a different object: the treatment coordinate of a pooled fixed-effect PPML population projection when treatment effects are proportional rather than additive.

The object of interest is deliberately narrow. Consider a finite cohort-time staggered-adoption design with positive limiting cohort shares, an exponential untreated mean with unit and time components, within-cohort baseline limits, proportional cohort-time treatment effects, and a full-rank collapsed fixed-effect design. The paper defines $\beta^\star(\delta)$ in Definition 3 as the treatment coordinate of the limiting collapsed Poisson pseudo-likelihood projection. This is the population analogue of the coefficient an applied pooled unit-and-time fixed-effect PPML

*Machine-generated by the CausalSmith research pipeline. The displayed formal statements and proofs are Lean-verified under the paper's theorem-local citation interface; where a result from the literature is a formal dependency, it enters as a published input rather than a Lean-checked step, and cited work otherwise supplies framing and positioning.

regression is trying to summarize. Its projection definition permits signed comparison weights and distinguishes it from causal average treatment effects built as convex aggregates of granular proportional effects.

The central question is whether the sign of this pooled PPML coefficient must agree with the signs of the underlying proportional effects. Sign reversal is possible. For any supported design satisfying the horizon, no-first-period-adopter, rank, and maintained population restrictions and whose cohort support contains $\{2, 3, 4, \infty\}$, Theorem 3 gives a primitive sign index Φ. Its sign is exactly the sign of $\beta^\star(\delta)$. The theorem's closed-form polynomial calculation is a separate specialization to the exact four-cohort support $\{2, 3, 4, \infty\}$. The general sign equivalence also shows that $\Phi < 0$ implies

$$\beta^\star(\delta) < 0 < PTT,$$

where PTT is the counterfactual-share proportional treatment-on-the-treated target. Thus the pooled fixed-effect PPML coefficient can have the opposite sign from every granular proportional effect while a positive-weight proportional target remains positive.

The mechanism is local as well as global. Theorem 1 shows that, under collapsed rank, the derivative of $\beta^\star(\delta)$ with respect to a treated cell's log multiplier has the sign of the treatment indicator residual after projecting treatment on the fixed effects with weights $q_{gt}\mu^\star_{gt}(\delta)$. These weights are generated by the fitted PPML mean through the pseudo-likelihood score and its curvature, producing a comparison geometry distinct from linear least squares. A treated cell with a negative weighted fixed-effect residual pushes the pooled coefficient downward when its proportional effect is increased. This is the multiplicative PPML analogue of a forbidden-comparison concern.

The paper also records the benchmark in which the pooled coefficient is well behaved. Under the same exponential untreated mean and collapsed rank conditions, if every treated cohort-time cell has the same log proportional effect δ_0, Theorem 2 shows that $\beta^\star(\delta) = \delta_0$ and that the fitted collapsed mean equals the true collapsed mean in every cell. This calibration locates the sign problem in the interaction between heterogeneous proportional effects and their compression into one pooled treatment coordinate in a staggered-adoption design.

The explicit witness is intentionally simple. Definition 6 considers four periods and four cohorts, $\{2, 3, 4, \infty\}$, with equal cohort shares, flat untreated means, and strictly positive treated-cell log effects. All treated cells have multiplier $101/100$ except cell $(2, 4)$, which has multiplier 4. Theorem 4 shows that this primitive belongs to the sign-reversal region in Definition 5. The special four-cohort calculation in Theorem 3 gives

$$\Phi = \frac{5x^2 + 12x - 2y - 15}{16},$$

and the witness has $\Phi = -11559/32000$. Hence the negative coefficient occurs with equal shares, strictly positive effects, and flat untreated time components.

The distinction between the pooled coefficient and a positive proportional target is important for interpretation. Counterfactual-share proportional treatment-on-the-treated targets, such as those studied by Moreau-Kastler [2025], aggregate granular proportional effects using positive exposure-based weights. Under strictly positive granular effects, those targets preserve the positive sign. The pooled fixed-effect PPML coefficient studied here is different: it is a pseudo-true projection coefficient in the sense of misspecified likelihood theory [White, 1982, Gouriéroux et al., 1984b,a]. Its sign is governed by the fixed-effect fit and the induced score equation, not by positive aggregation weights.

The contribution is a sharp population diagnostic. It characterizes the deterministic projection target, derives its local and global sign formulas, and proves that under stated staggered-adoption and multiplicative-mean restrictions its sign can oppose every underlying proportional effect. These results identify the interpretation problem that granular-effect estimation and sampling inference must address in subsequent work.

The rest of the paper proceeds as follows. The next section relates the result to linear and nonlinear DiD, proportional treatment-on-the-treated targets, misspecified likelihood, and PPML applications. The following section defines the finite-array and collapsed population projections and the sign-reversal region. A later section gives the local derivative result, the homogeneous benchmark, the primitive sign frontier, and the four-cohort witness. The final section discusses interpretation and extensions. The appendix contains the collapse lemma, the pseudo-true projection lemma, and proofs of the main results.

2 Related Literature

This paper contributes to the econometric literature on difference-in-differences by studying a particular population object: the treatment coordinate of a pooled fixed-effect Poisson pseudo-likelihood in a staggered-adoption design. The classical program-evaluation literature established difference-in-differences as a comparison of changes across treated and untreated groups [Ashenfelter and Card, 1985, Angrist and Pischke, 2009, Imbens and Wooldridge, 2009]. Semiparametric and nonlinear extensions clarify how identifying restrictions and target parameters change beyond additive linear conditional-mean models [Abadie, 2005, Athey and Imbens, 2006, Sant'Anna and Zhao, 2020]. The present paper studies the projection target generated by a deliberately pooled multiplicative regression and asks when that common coefficient can have the wrong sign even though all cohort-time proportional effects are positive.

A large recent literature shows that two-way fixed-effect linear regressions in staggered-adoption designs can aggregate heterogeneous treatment effects through signed or hard-to-interpret weights. Goodman-Bacon [2021] decomposes the linear two-way fixed-effect coefficient into two-by-two comparisons and highlights the role of comparisons between already-treated and newly treated groups. de Chaisemartin and D'Haultfoeuille [2020], Callaway and Sant'Anna [2021], Sun and Abraham [2021], and Borusyak et al. [2024] develop alternative estimands and estimators that use interpretable comparisons under suitable identifying restrictions. Those results are primarily about additive treatment effects and linear least-squares projections. Here the regression is Poisson pseudo-maximum likelihood with unit and time fixed effects, and treatment effects are proportional. Pooling across cohorts and periods therefore operates through a different mechanism: the pseudo-true Poisson coefficient responds to the fixed-effect fit and the induced comparison weights in the multiplicative mean model.

The analysis also relates to nonlinear difference-in-differences. Wooldridge [2023] specifies nonlinear conditional-mean models whose parameters or average partial effects are given a causal interpretation under maintained identifying restrictions. Roth and Sant'Anna [2020] and Roth et al. [2023] emphasize how identifying assumptions, functional-form restrictions, and target parameters must be separated. The present paper holds the multiplicative untreated-mean restriction fixed and asks what sign is carried by the common treatment coordinate when heterogeneous cohort-time multipliers are pooled in a misspecified fixed-effect PPML criterion. Its object is the pseudo-true projection coefficient $\beta^\star(\delta)$, which is distinct from a nonlinear-DiD effect parameter selected for causal interpretability.

The closest proportional-effect comparison is the counterfactual-share treatment-on-the-treated target studied by Moreau-Kastler [2025]. The objects differ in four steps. Its primitive effects are granular proportional effects; its target weights those effects by untreated counterfactual exposure; its sign is positive whenever every granular effect is positive; and its analysis is organized around identifying that causal aggregate. Here the same granular effects enter observed cohort-time means, but the target is the treatment coordinate solving the pooled fixed-effect PPML score equations. Those score equations induce fitted-mean curvature weights and residualized comparisons in place of counterfactual shares. Consequently, the positive proportional PTT and the potentially sign-reversing $\beta^\star(\delta)$ coexist as distinct summaries, and the sign-reversal result requires the pooled-score characterization developed here.

Methodologically, the paper builds on the theory of misspecified maximum likelihood and pseudo-likelihood. White [1982] formalizes pseudo-true parameters under misspecification, and Gouriéroux et al. [1984b,a] develop the pseudo-maximum-likelihood properties of Poisson estimators. Andrews [1988] provides general laws of large numbers for dependent and heterogeneous settings that underlie many asymptotic arguments. The formal results here characterize the deterministic population target and its causal interpretation, separating that question from convergence of an empirical estimator to its pseudo-true target.

Fixed-effect Poisson estimation is standard in applied microeconometrics and count-data models [Hausman et al., 1984, Cameron and Trivedi, 2013]. In international trade, PPML with high-dimensional fixed effects is a central empirical tool because it accommodates zeros and multiplicative gravity equations [Santos Silva and Tenreyro, 2006, 2011, Correia et al., 2020]. Gravity applications often combine multiplicative conditional means, rich fixed effects, and policy variation across country pairs or time [Anderson and van Wincoop, 2003, Head and Mayer, 2014, Yotov et al., 2016, Nagengast and Yotov, 2025]. Those features make PPML attractive, but they also make the pooled treatment coefficient easy to overinterpret when treatment timing is staggered and proportional effects vary across treated cells.

Recent work continues to refine fixed-effect and staggered-adoption estimators, including imputation and two-stage approaches for heterogeneous treatment effects [Gardner, 2022, Baker et al., 2025]. This paper complements that literature with a population characterization of when the coefficient produced by the familiar pooled fixed-effect PPML regression is negative despite strictly positive proportional effects. The result isolates the projection itself as the source of an unsafe causal summary and provides a diagnostic foundation for subsequent estimator and inference work.

3 Setup and Assumptions

Consider a finite-N panel with units $i = 1, \dots, N$ observed over calendar periods $t = 1, \dots, T$. Each unit belongs to an adoption cohort G_i in the finite support $\mathcal{C}$; the never-treated cohort is denoted by ∞. Treatment is absorbing: D_{gt} indicates whether cohort g is treated in period t, and $D_{it} = D_{G_i t}$. Let $I_{gN} = \{i : G_i = g\}$ and $n_{gN} = |I_{gN}|$. Expectations under the triangular-array population law are written $E_{\mathcal{P}_N}$.

The first restrictions keep the large-array design stable and impose the multiplicative untreated mean structure. Potential outcomes are $Y_{it}(0)$ and $Y_{it}(1)$. The untreated mean has a positive unit baseline b_{iN} and a calendar component γ_{t0}; $\bar{b}_{gN}$ is the within-cohort average baseline and $\bar{b}_g$ its limit.

Notation	Economic object
D_{gt}	treatment status of cohort g in period t
$B_{gt} = \bar{b}_g \exp(\gamma_{t0})$	untreated cohort-time mean
δ_{gt}	treated-cell log proportional effect
$m_{gt}(\delta) = B_{gt} \exp(D_{gt}\delta_{gt})$	observed cohort-time mean
$\beta^{\star}_N(\delta)$	finite-array population PPML treatment coordinate
$\beta^{\star}(\delta)$	limiting collapsed population PPML treatment coordinate
$\widetilde{W}_{gt}(\delta)$	fitted-mean-weighted residualized treatment
$\Phi(\delta)$	primitive index with the sign of $\beta^{\star}(\delta)$

Assumption 1. [`ass:cohort-share-limit`] (Cohort share limits) *For every $N \geq |\mathcal{C}|$ and every $g \in \mathcal{C}$,*

$$n_{gN} > 0 \quad \textit{and} \quad \frac{n_{gN}}{N} \to \pi_g \in (0,1).$$

Assumption 1 is the positive limiting cohort shares condition: every cohort remains represented asymptotically and contributes positive mass to the collapsed comparison [de Chaisemartin and D'Haultfoeuille, 2020]. The limit share is denoted π_g.

Assumption 2. [`ass:unit-untreated-exponential-mean`] (Untreated exponential mean) *For every unit $i = 1, \ldots, N$ and period $t = 1, \ldots, T$,*

$$E_{\mathcal{P}_N}[Y_{it}(0)] = b_{iN} \exp(\gamma_{t0}),$$

with the normalization $\gamma_{10} = 0$.

Assumption 2 is a unit-fixed-effect multiplicative parallel-trends restriction in an exponential mean model [Wooldridge, 2023]. The normalization fixes the location of the time effects and has no substantive content.

Assumption 3. [`ass:within-cohort-baseline-limit`] (Within-cohort baseline limits) *For every $g \in \mathcal{C}$, the deterministic triangular-array average*

$$\bar{b}_{gN} = n_{gN}^{-1} \sum_{i \in I_{gN}} b_{iN}$$

satisfies

$$\bar{b}_{gN} \to \bar{b}_g \in (0, \infty).$$

Assumption 3 is specific to this analysis. It is the condition that makes the cohort-time collapse deterministic: after averaging within a cohort, the only limiting untreated heterogeneity that remains is the cohort baseline $\bar{b}_g$.

Assumption 4. [`ass:proportional-effects`] (Proportional treatment effects) *For every unit $i = 1, \ldots, N$ and period $t = 1, \ldots, T$ with $D_{it} = 1$,*

$$E_{\mathcal{P}_N}[Y_{it}(1)] = E_{\mathcal{P}_N}[Y_{it}(0)] \exp(\delta_{G_i t}).$$

Assumption 4 is the cohort-time proportional treatment effects condition [Moreau-Kastler, 2025]. The log multiplier $\delta_{G_i t}$ may vary across treated cohorts and calendar periods, so the pooled coefficient below is not imposed to equal a common structural effect.

These assumptions induce a collapsed cohort-by-time representation. Define the untreated cohort-time mean $B_{gt} = \bar{b}_g \exp(\gamma_{t0})$, the collapsed cell mass $q_{gt} = \pi_g/T$, and the observed collapsed mean $m_{gt}(\delta) = B_{gt} \exp(D_{gt}\delta_{gt})$, where δ_{gt} denotes the cohort-time log multiplier on treated cells. Let r_{gt} collect the collapsed fixed-effect regressors and the treatment indicator, and let X_{gt} be the same vector with the treatment coordinate removed.

Assumption 5. [`ass:collapsed-design-rank`] (Collapsed design rank) *The collapsed second-moment matrix*

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} r_{gt} r_{gt}'$$

is positive definite.

Assumption 5 is the full-rank Poisson fixed-effect design condition [White, 1982]. It rules out exact collinearity in the collapsed regressors, so the population projection has a well-defined treatment coordinate rather than an unidentified linear combination of fixed effects and treatment.

The population object is a Poisson pseudo-likelihood projection defined through the conditional-mean criterion. For a finite array, let v_{itN} be the unit-time regressor vector containing unit effects, time effects, and treatment; let θ_N be the corresponding coefficient vector; and let $m_{itN}(\delta)$ be the observed unit-time mean generated by the untreated mean and proportional effects.

Definition 1. [`def:unit-fe-population-projection`] (Unit-FE Projection $\beta_N^\star(\delta)$) *Define $\beta_N^\star(\delta)$ as the last coordinate of a maximizer over $\theta_N \in \mathbb{R}^{N+T}$ of*

$$L_N(\theta_N; \delta) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ m_{itN}(\delta)\, v_{itN}' \theta_N - \exp(v_{itN}' \theta_N) \right\}.$$

Definition 1 is the finite-array population analogue of the unit fixed-effect PPML regression. The criterion $L_N(\theta_N; \delta)$ is evaluated at means rather than realized outcomes, and $\beta_N^\star(\delta)$ is the treatment coordinate of that population projection.

Definition 2. [`def:sample-unit-fe-ppml`] (Sample PPML Coefficient $\hat{\beta}_N$) *Let A_N be the set of admissible extended maximizers of*

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ Y_{it}\, v_{itN}' \theta_N - \exp(v_{itN}' \theta_N) \right\}$$

over unit-effect coordinates in $\mathbb{R} \cup \{-\infty\}$, time-effect coordinates in $\mathbb{R}$ satisfying the period-1 normalization, and $\beta \in \mathbb{R}$, with the convention $\exp(-\infty) = 0$. Admissibility means that the period-1 time effect is normalized to zero and that a unit-effect coordinate may equal $-\infty$ only for a unit whose whole observed outcome path satisfies $Y_{it} = 0$ for every t; for such a unit the extended objective contribution is defined to be zero. If A_N contains an element whose squared Euclidean norm over the finite time-effect and β coordinates is no larger than that of every element of A_N, then $\hat{\beta}_N$ is the β-coordinate of such a selected element. If no such minimum-finite-coordinate-norm admissible extended maximizer exists, set $\hat{\beta}_N = 0$.

The limiting projection works directly with the collapsed cohort-time cells. With θ denoting the collapsed coefficient vector, write the limiting collapsed Poisson criterion as $L(\theta;\delta)$. The finite collapsed analogue, used only to connect the array to this limit, is denoted $L_{Nc}(\theta;\delta)$.

Definition 3. [`def:collapsed-population-projection`] (Collapsed Projection $\theta^\star(\delta)$) *Define*

$$\theta^\star(\delta) \in \arg\max_{\theta\in\mathbb{R}^{|\mathcal{C}|+T}} L(\theta;\delta),$$

where

$$L(\theta;\delta) = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\left\{m_{gt}(\delta)\, r_{gt}'\theta - \exp(r_{gt}'\theta)\right\}.$$

The coefficient $\beta^\star(\delta)$ is the last coordinate of $\theta^\star(\delta)$, and the fitted collapsed mean is

$$\mu_{gt}^\star(\delta) = \exp\{r_{gt}'\theta^\star(\delta)\}.$$

Definition 3 gives the main object studied in the paper. The maximizer $\theta^\star(\delta)$, its treatment coordinate $\beta^\star(\delta)$, and the fitted mean $\mu_{gt}^\star(\delta)$ are pseudo-true quantities in the sense of misspecified likelihood theory [White, 1982, Gouriéroux et al., 1984b,a]. The target is therefore a deterministic population projection defined by the Poisson mean criterion.

The analysis centers on this deterministic population projection. A sample fixed-effect PPML regression replaces the means by observed outcomes and may also confront separation [Santos Silva and Tenreyro, 2006, 2011, Correia et al., 2020]; its sampling theory is a distinct extension. Accordingly, references below to a "pooled coefficient" mean the population coordinate $\beta^\star(\delta)$, unless explicitly stated otherwise.

The sign analysis uses the residualized treatment after partialling out the collapsed fixed effects with weights given by the fitted PPML mean. Let $\rho^\star(\delta)$ be the weighted least-squares coefficient from that auxiliary projection.

Definition 4. [`def:weighted-fwl-residual`] (Weighted Residual $\widetilde{W}_{gt}(\delta)$) *Define*

$$\rho^\star(\delta) \in \arg\min_{\rho\in\mathbb{R}^{|\mathcal{C}|+T-1}} \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\,\mu_{gt}^\star(\delta)\{D_{gt} - X_{gt}'\rho\}^2.$$

The weighted Frisch–Waugh–Lovell residualized treatment is

$$\widetilde{W}_{gt}(\delta) = D_{gt} - X_{gt}'\rho^\star(\delta).$$

Definition 4 gives the comparison weights that enter the local sign formula in the next section. The residual $\widetilde{W}_{gt}(\delta)$ is the fixed-effect projection residual of treatment under the curvature weights of the Poisson criterion and serves as a projection comparison weight.

The sign-reversal result is stated on designs with at least the four adoption statuses needed for the explicit construction. The next two assumptions isolate that scope and the positive-effect case of interest.

Assumption 6. [`ass:multicohort-frontier-scope`] (Multicohort frontier scope) *The cohort set satisfies*

$$\{2,3,4,\infty\} \subseteq \mathcal{C}.$$

Assumption 6 is specific to this analysis. It ensures that the design contains early, middle, late, and never-treated cohorts, the minimal structure used by the four-cohort construction.

Assumption 7. [ass:strict-positive-effects] (Strict Positive Effects) *For every cohort-period cell (g,t) with $D_{gt}=1$,*

$$\delta_{gt} > 0.$$

Assumption 7 is also specific to this analysis. It deliberately restricts attention to the strongest sign benchmark: every treated cell has a positive proportional effect, so any negative pooled coefficient is a property of the projection rather than a mixture of positive and negative underlying effects.

The global parameter region collects the maintained restrictions under which the limiting PPML treatment coordinate is negative despite those positive cell-level effects.

Definition 5. [def:global-sign-reversal-region] (Sign-Reversal Region $\mathcal{R}_T$) *Define $\mathcal{R}_T$ as the set of all triples*

$$\big((\pi_g)_{g\in\mathcal{C}}, (B_{gt})_{g\in\mathcal{C},\ 1\le t\le T}, (\delta_{gt})_{(g,t):D_{gt}=1}\big)$$

with $T \ge 4$, indexed by a cohort support $\mathcal{C}$ that contains $\{2,3,4,\infty\}$ and contains no first-period finite cohort, such that

$$0 < \pi_g < 1 \quad (g \in \mathcal{C}), \qquad \sum_{g\in\mathcal{C}} \pi_g = 1,$$

there exist $\bar{b}_g > 0$ for $g \in \mathcal{C}$ with $B_{gt} = \bar{b}_g \exp(\gamma_{t0})$ for every supported cell (g,t), $\delta_{gt} > 0$ for every treated supported cell (g,t) with $D_{gt}=1$, the collapsed design has full weighted rank in the sense that

$$\forall a \neq 0, \qquad 0 < \sum_{g\in\mathcal{C}} \sum_{t=1}^{T} \frac{\pi_g}{T} (r'_{gt} a)^2,$$

and

$$\beta^\star(\delta) < 0.$$

Definition 5 names the population configurations in which the pooled limiting coefficient has the opposite sign from every granular proportional treatment effect. The notation $\mathcal{R}_T$ is only a shorthand for this primitive set of shares, untreated means, and treated-cell multipliers.

The paper uses one explicit triangular array to show that this region is nonempty. Its role is evidentiary rather than calibrational: all baselines and untreated time effects are set to one, and the only large treated multiplier is placed in the period-4 cell of the cohort first treated in period 2.

Definition 6. [def:four-cohort-witness] (Four-cohort witness W_4) *Let W_4 be the cofinal triangular subsequence indexed by $N = 4m$, $m = 1,2,\ldots$, with $T = 4$ and*

$$\mathcal{C} = \{2,3,4,\infty\}.$$

For each such N,

$$n_{2N} = n_{3N} = n_{4N} = n_{\infty N} = m,$$

$b_{iN} = 1$ for every i, and $\gamma_{t0} = 0$ for every t. The treatment effects satisfy

$$\delta_{gt} = \log(101/100)$$

for every treated cell $(g,t) \neq (2,4)$, while

$$\delta_{2,4} = \log(4).$$

Definition 6 defines the witness W_4. Equal cohort shares and flat untreated means remove scale and composition as explanations for the sign reversal; the later result depends only on the staggered treatment pattern and the heterogeneous positive proportional effects encoded here.

Finally, the primitive sign characterization uses an algebraic index based on observed mean components. Let H denote the set of treated cohort-time cells when that notation is needed later.

Definition 7. [`def:frontier-elimination-handle`] (Frontier-elimination polynomial $\Phi(\delta)$) *Define*

$$\Phi(\delta) = MA - \sum_{g,t} D_{gt} R_g C_t,$$

where

$$h_{gt} = \pi_g B_{gt} \exp(D_{gt}\delta_{gt}), \qquad R_g = \sum_t h_{gt}, \qquad C_t = \sum_g h_{gt},$$

and

$$M = \sum_g R_g, \qquad A = \sum_{g,t} D_{gt} h_{gt}.$$

The relation between $\Phi(\delta)$ and $\beta^\star(\delta)$ is stated separately.

Definition 7 introduces the scalar $\Phi(\delta)$ used by the primitive sign result. The component h_{gt} is the cohort-share-weighted observed collapsed mean, equivalently $Tq_{gt}m_{gt}(\delta)$, while R_g, C_t, M, and A are its cohort sums, period sums, grand sum, and treated-cell sum. The main theorem in the next section supplies the sign relation for this scalar.

4 Main Results

The setup isolates a population object: the treatment coordinate of the collapsed fixed-effect PPML projection in Definition 3. The first result gives the local comparative static of that coefficient with respect to one treated cohort-time log multiplier. Its content is a nonlinear analogue of a forbidden-comparison diagnostic: after the Poisson fixed effects have been fit, the sign of the derivative is exactly the sign of the residualized treatment in that cell, where residualization uses the fitted PPML mean as the curvature weight.

Theorem 1. [`thm:sharp-ppml-forbidden-sign`] (Sharp PPML Sign) *Let T be positive and let $\mathcal{C}$ be a nonempty finite cohort set. Fix cohort shares $\pi_g \in (0,1)$, baseline levels $\bar{b}_g > 0$, time effects γ_t, and a proportional-effect array δ. Assume:*

- ***(Rank.)*** *The collapsed design satisfies Assumption 5.*
- ***(Cell.)*** *The cohort k belongs to $\mathcal{C}$, the period $s \in \{1, \dots, T\}$, and the cohort-period cell is treated: $D_{ks} = 1$.*

Write

$$\widetilde{W}_{ks}(\delta)$$

for the pseudo-true $q_{gt}\mu^\star_{gt}(\delta)$-weighted FWL residual of the treatment indicator in cell (k,s), and define

$$\mathcal{E}(\delta) = \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt}\, \mu^\star_{gt}(\delta)\, \widetilde{W}_{gt}(\delta)^2.$$

Then $\mathcal{E}(\delta) > 0$, and the one-coordinate path

$$x \mapsto \beta^\star\big(\delta \text{ with } \delta_{ks} \text{ replaced by } x\big)$$

is differentiable at $x = \delta_{ks}$ with derivative

$$\frac{q_{ks}\, B_{ks} \exp(\delta_{ks})\, \widetilde{W}_{ks}(\delta)}{\mathcal{E}(\delta)}.$$

Consequently,

$$\frac{\partial \beta^\star(\delta)}{\partial \delta_{ks}} < 0 \iff \widetilde{W}_{ks}(\delta) < 0, \qquad \frac{\partial \beta^\star(\delta)}{\partial \delta_{ks}} = 0 \iff \widetilde{W}_{ks}(\delta) = 0,$$

and

$$\frac{\partial \beta^\star(\delta)}{\partial \delta_{ks}} > 0 \iff \widetilde{W}_{ks}(\delta) > 0.$$

Assumption 5 makes the denominator $\mathcal{E}(\delta)$ strictly positive by preserving residual treatment variation after projecting on the fixed effects. Thus the local sign follows directly from the pseudo-true fit. A treated cell with a negative residual in Definition 4 pushes the pooled coefficient downward when its proportional effect is increased; a positive residual pushes it upward. This is the PPML counterpart to the comparison-weight concern in staggered-adoption linear regressions [Goodman-Bacon, 2021, de Chaisemartin and D'Haultfoeuille, 2020], with weights generated by the multiplicative pseudo-likelihood projection instead of least squares on the outcome.

The next result records the exact-calibration benchmark. When all treated cells have the same proportional log effect, the fixed-effect PPML projection recovers that common multiplier exactly.

Theorem 2. [`prop:homogeneous-effect-reduction`] (Homogeneous Effect Reduction) *Fix T, a finite cohort set $\mathcal{C}$, sampling laws $\{\mathcal{P}_N\}_N$, potential outcomes $Y_{it}(d)$, positive unit baselines b_{iN}, cohort shares $\pi_g \in (0,1)$, positive limiting cohort baselines $\bar{b}_g$, untreated time components γ_{t0}, cohort-time log effects δ_{gt}, and a scalar δ_0. Suppose that*

- ***(Untreated means.)*** *Assumption 2 holds.*
- ***(Collapsed rank.)*** *Assumption 5 holds for $\mathcal{C}$ and $\{\pi_g\}_g$.*
- ***(Common treated-cell effect.)*** *For every $g \in \mathcal{C}$ and every period t, if $D_{gt} = 1$, then $\delta_{gt} = \delta_0$.*

Then

$$\beta^\star(\delta) = \delta_0,$$

and, for every $g \in \mathcal{C}$ and every period t,

$$\mu^\star_{gt}(\delta) = B_{gt} \exp(D_{gt}\delta_0).$$

The homogeneous case is therefore an exact calibration point. If the proportional effect is common across treated cells, the treatment regressor completes the multiplicative mean specification after cohort and time effects are included, and the pooled coefficient has the expected

causal interpretation. Heterogeneous proportional effects generate the sign-reversal mechanism studied here, the setting in which modern staggered-adoption estimands use cohort-specific effects and interpretable aggregation [Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Borusyak et al., 2024].

The sharp local derivative identifies the cellwise mechanism. The next theorem complements it with a primitive scalar index whose sign is exactly the sign of the limiting pooled PPML coefficient under the maintained multicohort positive-effect scope. It also compares the pooled coefficient with the counterfactual-share proportional treatment-on-the-treated target, which remains positive when every granular proportional effect is positive [Moreau-Kastler, 2025].

Theorem 3. [thm:primitive-global-frontier] (Primitive Sign Frontier) *Fix a panel length T, a finite cohort support $\mathcal{C}$, triangular-array sampling laws $\mathcal{P}_N$, potential outcomes $Y_{it}(d)$, deterministic cohort labels G_i, positive unit baselines b_{iN}, limiting shares $\pi_g \in (0,1)$, positive limiting cohort baselines $\bar{b}_g$, untreated time effects γ_{t0}, and log proportional effects δ_{gt}. Suppose:*

- ***(Horizon and support.)*** *$T > 0$, $T \geq 4$, the never-treated cohort belongs to $\mathcal{C}$, and every finite supported cohort has adoption date different from the first period.*
- ***(Array support.)*** *For every N and every unit i, $G_i \in \mathcal{C}$.*
- ***(Share limits.)*** *Assumption 1 holds for $(\mathcal{C}, G_i, \pi_g)$.*
- ***(Untreated means.)*** *Assumption 2 holds for $(\mathcal{P}_N, Y_{it}(d), b_{iN}, \gamma_{t0})$.*
- ***(Baseline limits.)*** *Assumption 3 holds for $(\mathcal{C}, G_i, b_{iN}, \bar{b}_g)$.*
- ***(Proportional effects.)*** *Assumption 4 holds for $(\mathcal{P}_N, Y_{it}(d), G_i, \delta_{gt})$.*
- ***(Rank.)*** *Assumption 5 holds for $(\mathcal{C}, \pi_g)$.*
- ***(Frontier scope.)*** *Assumption 6 holds for $(T, \mathcal{C})$.*
- ***(Positive effects.)*** *Assumption 7 holds for $(\mathcal{C}, \delta_{gt})$.*

Define

$$h_{gt} = \pi_g \bar{b}_g \exp(\gamma_{t0}) \exp(D_{gt}\delta_{gt}), \qquad R_g = \sum_{t=1}^{T} h_{gt}, \qquad C_t = \sum_{g \in \mathcal{C}} h_{gt},$$

$$M = \sum_{g \in \mathcal{C}} R_g, \qquad A = \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} D_{gt} h_{gt}, \qquad \Phi = MA - \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} D_{gt} R_g C_t.$$

Then

$$\beta^\star(\delta) < 0 \iff \Phi < 0, \qquad \beta^\star(\delta) = 0 \iff \Phi = 0, \qquad 0 < \beta^\star(\delta) \iff 0 < \Phi.$$

Moreover, the primitive tuple induced by $(\pi_g, \bar{b}_g, \gamma_{t0}, \delta_{gt})$ belongs to $\mathcal{R}_T$ if and only if $\Phi < 0$.

For every $x > 1$ and $y > 1$, in the four-period support $\{2, 3, 4, \infty\}$ with equal shares, unit limiting baselines, zero untreated time effects, multiplier x on treated cells other than $(2, 4)$, and multiplier y at $(2, 4)$,

$$\Phi = \frac{5x^2 + 12x - 2y - 15}{16}.$$

Also,

$$\frac{5(101/100)^2 + 12(101/100) - 15}{2} = \frac{4441}{4000},$$

and the explicit witness W_4 has

$$\Phi = -\frac{11559}{32000}.$$

Let

$$m_{gt} = \bar{b}_g \exp(\gamma_{t0}) \exp(D_{gt}\delta_{gt}), \qquad H = \{(g,t) : g \in \mathcal{C},\ D_{gt} = 1\}.$$

For every $(g,t) \in H$,

$$B_{gt}^{obs} = B_{gt}, \qquad \tau_{gt} = \exp(\delta_{gt}) - 1, \qquad \omega_{gt} > 0.$$

The counterfactual-share weights sum to one:

$$\sum_{(g,t)\in H} \omega_{gt} = 1.$$

The counterfactual-share PTT satisfies

$$PTT = \frac{\sum_{(g,t)\in H} q_{gt} m_{gt}}{\sum_{(g,t)\in H} q_{gt} B_{gt}^{obs}} - 1.$$

Finally,

$$\Phi < 0 \quad \Longrightarrow \quad \beta^\star(\delta) < 0 < PTT.$$

Theorem 3 has two distinct scopes. The sign equivalence between Φ and the pooled coefficient applies to every support satisfying the theorem's assumptions, including supports that strictly contain $\{2,3,4,\infty\}$. The displayed polynomial in x and y, and the numerical value for W_4, specialize that general characterization to the exact equal-share four-cohort support. The index Φ serves as a population diagnostic conditional on known or specified primitives, while the counterfactual-share proportional PTT is the causal aggregate. In the same primitive configurations, the counterfactual-share weights are positive and sum to one, so the proportional PTT is a positive weighted average whenever all granular proportional effects are positive.

The explicit four-cohort construction makes the sign reversal concrete. The support contains cohorts first treated in periods 2, 3, and 4, together with a never-treated cohort; untreated means are flat and cohort shares are equal. Hence the example establishes the phenomenon under conventional baseline scaling and constant untreated time components.

Theorem 4. [`prop:four-cohort-sign-reversal`] (Four-Cohort Sign Reversal) *The four-cohort witness W_4 is constructed with cohort support $\mathcal{C} = \{2,3,4,\infty\}$, equal shares $\pi_g = 1/4$, limiting baselines $\bar{b}_g = 1$, untreated time component $\gamma_{t0} = 0$, and proportional-effect log multipliers*

$$\delta_{gt} = \begin{cases} \log 4, & (g,t) = (2,4), \\ \log(101/100), & D_{gt} = 1 \text{ and } (g,t) \neq (2,4), \\ 0, & D_{gt} = 0. \end{cases}$$

Then:

- ***(Region membership.)*** *The primitive W_4 belongs to the sign-reversal region $\mathcal{R}_4$ of Definition 5.*
- ***(Positive effects.)*** *Every treated supported cell has a strictly positive proportional-effect log multiplier: if $g \in \mathcal{C}$, $t \in \{1,2,3,4\}$, and $D_{gt} = 1$, then $\delta_{gt} > 0$.*
- ***(Unique largest treated effect.)*** *The treated cell $(2,4)$ has the strictly largest proportional-effect log multiplier: if $g \in \mathcal{C}$, $t \in \{1,2,3,4\}$, $D_{gt} = 1$, and $(g,t) \neq (2,4)$, then $\delta_{gt} < \delta_{2,4}$.*
- ***(Weighted FWL residuals.)*** *At the no-effect vector, the weighted FWL residual $\widetilde{W}_{gt}$ of Definition 4 equals*

$$\frac{1}{8}\begin{pmatrix} -3 & 3 & 1 & -1 \\ -1 & -3 & 3 & 1 \\ 1 & -1 & -3 & 3 \\ 3 & 1 & -1 & -3 \end{pmatrix},$$

  with rows $g = 2,3,4,\infty$ and columns $t = 1,2,3,4$. In particular, $\widetilde{W}_{2,4} = -1/8$.
- ***(Negative late-cell derivative at zero.)*** *Holding all other effects at zero,*

$$\frac{d}{dx}\,\beta^\star(\delta_{2,4} = x,\ \delta_{gt} = 0 \text{ for } (g,t) \neq (2,4))\Big|_{x=0} = -\frac{1}{10}.$$

- ***(Local negative derivative under positive effects.)*** *There exists $\varepsilon > 0$ such that, for every effect vector δ' satisfying $|\delta'_{gt}| < \varepsilon$ for all $g \in \mathcal{C}$ and $t \in \{1,2,3,4\}$, if δ' has strictly positive treated effects in the sense of Assumption 7, then*

$$\frac{d}{dx}\,\beta^\star\big(\delta' \text{ with } \delta'_{2,4} \text{ replaced by } x\big)\Big|_{x=\delta'_{2,4}} < 0.$$

The residual table in Theorem 4 identifies the period-4 cell $(2,4)$ of the early-treated cohort as locally negative at the no-effect vector, and the neighborhood statement extends that derivative sign to small strictly positive effects. The actual witness W_4, however, is a global sign-reversal example because Theorem 3 gives $\Phi < 0$ for its primitive tuple and Theorem 4 places that tuple in $\mathcal{R}_4$. In this example every treated cell has a positive proportional effect, and the largest effect is in that late calendar cell, yet the limiting pooled PPML coefficient is negative.

The main implication concerns interpretation under heterogeneous proportional effects: the single pooled coefficient is a pseudo-true projection whose sign is governed by fixed-effect residual comparisons. Interpretable causal summaries should therefore be built from granular proportional effects and positive aggregation weights, as in counterfactual-share targets, in place of inference from the sign of the pooled fixed-effect PPML coefficient alone [Moreau-Kastler, 2025, Wooldridge, 2023, Roth and Sant'Anna, 2020].

5 Discussion and Extensions

The results are motivated by empirical settings where the multiplicative conditional mean is substantively natural and the treatment timing is staggered. Gravity applications are a leading example: PPML specifications with high-dimensional fixed effects are standard because

trade flows are modeled through multiplicative resistance terms, exposure variables, and policy shifters, and because zeros are common in the outcome [Anderson and van Wincoop, 2003, Head and Mayer, 2014, Yotov et al., 2016, Santos Silva and Tenreyro, 2006, 2011, Correia et al., 2020]. In such designs, a pooled treatment indicator may look like a compact summary of a policy change. The results above show that, under heterogeneous proportional effects, the population coefficient of that regression is instead a fixed-effect Poisson projection whose sign can be governed by residualized comparisons across cohort-time cells.

The conclusion pinpoints how PPML summarizes a multiplicative mean model. The homogeneous benchmark in Theorem 2 shows that a common proportional effect is recovered exactly. With heterogeneous cohort-time proportional effects, Theorem 1 shows that increasing a positive treated-cell effect moves the pooled coefficient in the direction determined by the weighted fixed-effect residual for that cell. The interpretation problem therefore arises from compressing a collection of heterogeneous economic effects into a single pooled projection coordinate.

5.1 Limitations and future work

The primitive index Φ in Theorem 3 is a population diagnostic conditional on known or specified primitives. It determines whether the limiting fixed-effect PPML treatment coordinate is negative, zero, or positive. Estimation of those primitives, implementation of Φ as a specification test, and construction of positive-weight estimators from granular proportional effects are the main next steps.

That distinction is important for empirical interpretation. A negative pooled PPML coefficient is sometimes read as evidence that the treatment reduced the conditional mean. Under the conditions of Theorem 3, the same primitive configuration can instead have $\beta^\star(\delta) < 0$ and a positive counterfactual-share proportional treatment-on-the-treated target. The sign reversal therefore identifies the pooled fixed-effect PPML coefficient as a projection summary distinct from positive-weight proportional targets.

The positive counterfactual-share target remains closer to the causal object under the maintained proportional-effect restrictions because it aggregates treated-cell effects with weights tied to untreated counterfactual exposure [Moreau-Kastler, 2025]. When each granular proportional effect is positive, such an aggregation preserves the sign by construction. Theorem 3 formalizes this contrast in the same population environment: the pooled coefficient can be negative while PTT is positive. Empirical conclusions should therefore distinguish the regression projection from the proportional treatment-on-the-treated estimand.

The four-cohort witness in Theorem 4 establishes the sign reversal with equal cohort shares, flat untreated time effects, and positive treated-cell effects. Staggered timing and heterogeneous proportional effects place a large positive effect in a cell with a negative residualized treatment value. This makes the phenomenon relevant for conventional staggered-adoption designs, including policy settings where treatment intensity or exposure plausibly varies across treated cohorts and periods.

The population characterization also clarifies the choice that precedes estimation. A researcher may intend a common log multiplier, a collection of granular proportional effects, or a positive-weight proportional aggregate; the pooled projection equals the first under homogeneity and can differ sharply from the third under heterogeneity. The primitive index Φ establishes this distinction theoretically. Future work can turn that diagnostic into an operational procedure by estimating granular effects, constructing a sampling distribution for Φ, assessing separation, and determining how often the sign-reversal region occurs in empirically calibrated designs [San-

tos Silva and Tenreyro, 2006, 2011, Correia et al., 2020, Wooldridge, 2023, Nagengast and Yotov, 2025].

Appendices

A Proofs, Auxiliary Lemmas, and Verification Note

This appendix records the two auxiliary population results used to connect the unit-level fixed-effect regression to the collapsed cohort-time analysis. The ordering mirrors the logic of the main text. The collapse result first justifies replacing the finite-N unit fixed-effect projection by its cohort-time counterpart. The projection result then gives the first-order conditions for the limiting collapsed pseudo-true parameter. The main theorem proofs use these statements before establishing the primitive sign characterization in Theorem 3; the four-cohort sign-reversal result in Theorem 4 is then read as a consequence of that characterization together with the explicit witness construction.

The first auxiliary result formalizes the reduction from a unit fixed-effect PPML population criterion to a collapsed criterion. Its role is to show that, under the same cohort-share, untreated-mean, baseline-limit, proportional-effect, and rank restrictions used in Assumption 1, the treatment coordinate is invariant to within-cohort rearrangements of unit baselines that leave cohort averages unchanged. Thus the finite-array object in Definition 1 and the limiting object in Definition 3 are linked through the cohort-time means that drive the sign analysis.

Proposition 1. [`lem:unit-fe-collapse`] (Unit FE Collapse) *Let T be a number of periods and let $\mathcal{C}$ be a finite cohort set. Suppose:*

- ***(Horizon.)*** *$T \geq 4$.*
- ***(Cohort support.)*** *The never-treated cohort belongs to $\mathcal{C}$, and every finite cohort in $\mathcal{C}$ is dated after the first period.*
- ***(Array support.)*** *For every N and every unit i, the cohort label G_i belongs to $\mathcal{C}$.*
- ***(Positive primitives.)*** *The unit baselines b_{iN} and the limiting within-cohort baselines $\bar{b}_g$ are strictly positive, and each limiting cohort share satisfies $\pi_g \in (0,1)$.*
- ***(Share limits.)*** *Assumption 1 holds for G_i and π_g.*
- ***(Untreated means.)*** *Assumption 2 holds for $Y_{it}(d)$, $\mathcal{P}_N$, b_{iN}, and γ_{t0}.*
- ***(Baseline limits.)*** *Assumption 3 holds for G_i, b_{iN}, and $\bar{b}_g$.*
- ***(Proportional effects.)*** *Assumption 4 holds for $Y_{it}(d)$, $\mathcal{P}_N$, G_i, and δ_{gt}.*
- ***(Collapsed rank.)*** *Assumption 5 holds for $\mathcal{C}$ and π_g.*

Then, for every N with $|\mathcal{C}| \leq N$, the finite-array unit-FE population criterion $L_N(\theta_N;\delta)$ of Definition 1 has a unique global maximizer, the collapsed finite-array criterion $L_{Nc}(\theta;\delta)$ has a unique global maximizer, and

$$\beta_N^\star(\delta) = \text{the treatment coordinate of the unique maximizer of } L_{Nc}(\theta;\delta).$$

Moreover,

$$\beta_N^\star(\delta) \longrightarrow \beta^\star(\delta) \qquad \text{as } N \to \infty.$$

Finally, for every N with $|\mathcal{C}| \leq N$ and every alternative positive baseline array b'_{iN}, if

$$\bar{b}'_{gN} = \bar{b}_{gN} \qquad \text{for every } g \in \mathcal{C},$$

then replacing b_{iN} by b'_{iN} leaves the finite-array unit-FE coefficient unchanged:

$$\beta_N^\star(\delta; b') = \beta_N^\star(\delta; b).$$

The uniqueness clauses are population pseudo-likelihood statements in the usual misspecified-likelihood sense [White, 1982, Gouriéroux et al., 1984b,a]. The final invariance clause is specific to the cohort-time collapse used here: once unit effects absorb individual baseline levels, only the within-cohort baseline average enters the collapsed mean relevant for the treatment coordinate. This is why the later sign results can be stated in terms of cohort shares, untreated cohort-time means, and proportional-effect multipliers rather than the full list of unit baselines.

The second auxiliary result records the first-order conditions for the collapsed pseudo-true PPML projection. These score equations are the fixed-effect and treatment equations behind the derivative formula in Theorem 1 and the algebraic sign calculation in Theorem 3.

Lemma 1. [`lem:pseudo-true-ppml-projection`] (Unique PPML Projection) *Fix a number of periods T, a finite cohort set $\mathcal{C}$, cohort shares $\pi_g \in (0,1)$, positive limiting baseline means $\bar{b}_g > 0$, time components γ_{t0}, and proportional-effect log multipliers δ_{gt}. Suppose the collapsed design rank condition in Assumption 5 holds for $(T, \mathcal{C}, \pi)$. Let $\theta^\star(\delta)$, $L(\theta;\delta)$, $m_{gt}(\delta)$, and $\mu^\star_{gt}(\delta)$ be as in Definition 3. Then $\theta^\star(\delta)$ is the unique global maximizer of $L(\theta;\delta)$. Moreover, for every collapsed nuisance regressor index j,*

$$\sum_{g\in\mathcal{C}} \sum_{t=1}^{T} q_{gt} X_{gt,j} \{m_{gt}(\delta) - \mu^\star_{gt}(\delta)\} = 0,$$

and the treatment score also vanishes:

$$\sum_{g\in\mathcal{C}} \sum_{t=1}^{T} q_{gt} D_{gt} \{m_{gt}(\delta) - \mu^\star_{gt}(\delta)\} = 0.$$

Proof of Lemma 1. Throughout, a *supported cell* is a pair (g,t) with $g \in \mathcal{C}$ and $t \in \{1,\dots,T\}$, and a collapsed parameter is written $\theta = (a, \beta)$ with nuisance block a and treatment coordinate β, so that the collapsed linear index is

$$r'_{gt}\theta = \sum_j X_{gt,j} a_j + D_{gt}\beta.$$

The collapsed cell mass and observed cohort-time mean are

$$q_{gt} = \frac{\pi_g}{T}, \qquad m_{gt}(\delta) = B_{gt} \exp(D_{gt}\delta_{gt}), \qquad B_{gt} = \bar{b}_g \exp(\gamma_{t0}).$$

Definition 3 defines the limiting criterion $L(\theta;\delta)$, the collapsed projection $\theta^\star(\delta)$, and the fitted collapsed mean

$$\mu^\star_{gt}(\delta) = \exp\{r'_{gt}\theta^\star(\delta)\}.$$

Step 1: the supported-cell table is nonempty. Take the collapsed parameter $\theta^0 = (0, 1)$, whose nuisance block is zero and whose treatment coordinate is one, so $\theta^0 \neq 0$. If $T = 0$ or $\mathcal{C}$ is empty, then

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} \, (r'_{gt}\theta^0)^2 = 0,$$

because the sum is empty; this contradicts Assumption 5, which requires the displayed quantity to be strictly positive at every nonzero collapsed parameter. Hence $T > 0$ and $\mathcal{C}$ is nonempty, and at least one supported cell exists.

Step 2: masses and observed means are strictly positive. For every supported cell,

$$q_{gt} = \frac{\pi_g}{T} > 0, \qquad m_{gt}(\delta) = \bar{b}_g \exp(\gamma_{t0}) \exp(D_{gt}\delta_{gt}) > 0,$$

the first because $\pi_g \in (0, 1)$ and $T > 0$ by Step 1, the second because $\bar{b}_g > 0$ and the exponential is positive.

Step 3: the collapsed design is injective. If two collapsed parameters θ_1, θ_2 satisfy $r'_{gt}\theta_1 = r'_{gt}\theta_2$ at every supported cell, then $h = \theta_1 - \theta_2$ satisfies $r'_{gt}h = 0$ at every supported cell by linearity of $\theta \mapsto r'_{gt}\theta$, whence

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} \, (r'_{gt}h)^2 = 0.$$

By Assumption 5 this forces $h = 0$, i.e. $\theta_1 = \theta_2$. Thus the linear map $\theta \mapsto (r'_{gt}\theta)_{(g,t)}$ from $\mathbb{R}^{|\mathcal{C}|+T}$ into the space of supported-cell tables is injective.

Step 4: existence and uniqueness of the maximizer. Written cellwise,

$$L(\theta; \delta) = \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} \{ m_{gt}(\delta) \, \eta_{gt} - \exp(\eta_{gt}) \}, \qquad \eta_{gt} = r'_{gt}\theta,$$

so $L(\cdot; \delta)$ is a finite positive-mass, positive-mean Poisson criterion composed with the injective linear design of Step 3. Two elementary cell facts drive the argument. First, for every $m > 0$ and every real η,

$$m\eta - \exp(\eta) \leq \frac{(m+1)^2}{2} - \min(m, 1) \, |\eta|,$$

which follows from $\exp(\eta) \geq 0$ for $\eta \leq 0$ and from $\exp(\eta) \geq \eta^2/2$ for $\eta \geq 0$. Second, for every $q > 0$, every m, and every pair $\eta \neq \eta'$,

$$\frac{q\{m\eta - \exp(\eta)\} + q\{m\eta' - \exp(\eta')\}}{2} < q \left\{ m \, \frac{\eta + \eta'}{2} - \exp\left(\frac{\eta + \eta'}{2} \right) \right\},$$

by strict convexity of the exponential. The first display supplies the tail bound used for existence. Applying it cell by cell bounds $L(\theta; \delta)$ above by a constant minus the norm of the weighted index table

$$\left(q_{gt} \min\{ m_{gt}(\delta), 1 \} \, r'_{gt}\theta \right)_{(g,t)}.$$

The map sending θ to this table is injective by Step 3 and the strict positivity from Step 2, so finite-dimensional norm equivalence bounds this table norm below by a positive multiple of $\|\theta\|$. Hence $L(\theta; \delta) \to -\infty$ as $\|\theta\| \to \infty$; since $L(\cdot; \delta)$ is continuous, a global maximizer exists on a

large enough closed ball. The second display gives the uniqueness step: if $\theta \neq \theta'$ were two global maximizers, injectivity would give a supported cell with $r'_{gt}\theta \neq r'_{gt}\theta'$, and the strict inequality in that cell together with the weak inequality in the others would make the midpoint $(\theta+\theta')/2$ strictly better, a contradiction. Hence

$$L(\cdot;\delta) \text{ has exactly one global maximizer.}$$

Step 5: identification of the projection with the unique maximizer. By Definition 3, $\theta^\star(\delta)$ is a maximizer of $L(\cdot;\delta)$. Step 4 shows that this maximizer is unique. Therefore $L(\theta;\delta) \leq L(\theta^\star(\delta);\delta)$ for all θ, and $L(\theta;\delta) = L(\theta^\star(\delta);\delta)$ implies $\theta = \theta^\star(\delta)$. This is the asserted unique global maximality.

Step 6: every directional score vanishes. Fix a direction $d \in \mathbb{R}^{|\mathcal{C}|+T}$ and consider $s \mapsto L(\theta^\star(\delta) + sd;\delta)$. In each cell the index is affine in s,

$$r'_{gt}\{\theta^\star(\delta) + sd\} = r'_{gt}\theta^\star(\delta) + s\, r'_{gt}d,$$

and $\eta \mapsto m\eta - \exp(\eta)$ is differentiable with derivative $m - \exp(\eta)$. Differentiating the finite sum term by term at $s = 0$ and using $\mu^\star_{gt}(\delta) = \exp\{r'_{gt}\theta^\star(\delta)\}$,

$$\frac{d}{ds}L(\theta^\star(\delta) + sd;\delta)\Big|_{s=0} = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\,(r'_{gt}d)\{m_{gt}(\delta) - \mu^\star_{gt}(\delta)\}.$$

By Step 5 the scalar function $s \mapsto L(\theta^\star(\delta)+sd;\delta)$ attains its maximum at $s = 0$, so its derivative there is zero:

$$\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\,(r'_{gt}d)\{m_{gt}(\delta) - \mu^\star_{gt}(\delta)\} = 0.$$

Step 7: the two stated score equations. Take d to be the unit vector in nuisance coordinate j, so that $r'_{gt}d = X_{gt,j}$; Step 6 gives

$$\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}X_{gt,j}\{m_{gt}(\delta) - \mu^\star_{gt}(\delta)\} = 0.$$

Take instead d to be the pure treatment direction, with zero nuisance block and treatment coordinate one, so that $r'_{gt}d = D_{gt}$; Step 6 gives

$$\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}D_{gt}\{m_{gt}(\delta) - \mu^\star_{gt}(\delta)\} = 0.$$

□

Lemma 1 is the collapsed analogue of the pseudo-true score condition for a Poisson quasi-likelihood projection. The rank condition rules out flat directions in the fixed-effect and treatment regressors, while positivity of the cell means keeps the exponential criterion strictly concave on the relevant design span. The argument uses the mean criterion characteristic of standard pseudo-maximum-likelihood analysis [White, 1982, Gouriéroux et al., 1984a].

Combining Proposition 1 and Lemma 1 fixes the population target before the sign results are applied. The derivative result in Theorem 1 uses the score equations to express the response

of the treatment coordinate through the weighted residualized treatment. The homogeneous benchmark in Theorem 2 then identifies the correctly specified common-effect case. The primitive sign theorem, Theorem 3, comes next because it supplies the global sign criterion used to place the explicit four-cohort construction in $\mathcal{R}_4$. With that ordering, Theorem 4 becomes the concrete specialization of the sign frontier to the witness in Definition 6.

Reproducibility note. The formal supplement verifies the displayed population claims in Lean 4 and records the correspondence between mathematical statements and declarations. Its deterministic scope covers the finite-dimensional projection, collapse, derivative-sign, primitive-sign, and four-cohort calculations. Sampling consistency, inference, and empirical identification of the primitive means and proportional effects form the natural empirical extension.

B Proofs of the main results

Proof of Proposition 1. Throughout, a collapsed parameter $\theta \in \mathbb{R}^{|\mathcal{C}|+T}$ is written through the levels it induces,

$$c_g(\theta) = \begin{cases} \theta_{\text{int}}, & g = \infty, \\ \theta_{\text{int}} + \theta_{\text{coh},g}, & g \neq \infty, \end{cases} \qquad w_t(\theta) = \begin{cases} 0, & t = 1, \\ \theta_{\text{time},t}, & t \neq 1, \end{cases} \qquad \beta(\theta) = \theta_{\text{trt}},$$

so that for every supported cell (g,t), $g \in \mathcal{C}$,

$$r_{gt}'\theta = c_g(\theta) + w_t(\theta) + \beta(\theta) D_{gt}.$$

A unit parameter $\theta_N \in \mathbb{R}^{N+T}$ is written the same way,

$$\alpha_i(\theta_N) = \begin{cases} \theta_{N,\text{int}}, & i = 1, \\ \theta_{N,\text{int}} + \theta_{N,\text{unit},i}, & i \neq 1, \end{cases} \qquad w_t(\theta_N) = \begin{cases} 0, & t = 1, \\ \theta_{N,\text{time},t}, & t \neq 1, \end{cases} \qquad \beta(\theta_N) = \theta_{N,\text{trt}},$$

so that

$$v_{itN}'\theta_N = \alpha_i(\theta_N) + w_t(\theta_N) + \beta(\theta_N) D_{it}, \qquad D_{it} = D_{G_i t}.$$

Both identities are the statement that the intercept, dummy, and treatment coordinates of the regressor pick out exactly these terms.

Write the finite-array collapsed primitives as

$$q_{gtN} = \frac{n_{gN}}{NT}, \qquad m_{gtN}(\delta) = \bar{b}_{gN} \exp(\gamma_{t0}) \exp(D_{gt}\delta_{gt}),$$

$$L_{Nc}(\theta;\delta) = \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gtN} \{ m_{gtN}(\delta)\, r_{gt}'\theta - \exp(r_{gt}'\theta) \},$$

and write the unit-time observed mean entering the criterion $L_N(\theta_N;\delta)$ of Definition 1 as

$$m_{itN}(\delta) = b_{iN} \exp(\gamma_{t0} + D_{it}\delta_{G_i t}),$$

which is the form generated by the untreated mean of Assumption 2 together with the proportional effects of Assumption 4.

Step 0: standing positivity. The horizon assumption $T \geq 4$ gives $T > 0$, and $\infty \in \mathcal{C}$ makes $\mathcal{C}$ nonempty, so at least one supported cell exists. Fix N with $|\mathcal{C}| \leq N$. Assumption 1 then supplies $N > 0$ and $n_{gN} > 0$ for every $g \in \mathcal{C}$; every unit label lies in $\mathcal{C}$ by the array-support hypothesis. Consequently, for every $g \in \mathcal{C}$ and every t,

$$q_{gtN} = \frac{n_{gN}}{NT} > 0, \qquad \bar{b}_{gN} = n_{gN}^{-1} \sum_{i \in I_{gN}} b_{iN} > 0, \qquad m_{gtN}(\delta) > 0, \qquad m_{itN}(\delta) > 0,$$

the last two because they are products of positive baselines and exponentials.

Step 1: the collapsed designs are injective, and L_{Nc} has a unique maximizer. If $r'_{gt}\theta_1 = r'_{gt}\theta_2$ at every supported cell, then $a = \theta_1 - \theta_2$ has $r'_{gt}a = 0$ at every supported cell, so $\sum_{g \in \mathcal{C}} \sum_t q_{gt}(r'_{gt}a)^2 = 0$, and Assumption 5 forces $a = 0$. Thus $\theta \mapsto (r'_{gt}\theta)_{(g,t)}$ is injective. With positive masses q_{gtN}, positive means $m_{gtN}(\delta)$, and this injective design, the finite collapsed criterion is a finite positive-mass, positive-mean Poisson criterion, so it has exactly one global maximizer θ_N^c: each cell contribution $\eta \mapsto m\eta - \exp(\eta)$ is strictly concave and satisfies $m\eta - \exp(\eta) \leq (m+1)^2/2 - \min(m, 1)|\eta|$, which by injectivity makes $L_{Nc}(\cdot; \delta)$ continuous, coercive, and strictly concave. Because θ_N^c maximizes and the criterion is differentiable, every collapsed directional score vanishes: for every $a \in \mathbb{R}^{|\mathcal{C}|+T}$,

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gtN}\,(r'_{gt}a)\{m_{gtN}(\delta) - \exp(r'_{gt}\theta_N^c)\} = 0.$$

Step 2: lifting the collapsed maximizer to unit fixed effects. Define the unit parameter θ_N^{lift} by

$$\alpha_i(\theta_N^{\text{lift}}) = \log \frac{b_{iN}}{\bar{b}_{G_iN}} + c_{G_i}(\theta_N^c), \qquad w_t(\theta_N^{\text{lift}}) = w_t(\theta_N^c), \qquad \beta(\theta_N^{\text{lift}}) = \beta(\theta_N^c).$$

(Any prescribed levels are realized by a unique choice of intercept, dummy, and treatment coordinates.) By the two level identities,

$$v'_{itN}\theta_N^{\text{lift}} = \log \frac{b_{iN}}{\bar{b}_{G_iN}} + r'_{G_it}\theta_N^c.$$

Exponentiating and using $m_{itN}(\delta) = (b_{iN}/\bar{b}_{G_iN})\, m_{G_it,N}(\delta)$, which holds because $m_{itN}(\delta) = b_{iN} \exp(\gamma_{t0}) \exp(D_{G_it}\delta_{G_it})$ and $m_{G_it,N}(\delta) = \bar{b}_{G_iN} \exp(\gamma_{t0}) \exp(D_{G_it}\delta_{G_it})$, each unit residual is the corresponding cell residual scaled by the baseline ratio:

$$m_{itN}(\delta) - \exp(v'_{itN}\theta_N^{\text{lift}}) = \frac{b_{iN}}{\bar{b}_{G_iN}} \left\{ m_{G_it,N}(\delta) - \exp(r'_{G_it}\theta_N^c) \right\}.$$

Step 3: unit scores at the lift are collapsed scores. For every $g \in \mathcal{C}$, the definition of $\bar{b}_{gN}$ gives

$$\sum_{i \in I_{gN}} \frac{b_{iN}}{\bar{b}_{gN}} = n_{gN}.$$

Given a unit direction $d \in \mathbb{R}^{N+T}$, define the collapsed direction $d^c \in \mathbb{R}^{|\mathcal{C}|+T}$ by its levels

$$c_g(d^c) = n_{gN}^{-1} \sum_{i \in I_{gN}} \frac{b_{iN}}{\bar{b}_{gN}}\, \alpha_i(d), \qquad w_t(d^c) = w_t(d), \qquad \beta(d^c) = \beta(d).$$

Then

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(v'_{itN}d)\left\{m_{itN}(\delta)-\exp(v'_{itN}\theta_N^{\text{lift}})\right\}=\sum_{g\in\mathcal{C}}\sum_{t=1}^{T}q_{gtN}\,(r'_{gt}d^c)\left\{m_{gtN}(\delta)-\exp(r'_{gt}\theta_N^c)\right\}.$$

Indeed, substituting the residual identity of Step 2 and grouping units by cohort (legitimate because every G_i lies in $\mathcal{C}$), the inner sum over $i\in I_{gN}$ at fixed (g,t) is

$$(NT)^{-1}\left\{m_{gtN}(\delta)-\exp(r'_{gt}\theta_N^c)\right\}\sum_{i\in I_{gN}}\frac{b_{iN}}{\bar{b}_{gN}}\left\{\alpha_i(d)+w_t(d)+\beta(d)D_{gt}\right\},$$

and the displayed baseline-ratio identity turns that bracketed sum into $n_{gN}\{c_g(d^c)+w_t(d^c)+\beta(d^c)D_{gt}\}=n_{gN}\,r'_{gt}d^c$, while $n_{gN}/(NT)=q_{gtN}$. Applying the vanishing collapsed score of Step 1 to $a=d^c$, we conclude that every unit directional score at θ_N^{lift} vanishes:

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(v'_{itN}d)\left\{m_{itN}(\delta)-\exp(v'_{itN}\theta_N^{\text{lift}})\right\}=0\qquad\text{for every } d.$$

Step 4: the lift maximizes the unit criterion. For a Poisson criterion with nonnegative weights, vanishing of all directional scores is sufficient for global maximality:

$$L_N(\theta_N;\delta)\le L_N(\theta_N^{\text{lift}};\delta)\qquad\text{for every }\theta_N\in\mathbb{R}^{N+T},$$

because the tangent-line bound $\exp(\eta')\ge\exp(\eta)\{1+(\eta'-\eta)\}$ makes each cell increment at most its linearization, and the sum of the linearizations is the score in the direction $\theta_N-\theta_N^{\text{lift}}$, which is zero by Step 3.

Step 5: the unit design is injective. Suppose $v'_{itN}d=0$ for every unit i and every period t. For each $g\in\mathcal{C}$ choose a representative unit $j_g\in I_{gN}$; this is possible because $n_{gN}>0$. Define a collapsed parameter a by $c_g(a)=\alpha_{j_g}(d)$, $w_t(a)=w_t(d)$, $\beta(a)=\beta(d)$. Then for every supported cell,

$$r'_{gt}a=\alpha_{j_g}(d)+w_t(d)+\beta(d)D_{gt}=v'_{j_gtN}d=0,$$

using $G_{j_g}=g$. Hence $\sum_{g\in\mathcal{C}}\sum_t q_{gt}(r'_{gt}a)^2=0$, so Assumption 5 gives $a=0$; in particular $\beta(d)=0$ and $w_t(d)=0$ for every t. Evaluating $0=v'_{i1N}d=\alpha_i(d)+w_1(d)+\beta(d)D_{i1}$ in the first period and substituting $w_1(d)=0$ and $\beta(d)=0$ gives $\alpha_i(d)=0$ for every unit i. All levels of d vanish, so $d=0$, and the unit design $\theta_N\mapsto(v'_{itN}\theta_N)_{i,t}$ is injective.

Step 6: unique unit maximizer and equality of treatment coordinates. The unit criterion $L_N(\cdot;\delta)$ is a finite Poisson criterion with the constant positive weights $(NT)^{-1}$, positive means $m_{itN}(\delta)$, and, by Step 5, an injective design; exactly as in Step 1 it therefore has a unique global maximizer. By Step 4 that maximizer is θ_N^{lift}. Hence $L_N(\cdot;\delta)$ and $L_{Nc}(\cdot;\delta)$ each have a unique global maximizer, and

$$\beta_N^\star(\delta)=\beta(\theta_N^{\text{lift}})=\beta(\theta_N^c),$$

i.e. $\beta_N^\star(\delta)$ is the treatment coordinate of the unique maximizer of $L_{Nc}(\theta;\delta)$.

Step 7: convergence to the limiting collapsed projection. Fix a supported cell (g,t). By Assumption 1,

$$q_{gtN}=\frac{n_{gN}/N}{T}\longrightarrow\frac{\pi_g}{T}=q_{gt},$$

and by Assumption 3, since $\exp(\gamma_{t0})\exp(D_{gt}\delta_{gt})$ does not depend on N,

$$m_{gtN}(\delta) = \bar{b}_{gN}\exp(\gamma_{t0})\exp(D_{gt}\delta_{gt}) \longrightarrow \bar{b}_g\exp(\gamma_{t0})\exp(D_{gt}\delta_{gt}) = m_{gt}(\delta).$$

Write the finite set of supported cells as $z = (g,t)$, put $q_z = q_{gt}$, $m_z = m_{gt}(\delta)$, and write $\eta_z(\theta) = r'_{gt}\theta$ for the supported-cell linear predictor. The preceding convergence is cellwise convergence of (q_{zN}, m_{zN}) to the strictly positive array (q_z, m_z), and the map $\theta \mapsto (\eta_z(\theta))_z$ is injective by Step 1. Define

$$s_z = \frac{q_z}{2}\min\left\{\frac{m_z}{2}, 1\right\}, \qquad (B\theta)_z = s_z\eta_z(\theta).$$

Since every s_z is positive, B is injective. Finite-dimensional norm equivalence therefore gives a constant $\kappa > 0$ such that

$$\kappa\|\theta\| \le \|B\theta\| \qquad \text{for every } \theta.$$

For all sufficiently large N, uniformly over supported cells,

$$\frac{q_z}{2} < q_{zN} < 2q_z, \qquad \frac{m_z}{2} < m_{zN} < 2m_z.$$

Set

$$U_0 = \sum_z 2q_z\frac{(2m_z+1)^2}{2}, \qquad Q_0 = \sum_z 2q_z, \qquad R_0 = \frac{U_0+Q_0}{\kappa}.$$

The one-cell bound $m\eta - \exp(\eta) \le (m+1)^2/2 - \min(m,1)|\eta|$, combined with the displayed eventual coefficient bounds, gives for every θ and all sufficiently large N,

$$L_{Nc}(\theta;\delta) \le U_0 - \|B\theta\|.$$

The same coefficient bounds give $-Q_0 \le L_{Nc}(0;\delta)$. Since θ^c_N maximizes $L_{Nc}(\cdot;\delta)$,

$$-Q_0 \le L_{Nc}(0;\delta) \le L_{Nc}(\theta^c_N;\delta) \le U_0 - \|B\theta^c_N\|,$$

and therefore $\|\theta^c_N\| \le R_0$ eventually. Thus the finite maximizers are eventually contained in the compact ball $K = \{\theta : \|\theta\| \le R_0\}$.

Because there are finitely many supported cells, the cellwise convergence of (q_{zN}, m_{zN}) also gives uniform convergence $L_{Nc}(\cdot;\delta) \to L(\cdot;\delta)$ on this compact set, and L is continuous there. By Lemma 1, the limiting criterion $L(\cdot;\delta)$ has the unique global maximizer $\theta^\star(\delta)$, hence the unique maximizer on $K \cup \{\theta^\star(\delta)\}$. The compact-containment bound, compact-uniform convergence, eventual global maximality of θ^c_N, and this uniqueness yield

$$\theta^c_N \longrightarrow \theta^\star(\delta).$$

Taking the treatment coordinate, which is a continuous function of the parameter, gives $\beta(\theta^c_N) \to \beta^\star(\delta)$; since $\beta^\star_N(\delta) = \beta(\theta^c_N)$ for every $N \ge |\mathcal{C}|$ by Step 6,

$$\beta^\star_N(\delta) \longrightarrow \beta^\star(\delta).$$

Step 8: invariance to within-cohort baseline rearrangement. Fix $N \ge |\mathcal{C}|$ and a positive baseline array b'_{iN} with $\bar{b}'_{gN} = \bar{b}_{gN}$ for every $g \in \mathcal{C}$. The finite collapsed observed mean depends on the baselines only through the within-cohort average,

$$m_{gtN}(\delta; b') = \bar{b}'_{gN}\exp(\gamma_{t0})\exp(D_{gt}\delta_{gt}) = \bar{b}_{gN}\exp(\gamma_{t0})\exp(D_{gt}\delta_{gt}) = m_{gtN}(\delta; b),$$

and the masses q_{gtN} do not involve the baselines at all. Hence $L_{Nc}(\cdot;\delta;b') = L_{Nc}(\cdot;\delta;b)$ as functions of θ, so their unique maximizers coincide. Applying Step 6 once with b' and once with b,

$$\beta^\star_N(\delta;b') = \beta(\theta^c_N) = \beta^\star_N(\delta;b).$$

□

Proof of Theorem 1. Write $W_{gt} = \widetilde{W}_{gt}(\delta)$, $\mu^\star_{gt} = \mu^\star_{gt}(\delta)$, and

$$\mathcal{E} = \sum_{g\in\mathcal{C}} \sum_{t=1}^{T} q_{gt}\mu^\star_{gt} W^2_{gt}.$$

A collapsed parameter is written $\vartheta = (\eta, \beta)$ with $r'_{gt}\vartheta = \sum_j X_{gt,j}\eta_j + D_{gt}\beta$, and $\rho^\star(\delta)$ is the weighted nuisance coefficient of Definition 4, so that

$$W_{gt} = D_{gt} - \sum_j X_{gt,j}\rho^\star_j(\delta).$$

Step 1: $\mathcal{E} > 0$. Each summand is nonnegative, since $q_{gt} > 0$, $\mu^\star_{gt} = \exp\{r'_{gt}\theta^\star(\delta)\} > 0$, and $W^2_{gt} \geq 0$. If $\mathcal{E} = 0$, every summand vanishes, and positivity of $q_{gt}\mu^\star_{gt}$ forces $W_{gt} = 0$ at every supported cell. Consider the collapsed parameter

$$\bar{\vartheta} = (-\rho^\star(\delta), 1),$$

whose treatment coordinate is one, so $\bar{\vartheta} \neq 0$. Its collapsed index is exactly the residual,

$$r'_{gt}\bar{\vartheta} = D_{gt} - \sum_j X_{gt,j}\rho^\star_j(\delta) = W_{gt} = 0 \qquad \text{at every supported cell,}$$

whence $\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}(r'_{gt}\bar{\vartheta})^2 = 0$, contradicting Assumption 5. Thus $\mathcal{E} > 0$.

Step 2: the perturbed criterion. For $x \in \mathbb{R}$ let δ^x be δ with the coordinate δ_{ks} replaced by x, and write $x_0 = \delta_{ks}$. Because $D_{ks} = 1$ and $m_{gt}(\delta) = B_{gt}\exp(D_{gt}\delta_{gt})$, the perturbation moves exactly one collapsed mean:

$$m_{ks}(\delta^x) = B_{ks}\exp(x), \qquad m_{gt}(\delta^x) = m_{gt}(\delta) \quad \text{for } (g,t) \neq (k,s),$$

Setting

$$p_{gt} = \begin{cases} B_{ks}\exp(\delta_{ks}), & (g,t) = (k,s), \\ 0, & (g,t) \neq (k,s), \end{cases} \qquad \text{we have} \qquad \frac{d}{dx}\, m_{gt}(\delta^x)\Big|_{x=x_0} = p_{gt}.$$

Step 3: the score system and its solution branch. For a direction d put

$$S_d(x,\theta) = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}(r'_{gt}d)\left\{m_{gt}(\delta^x) - \exp(r'_{gt}\theta)\right\}.$$

By Lemma 1 applied at the effect array δ^x, the criterion $L(\cdot;\delta^x)$ has a unique global maximizer, and the nuisance and treatment score equations imply by linearity that $S_d(x,\theta) = 0$ for every direction d at that maximizer. Conversely, vanishing of every directional score is sufficient for

global maximality of a Poisson criterion with nonnegative weights, by the tangent-line bound $\exp(\eta') \geq \exp(\eta)\{1 + (\eta' - \eta)\}$. The system is smooth in (x, θ), it is solved at $(x_0, \theta^\star(\delta))$, and its derivative in θ at that point is the bilinear form

$$d, e \;\mapsto\; -\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\,\mu^\star_{gt}\,(r'_{gt}d)(r'_{gt}e),$$

which is negative definite: if $e \neq 0$, the injectivity of $\theta \mapsto (r'_{gt}\theta)_{(g,t)}$, obtained from Assumption 5 as in Lemma 1, gives a supported cell with $r'_{gt}e \neq 0$. Hence the derivative is invertible. The implicit function theorem therefore supplies a continuously differentiable $x \mapsto \theta(x)$ defined near x_0, with $\theta(x_0) = \theta^\star(\delta)$ and $S_d(x, \theta(x)) = 0$ for every d; by the characterization above and uniqueness of the maximizer, $\theta(x) = \theta^\star(\delta^x)$ for x near x_0. In particular $x \mapsto \beta^\star(\delta^x)$ is differentiable at x_0, with derivative the treatment coordinate v_β of $v = \theta'(x_0)$. Differentiating $S_d(x, \theta(x)) = 0$ at $x = x_0$ and using the one-cell derivative display above gives the linearized score: for every direction d,

$$\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}(r'_{gt}d)\left\{p_{gt} - \mu^\star_{gt}\,(r'_{gt}v)\right\} = 0.$$

Step 4: solving the linearized score by weighted FWL. Put

$$K = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\mu^\star_{gt} > 0, \qquad \psi_{gt} = \frac{q_{gt}\mu^\star_{gt}}{K}, \qquad \langle f, h\rangle = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T}\psi_{gt}f_{gt}h_{gt},$$

and define the mean-normalized one-cell source and the nuisance part of $v = (v_\eta, v_\beta)$ by

$$u_{gt} = \frac{p_{gt}}{\mu^\star_{gt}}, \qquad \alpha_{gt} = \sum_j X_{gt,j}v_{\eta,j},$$

so that $r'_{gt}v = \alpha_{gt} + v_\beta D_{gt}$ and α lies in the span of the nuisance columns $X_{\cdot,j}$. Because $\mu^\star_{gt} > 0$ and $K\psi_{gt} = q_{gt}\mu^\star_{gt}$, each summand of the linearized score factors as

$$q_{gt}(r'_{gt}d)\left\{p_{gt} - \mu^\star_{gt}\,(r'_{gt}v)\right\} = K\,\psi_{gt}\,(r'_{gt}d)\left\{u_{gt} - v_\beta D_{gt} - \alpha_{gt}\right\},$$

so dividing the whole linearized score by $K > 0$ rewrites it as

$$\langle u - v_\beta D - \alpha,\ (r'_{gt}d)_{g,t}\rangle = 0 \qquad \text{for every direction } d.$$

Taking d to be the pure treatment direction gives $\langle u - v_\beta D - \alpha,\ D\rangle = 0$; taking $d = (\rho, 0)$ for arbitrary ρ gives $\langle u - v_\beta D - \alpha,\ h\rangle = 0$ for every h in the nuisance span. The coefficient $\rho^\star(\delta)$ in Definition 4 minimizes a finite weighted least-squares projection of D on the nuisance span. Its normal equations give

$$\langle W, h\rangle = 0 \quad \text{for every } h \text{ in the nuisance span}, \qquad D - W \text{ lies in the nuisance span}.$$

Consequently $\langle W, D\rangle = \langle W, W\rangle$, and, since α also lies in the nuisance span, $\langle W, \alpha\rangle = 0$. Combining the score normal equations with $D - W$ in the nuisance span gives

$$0 = \langle u - v_\beta D - \alpha,\ D\rangle - \langle u - v_\beta D - \alpha,\ D - W\rangle = \langle u - v_\beta D - \alpha,\ W\rangle.$$

Expanding this last display and using the projection orthogonality identities yields

$$0 = \langle W,\ u - v_\beta D - \alpha \rangle = \langle W, u \rangle - v_\beta \langle W, W \rangle.$$

Now $\langle W, W \rangle = \mathcal{E}/K > 0$ by the positivity of $\mathcal{E}$ shown above, and, since p is supported on the single cell (k, s),

$$\langle W, u \rangle = \frac{1}{K} \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} \mu^\star_{gt} W_{gt} \frac{p_{gt}}{\mu^\star_{gt}} = \frac{q_{ks} B_{ks} \exp(\delta_{ks}) W_{ks}}{K}.$$

Dividing,

$$v_\beta = \frac{\langle W, u \rangle}{\langle W, W \rangle} = \frac{q_{ks} B_{ks} \exp(\delta_{ks}) W_{ks}}{\mathcal{E}}.$$

Step 5: the derivative formula. Combining the differentiability branch with the FWL calculation,

$$\frac{d}{dx}\, \beta^\star\big(\delta \text{ with } \delta_{ks} \text{ replaced by } x\big)\Big|_{x = \delta_{ks}} = v_\beta = \frac{q_{ks} B_{ks} \exp(\delta_{ks}) \widetilde{W}_{ks}(\delta)}{\mathcal{E}(\delta)},$$

which is the asserted differentiability at $x = \delta_{ks}$ together with the displayed derivative.

Step 6: the sign equivalences. The derivative is cW_{ks} with

$$c = \frac{q_{ks} B_{ks} \exp(\delta_{ks})}{\mathcal{E}} > 0,$$

because $q_{ks} > 0$, $B_{ks} = \bar{b}_k \exp(\gamma_{s0}) > 0$, $\exp(\delta_{ks}) > 0$, and $\mathcal{E} > 0$ as shown above. Multiplication by a positive scalar preserves and reflects each of the three order relations, so

$$cW_{ks} < 0 \iff W_{ks} < 0, \qquad cW_{ks} = 0 \iff W_{ks} = 0, \qquad 0 < cW_{ks} \iff 0 < W_{ks},$$

which, with $W_{ks} = \widetilde{W}_{ks}(\delta)$, are exactly the three asserted equivalences. □

Proof of Theorem 4. Throughout, $T = 4$, $\mathcal{C} = \{2, 3, 4, \infty\}$, $\pi_g = 1/4$, $\bar{b}_g = 1$, and $\gamma_{t0} = 0$, so that

$$B_{gt} = \bar{b}_g \exp(\gamma_{t0}) = 1, \qquad q_{gt} = \frac{\pi_g}{T} = \frac{1}{16} \qquad \text{for every supported cell } (g, t).$$

Treatment is absorbing from the adoption date on, so $D_{gt} = 1$ exactly when $g \le t$, and $D_{\infty t} = 0$; in table form, with rows $g = 2, 3, 4, \infty$ and columns $t = 1, 2, 3, 4$,

$$(D_{gt}) = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

A collapsed parameter a is written through its levels: the intercept a_{int}, the cohort dummies $a_{\text{coh},g}$ for $g \neq \infty$, the period dummies $a_{\text{time},t}$ for $t \neq 1$, and the treatment coordinate a_{trt}, so that

$$r'_{gt} a = \begin{cases} a_{\text{int}} + a_{\text{time},t} + a_{\text{trt}} D_{gt}, & g = \infty, \\ a_{\text{int}} + a_{\text{coh},g} + a_{\text{time},t} + a_{\text{trt}} D_{gt}, & g \neq \infty, \end{cases}$$

with the convention $a_{\text{time},1} = 0$, the period-1 dummy being absent.

Step 1: the collapsed design has full rank. Let a satisfy $\sum_{g\in\mathcal{C}}\sum_{t=1}^{4} q_{gt}(r_{gt}'a)^2 = 0$. Every summand is nonnegative and every $q_{gt} = 1/16$ is nonzero, so

$$r_{gt}'a = 0 \qquad \text{for every supported cell } (g,t).$$

Evaluate this at four families of cells, using the displayed level formula and the treatment table. At $(\infty,1)$: all dummies vanish and $D_{\infty 1} = 0$, so $a_{\text{int}} = 0$. At $(g,1)$ for $g \in \{2,3,4\}$: the period-1 dummy vanishes and $D_{g1} = 0$, so $a_{\text{coh},g} = -a_{\text{int}} = 0$. At (∞,t) for $t \in \{2,3,4\}$: the cohort dummy vanishes and $D_{\infty t} = 0$, so $a_{\text{time},t} = -a_{\text{int}} = 0$. Finally at $(2,2)$, where $D_{2,2} = 1$: all dummies and the intercept vanish, so $a_{\text{trt}} = 0$. Hence $a = 0$, which is Assumption 5 for this design.

Step 2: region membership. We verify each clause of Definition 5. The horizon clause is $4 \le T = 4$. The support clause holds because $\infty \in \mathcal{C}$ and the finite cohorts $2,3,4$ all adopt after the first period; those same three cohorts together with ∞ give the frontier scope in Assumption 6. The shares satisfy

$$0 < \pi_g = \frac{1}{4} < 1, \qquad \sum_{g\in\mathcal{C}} \pi_g = 4 \cdot \frac{1}{4} = 1.$$

The untreated-mean clause is witnessed by $\bar{b}_g = 1$, since $B_{gt} = 1 = 1 \cdot \exp(0)$ in every supported cell. The positive-effect clause is Step 3 below, and the rank clause is Step 1. It remains to check $\beta^\star(\delta) < 0$.

Evaluate the polynomial of Definition 7 at the present primitives. Write $x = \exp(\log(101/100)) = 101/100$ for the common treated multiplier off cell $(2,4)$ and $y = \exp(\log 4) = 4$ for the multiplier at $(2,4)$. Since $\pi_g = 1/4$ and $B_{gt} = 1$, the components $h_{gt} = \pi_g B_{gt} \exp(D_{gt}\delta_{gt})$ are

$$(4h_{gt}) = \begin{pmatrix} 1 & x & x & y \\ 1 & 1 & x & x \\ 1 & 1 & 1 & x \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

again with rows $g = 2,3,4,\infty$ and columns $t = 1,\dots,4$, so that

$$4R_2 = 1 + 2x + y, \quad 4R_3 = 2 + 2x, \quad 4R_4 = 3 + x, \quad 4R_\infty = 4,$$

$$4C_1 = 4, \quad 4C_2 = 3 + x, \quad 4C_3 = 2 + 2x, \quad 4C_4 = 1 + 2x + y,$$

$$4M = 10 + 5x + y, \qquad 4A = 5x + y.$$

The treated cells are $(2,2),(2,3),(2,4),(3,3),(3,4),(4,4)$, so

$$16\sum_{g\in\mathcal{C}}\sum_{t=1}^{4} D_{gt}R_gC_t = (1+2x+y)(6+5x+y) + (2+2x)(3+4x+y) + (3+x)(1+2x+y),$$

which expands to $20x^2 + 10xy + y^2 + 38x + 12y + 15$, while $16MA = (10+5x+y)(5x+y) = 25x^2 + 10xy + y^2 + 50x + 10y$. Subtracting,

$$\Phi = \frac{5x^2 + 12x - 2y - 15}{16} = \frac{5(101/100)^2 + 12(101/100) - 2\cdot 4 - 15}{16} = -\frac{11559}{32000} < 0.$$

The collapsed sign comparison used here is the beta-zero frontier-elimination comparison underlying Theorem 3. In this specialization, let $\theta^0 = (u^0, 0)$ be the fixed-effect parameter whose fitted cell mean is

$$\mu^0_{gt} = \frac{R_g C_t}{\pi_g M}.$$

The nuisance scores at θ^0 vanish by the row and column identities defining R_g, C_t, and M. The treatment score at θ^0 in the limiting collapsed criterion of Definition 3 is

$$S_0 = \sum_{g \in \mathcal{C}} \sum_{t=1}^{4} q_{gt} D_{gt} \{m_{gt}(\delta) - \mu^0_{gt}\} = \frac{\Phi}{TM}.$$

Here $TM > 0$, since $T = 4$ and every h_{gt} is positive. The weights q_{gt} and means $m_{gt}(\delta)$ are positive, and the rank calculation above makes the collapsed design map injective. If $\theta^\star = (u^\star, \beta^\star)$ is the maximizer in Definition 3 and $v = \theta^\star - \theta^0$, the score at $\theta^\star$ is zero. Whenever $v \neq 0$, strict monotonicity of the exponential along the nonzero fitted-index direction gives

$$0 < \sum_{g \in \mathcal{C}} \sum_{t=1}^{4} q_{gt} \, (r'_{gt} v) \{\exp(r'_{gt} \theta^\star) - \exp(r'_{gt} \theta^0)\} = \beta^\star S_0.$$

If $S_0 = 0$, then all score coordinates vanish at θ^0, so θ^0 is the unique collapsed maximizer and $\beta^\star = 0$. Therefore

$$\beta^\star(\delta) < 0 \iff \Phi < 0, \qquad \beta^\star(\delta) = 0 \iff \Phi = 0, \qquad 0 < \beta^\star(\delta) \iff 0 < \Phi.$$

For the present primitives $\Phi < 0$, hence $\beta^\star(\delta) < 0$. Since the primitive tuple entering Definition 5 is exactly the restriction of $(\pi_g, B_{gt}, \delta_{gt})$ to the supported and treated cells, its treatment coordinate is this same $\beta^\star(\delta)$. All clauses hold, so $W_4 \in \mathcal{R}_4$.

Step 3: strict positivity of the treated effects. If (g, t) is supported with $D_{gt} = 1$, then $\delta_{gt} \in \{\log 4, \log(101/100)\}$, and both are strictly positive because $4 > 1$ and $101/100 > 1$.

Step 4: cell $(2, 4)$ has the strictly largest treated effect. The cell $(2, 4)$ is treated, since $2 \leq 4$, and $\delta_{2,4} = \log 4$. Every other treated supported cell has $\delta_{gt} = \log(101/100)$, and $101/100 < 4$ with the logarithm strictly increasing on the positive reals gives

$$\delta_{gt} = \log(101/100) < \log 4 = \delta_{2,4}.$$

Step 5: the weighted FWL residual table at the no-effect vector. Take the effect vector identically zero. Theorem 2 applies to it with common treated multiplier $\delta_0 = 0$: its common-treated-effect hypothesis holds because every effect is zero, its rank hypothesis is Step 1, and its untreated-mean hypothesis is discharged by the auxiliary array in which every unit's untreated and treated outcomes are deterministically 1, with unit baselines $b_{iN} = 1$ and the time components $\gamma_{t0} = 0$ already fixed here, so that $E_{\mathcal{P}_N}[Y_{it}(0)] = 1 = b_{iN} \exp(\gamma_{t0})$ with the normalization $\gamma_{10} = 0$. Its conclusion gives, for every supported cell,

$$\mu^\star_{gt}(0) = B_{gt} \exp(D_{gt} \cdot 0) = 1.$$

Hence the projection weights of Definition 4 are

$$w_{gt} = q_{gt} \mu^\star_{gt}(0) = \frac{1}{16} \qquad \text{for every supported cell.}$$

Let V_{gt} denote the entry of the displayed table,

$$V = \frac{1}{8}\begin{pmatrix} -3 & 3 & 1 & -1 \\ -1 & -3 & 3 & 1 \\ 1 & -1 & -3 & 3 \\ 3 & 1 & -1 & -3 \end{pmatrix},$$

with rows $g = 2, 3, 4, \infty$ and columns $t = 1, 2, 3, 4$. Two facts about V are checked by direct summation of these sixteen numbers:

$$\sum_{t=1}^{4} V_{gt} = 0 \quad (g \in \mathcal{C}), \qquad \sum_{g \in \mathcal{C}} V_{gt} = 0 \quad (t = 1, \ldots, 4).$$

First, $D - V$ lies in the fixed-effect nuisance span. Indeed, with

$$(c_2, c_3, c_4, c_\infty) = \Big(\frac{3}{4}, \frac{1}{2}, \frac{1}{4}, 0\Big), \qquad (s_1, s_2, s_3, s_4) = \Big(0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\Big),$$

one checks cell by cell that

$$D_{gt} - V_{gt} = -\frac{3}{8} + c_g + s_t \qquad \text{for every supported cell,}$$

and the right-hand side is the value at (g, t) of the nuisance combination with intercept coefficient $-3/8$, cohort-dummy coefficients c_g (with $c_\infty = 0$ absorbed by the intercept), and period-dummy coefficients s_t (with $s_1 = 0$); every such additive intercept-plus-cohort-plus-period array is a linear combination of the columns $X_{gt,j}$. Second, V is orthogonal to that span under the weights w_{gt}. Each nuisance regressor is either the constant 1, the indicator of one cohort, or the indicator of one period; because the weights are the same constant $1/16$ in every cell, the corresponding weighted inner products are $\frac{1}{16}\sum_{g,t} V_{gt}$, $\frac{1}{16}\sum_t V_{g_0 t}$, and $\frac{1}{16}\sum_g V_{g t_0}$, all zero by the row and column sums above; orthogonality extends from these spanning columns to their whole span by linearity:

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{4} w_{gt} V_{gt} X_{gt,j} = 0 \qquad \text{for every nuisance coordinate } j.$$

The two facts together say that $D - V$ is the w-weighted orthogonal projection of D onto the nuisance span, which by Definition 4 means that the weighted FWL residual is the remainder:

$$\widetilde{W}_{gt}(0) = D_{gt} - (D_{gt} - V_{gt}) = V_{gt}.$$

Step 6: the late-cell residual. Reading off the entry in row $g = 2$, column $t = 4$ of V,

$$\widetilde{W}_{2,4}(0) = -\frac{1}{8}.$$

Step 7: the exact derivative at the no-effect vector. Since every weight is $1/16$ and $\widetilde{W}_{gt}(0) = V_{gt}$, the residual energy of Theorem 1 is

$$\mathcal{E}(0) = \sum_{g \in \mathcal{C}} \sum_{t=1}^{4} q_{gt} \mu^\star_{gt}(0) \widetilde{W}_{gt}(0)^2 = \frac{1}{16} \sum_{g \in \mathcal{C}} \sum_{t=1}^{4} V_{gt}^2 = \frac{1}{16} \cdot \frac{80}{64} = \frac{5}{64},$$

because each row of $8V$ consists of two entries of absolute value 3 and two of absolute value 1, contributing $9+9+1+1 = 20$ per row and 80 in total. The cell $(2,4)$ is treated, and Assumption 5 holds by Step 1, so Theorem 1 applies at $(k,s) = (2,4)$ and gives, with $q_{2,4} = 1/16$, $B_{2,4} = 1$, $\delta_{2,4} = 0$,

$$\frac{d}{dz}\,\beta^\star(\delta_{2,4} = z,\ \delta_{gt} = 0 \text{ for } (g,t) \neq (2,4))\Big|_{z=0} = \frac{q_{2,4}B_{2,4}\exp(0)\widetilde{W}_{2,4}(0)}{\mathcal{E}(0)} = \frac{(1/16)\cdot 1\cdot 1\cdot(-1/8)}{5/64} = -\frac{1}{10}.$$

Step 8: a neighborhood of the no-effect vector. The map $\delta \mapsto \widetilde{W}_{2,4}(\delta)$ is continuous at the zero array: the collapsed projection $\theta^\star(\delta)$, hence each fitted mean $\mu^\star_{gt}(\delta)$, depends continuously on δ because it is the unique maximizer of a criterion whose finitely many positive means move continuously with δ; under Assumption 5 the weighted nuisance Gram matrix built from those weights is invertible, and the residual is a continuous rational function of its entries. Since $\widetilde{W}_{2,4}(0) = -1/8 < 0$ and the cells are finite in number, continuity supplies $\varepsilon > 0$ such that

$$|\delta_{gt}| < \varepsilon \text{ at every cell} \quad\Longrightarrow\quad \widetilde{W}_{2,4}(\delta) < 0.$$

Now take any effect array δ' with $|\delta'_{gt}| < \varepsilon$ for every $g \in \mathcal{C}$ and $t \in \{1,\dots,4\}$ and with strictly positive treated effects in the sense of Assumption 7, and define its supported truncation

$$\delta^s_{gt} = \begin{cases} \delta'_{gt}, & g \in \mathcal{C}, \\ 0, & \text{otherwise.} \end{cases}$$

Then $|\delta^s_{gt}| < \varepsilon$ at every cell, so $\widetilde{W}_{2,4}(\delta^s) < 0$. Moreover δ^s and δ' agree on all supported cells, and the criterion $L(\cdot;\delta)$ – hence $\theta^\star(\delta)$, the fitted means, the weights, and the residual – involves only supported cells, so

$$\widetilde{W}_{2,4}(\delta') = \widetilde{W}_{2,4}(\delta^s) < 0.$$

Applying Theorem 1 once more at the treated cell $(2,4)$, now at the effect array δ', the derivative equals $q_{2,4}B_{2,4}\exp(\delta'_{2,4})\widetilde{W}_{2,4}(\delta')/\mathcal{E}(\delta')$ with a strictly positive prefactor and denominator, so it has the sign of $\widetilde{W}_{2,4}(\delta')$:

$$\frac{d}{dz}\,\beta^\star\big(\delta' \text{ with } \delta'_{2,4} \text{ replaced by } z\big)\Big|_{z=\delta'_{2,4}} < 0.$$

□

Proof of Theorem 2. For a collapsed parameter $\theta \in \mathbb{R}^{|\mathcal{C}|+T}$ and a supported cell (g,t), write the collapsed linear index as

$$\eta_{gt}(\theta) = r'_{gt}\theta,$$

so that, by Definition 3,

$$L(\theta;\delta) = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\left\{m_{gt}(\delta)\,\eta_{gt}(\theta) - \exp(\eta_{gt}(\theta))\right\}, \qquad \mu^\star_{gt}(\delta) = \exp\{\eta_{gt}(\theta^\star(\delta))\}.$$

The map $\theta \mapsto \eta_{gt}(\theta)$ is linear, so $\eta_{gt}(\theta) - \eta_{gt}(\theta') = \eta_{gt}(\theta - \theta')$ for all θ, θ'.

Step 1: the candidate parameter. Let ∞ be the never-treated reference cohort and define $\theta^0 = (a^0, \delta_0)$ coordinatewise by

$$a^0_{\text{int}} = \log \bar{b}_\infty, \qquad a^0_{\text{coh},c} = \log \bar{b}_c - \log \bar{b}_\infty \quad (c \in \mathcal{C},\ c \neq \infty), \qquad a^0_{\text{time},u} = \gamma_{u0} \quad (u \neq 1),$$

with treatment coordinate δ_0. Assumption 2 carries the normalization $\gamma_{10} = 0$. Hence, for every period t,

$$\begin{cases} 0, & t = 1, \\ a^0_{\text{time},t}, & t \neq 1, \end{cases} \quad = \quad \gamma_{t0},$$

since for $t = 1$ all period dummies are switched off and $\gamma_{10} = 0$, while for $t \neq 1$ exactly the dummy of period t contributes. Likewise, for every $g \in \mathcal{C}$,

$$\begin{cases} 0, & g = \infty, \\ a^0_{\text{coh},g}, & g \neq \infty, \end{cases} \quad = \quad \begin{cases} 0, & g = \infty, \\ \log \bar{b}_g - \log \bar{b}_\infty, & g \neq \infty, \end{cases}$$

because exactly the dummy of cohort g contributes when $g \neq \infty$, and none when $g = \infty$.

Step 2: the index at θ^0. Adding the intercept, cohort, period, and treatment contributions in the two cases $g = \infty$ and $g \neq \infty$ gives, for every $g \in \mathcal{C}$ and every period t,

$$\eta_{gt}(\theta^0) = \log \bar{b}_g + \gamma_{t0} + D_{gt}\delta_0.$$

Step 3: the homogeneous-effect hypothesis in product form. For every $g \in \mathcal{C}$ and every t,

$$D_{gt}\delta_{gt} = D_{gt}\delta_0 :$$

if $D_{gt} = 1$ the common treated-cell hypothesis gives $\delta_{gt} = \delta_0$, and if $D_{gt} = 0$ both sides vanish.

Step 4: the criterion is exactly specified at θ^0. Using $B_{gt} = \bar{b}_g \exp(\gamma_{t0})$ together with Steps 2 and 3, the observed collapsed mean satisfies

$$m_{gt}(\delta) = B_{gt} \exp(D_{gt}\delta_{gt}) = \exp\{\log \bar{b}_g + \gamma_{t0} + D_{gt}\delta_0\} = \exp\{\eta_{gt}(\theta^0)\}.$$

In particular $m_{gt}(\delta) > 0$ and

$$\log m_{gt}(\delta) = \eta_{gt}(\theta^0) \qquad \text{for every supported cell.}$$

Step 5: the scalar Poisson cell inequality. For every $m > 0$ and every real η,

$$m\eta - \exp(\eta) \leq m \log m - \exp(\log m),$$

with strict inequality when $\eta \neq \log m$. Apply $1 + x \leq \exp(x)$, strict for $x \neq 0$, at $x = \eta - \log m$, multiply by $m > 0$, and rearrange using $\exp(\eta) = m \exp(\eta - \log m)$.

Step 6: the cell masses are strictly positive. First $T > 0$: if $T = 0$, then for the nonzero collapsed parameter with zero nuisance block and treatment coordinate one the weighted sum $\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt}(r'_{gt}a)^2$ is an empty sum, hence zero, contradicting Assumption 5. Since $\pi_g \in (0, 1)$, it follows that

$$q_{gt} = \frac{\pi_g}{T} > 0 \qquad \text{for every supported cell.}$$

Step 7: θ^0 is a global maximizer. Fix any θ. Applying Step 5 cell by cell with $m = m_{gt}(\delta) > 0$ and $\eta = \eta_{gt}(\theta)$, and using $\log m_{gt}(\delta) = \eta_{gt}(\theta^0)$ from Step 4,

$$m_{gt}(\delta)\eta_{gt}(\theta) - \exp(\eta_{gt}(\theta)) \leq m_{gt}(\delta)\eta_{gt}(\theta^0) - \exp\big(\eta_{gt}(\theta^0)\big).$$

Multiplying by $q_{gt} > 0$ and summing over supported cells,

$$L(\theta;\delta) \le L(\theta^0;\delta).$$

Step 8: the maximizer is unique. Suppose $L(\theta;\delta) = L(\theta^0;\delta)$ and $\theta \neq \theta^0$. Then $\theta - \theta^0 \neq 0$, so Assumption 5 gives

$$\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}\left(\eta_{gt}(\theta-\theta^0)\right)^2 > 0,$$

and by linearity $\eta_{gt}(\theta-\theta^0) = \eta_{gt}(\theta) - \eta_{gt}(\theta^0)$, so some supported cell (g_0,t_0) has

$$\eta_{g_0t_0}(\theta) \neq \eta_{g_0t_0}(\theta^0) = \log m_{g_0t_0}(\delta).$$

At that cell the strict half of Step 5 applies, and all other cells obey the weak inequality of Step 7. Since $q_{g_0t_0} > 0$, summing gives

$$L(\theta;\delta) < L(\theta^0;\delta),$$

contradicting the assumed equality. Hence $\theta = \theta^0$.

Step 9: identification of the projection. By Steps 7 and 8 the criterion $L(\cdot;\delta)$ has θ^0 as its unique global maximizer, and Definition 3 selects $\theta^\star(\delta)$ among the global maximizers, so

$$\theta^\star(\delta) = \theta^0.$$

Step 10: the two conclusions. The treatment coordinate of θ^0 is δ_0, so

$$\beta^\star(\delta) = \delta_0,$$

and for every $g \in \mathcal{C}$ and every period t, by Step 2,

$$\mu^\star_{gt}(\delta) = \exp\{\eta_{gt}(\theta^\star(\delta))\} = \exp\{\log \bar{b}_g + \gamma_{t0} + D_{gt}\delta_0\} = B_{gt}\exp(D_{gt}\delta_0).$$

□

Proof of Theorem 3. Write $q_{gt} = \pi_g/T$ and $m_{gt} = m_{gt}(\delta) = B_{gt}\exp(D_{gt}\delta_{gt})$, with $B_{gt} = \bar{b}_g\exp(\gamma_{t0})$. Then the quantities of Definition 7 are

$$h_{gt} = \pi_g m_{gt} = T q_{gt} m_{gt}, \qquad R_g = \sum_{t=1}^{T} h_{gt}, \qquad C_t = \sum_{g\in\mathcal{C}} h_{gt},$$

$$M = \sum_{g\in\mathcal{C}} R_g, \qquad A = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} D_{gt}h_{gt}, \qquad \Phi = MA - \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} D_{gt}R_gC_t.$$

A collapsed parameter is written $\theta = (a,\beta)$, with nuisance block a and treatment coordinate β, and $r'_{gt}\theta = X'_{gt}a + D_{gt}\beta$. Finally set

$$B^{obs}_{gt} = \frac{m_{g1}m_{\infty t}}{m_{\infty 1}}, \qquad H = \{(g,t) : g \in \mathcal{C},\ D_{gt} = 1\}, \qquad Z = \sum_{(g,t)\in H} q_{gt}B^{obs}_{gt},$$

$$\tau_{gt} = \frac{m_{gt}}{B^{obs}_{gt}} - 1, \qquad \omega_{gt} = \frac{q_{gt}B^{obs}_{gt}}{Z}, \qquad PTT = \sum_{(g,t)\in H} \omega_{gt}\tau_{gt}.$$

Step 1: positivity of the primitive margins. Since $T \geq 4$, we have $T > 0$, and since $\infty \in \mathcal{C}$, the support $\mathcal{C}$ is nonempty. Every $h_{gt} = \pi_g \bar{b}_g \exp(\gamma_{t0}) \exp(D_{gt}\delta_{gt})$ is a product of strictly positive factors. Hence

$$h_{gt} > 0, \qquad R_g > 0, \qquad C_t > 0, \qquad M > 0.$$

Also, summing h_{gt} in the two possible orders gives

$$\sum_{t=1}^{T} C_t = M = \sum_{g \in \mathcal{C}} R_g.$$

Step 2: the zero-treatment row-column fit. Define the collapsed parameter $\theta^0 = (a^0, 0)$, whose treatment coordinate is zero and whose nuisance levels are

$$a^0_{\text{int}} = \log \frac{R_\infty}{\pi_\infty} + \log C_1 - \log M, \qquad a^0_{\text{coh},g} = \log \frac{R_g}{\pi_g} - \log \frac{R_\infty}{\pi_\infty} \quad (g \neq \infty),$$

$$a^0_{\text{time},t} = \log C_t - \log C_1 \quad (t \neq 1).$$

Adding the intercept, cohort, and period contributions, with the cohort dummy absent for $g = \infty$ and the period dummy absent for $t = 1$, gives, for every supported cell,

$$r'_{gt}\theta^0 = \log \frac{R_g}{\pi_g} + \log C_t - \log M, \qquad \exp(r'_{gt}\theta^0) = \frac{R_g C_t}{\pi_g M},$$

where the logarithms and division use the positivity in Step 1.

Step 3: the conditional residuals and their margins. Define the mass-weighted residual of θ^0 in cell (g, t) by

$$e_{gt} = q_{gt}\{m_{gt} - \exp(r'_{gt}\theta^0)\} = \frac{\pi_g}{T}\left\{m_{gt} - \frac{R_g C_t}{\pi_g M}\right\} = \frac{1}{T}\left\{h_{gt} - \frac{R_g C_t}{M}\right\}.$$

Using $\sum_t C_t = M$ and $\sum_{g \in \mathcal{C}} R_g = M$ from Step 1, its row and column sums vanish:

$$\sum_{t=1}^{T} e_{gt} = \frac{1}{T}\left\{R_g - \frac{R_g}{M}\sum_{t=1}^{T} C_t\right\} = 0 \qquad (g \in \mathcal{C}),$$

$$\sum_{g \in \mathcal{C}} e_{gt} = \frac{1}{T}\left\{C_t - \frac{C_t}{M}\sum_{g \in \mathcal{C}} R_g\right\} = 0 \qquad (t = 1, \ldots, T).$$

Step 4: the scores at θ^0. Every collapsed nuisance regressor $X_{\cdot,j}$ is either the constant 1, the indicator of one cohort $g_0 \neq \infty$, or the indicator of one period $t_0 \neq 1$. The corresponding nuisance score at θ^0 is therefore $\sum_{g,t} e_{gt}$, $\sum_t e_{g_0 t}$, or $\sum_g e_{g t_0}$, and these vanish by Step 3:

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} X_{gt,j}\{m_{gt} - \exp(r'_{gt}\theta^0)\} = 0 \qquad \text{for every nuisance coordinate } j.$$

Because the nuisance block is finite-dimensional and $X'_{gt}u$ is the corresponding linear combination of the coordinate regressors, the same conclusion holds for every nuisance direction u:

$$\sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt}(X'_{gt}u)\{m_{gt} - \exp(r'_{gt}\theta^0)\} = \sum_j u_j \sum_{g \in \mathcal{C}} \sum_{t=1}^{T} q_{gt} X_{gt,j}\{m_{gt} - \exp(r'_{gt}\theta^0)\} = 0.$$

The treatment score at θ^0 is

$$S = \sum_{g\in\mathcal{C}}\sum_{t=1}^{T} q_{gt}D_{gt}\{m_{gt} - \exp(r'_{gt}\theta^0)\} = \frac{1}{T}\left\{A - \frac{1}{M}\sum_{g\in\mathcal{C}}\sum_{t=1}^{T} D_{gt}R_gC_t\right\} = \frac{\Phi}{TM},$$

and $TM > 0$ by Step 1.

Step 5: the pseudo-true coefficient has the sign of S. The finite-dimensional Poisson sign comparison used here is the following. For a finite index set with strictly positive weights q_i, strictly positive means m_i, and an injective linear design map $\mathscr{A}$, suppose that a zero-treatment point $(u^0, 0)$ clears every nuisance-direction score,

$$\sum_i q_i\mathscr{A}(u,0)_i\{m_i - \exp(\mathscr{A}(u^0,0)_i)\} = 0 \qquad \text{for every nuisance direction } u.$$

If

$$s_0 = \sum_i q_i\mathscr{A}(0,1)_i\{m_i - \exp(\mathscr{A}(u^0,0)_i)\}$$

and β is the treatment coordinate of the unique maximizer of $\sum_i q_i\{m_i(\mathscr{A}\theta)_i - \exp((\mathscr{A}\theta)_i)\}$, then

$$\beta < 0 \iff s_0 < 0, \qquad \beta = 0 \iff s_0 = 0, \qquad 0 < \beta \iff 0 < s_0.$$

Indeed, the score at a maximizer vanishes in every direction. Writing $\widehat{\theta}$ for the maximizer, $\theta^0 = (u^0, 0)$, and $v = \widehat{\theta} - \theta^0$, the score identities and the nuisance-direction condition give

$$\widehat{\beta}s_0 = \sum_i q_i(\mathscr{A}v)_i\{\exp((\mathscr{A}\widehat{\theta})_i) - \exp((\mathscr{A}\theta^0)_i)\}.$$

The summands on the right are nonnegative, and at least one is strictly positive when $v \neq 0$, by injectivity of $\mathscr{A}$ and strict monotonicity of the exponential. If $s_0 = 0$, all scores at θ^0 vanish, and the tangent-line inequality for the exponential makes θ^0 globally maximizing; uniqueness gives the zero treatment coordinate.

Apply this comparison with index set $\mathcal{C}\times\{1,\dots,T\}$, design $\mathscr{A}\theta = (r'_{gt}\theta)_{g,t}$, and $(u^0, 0) = \theta^0$. The weights and means are positive by Step 1, Step 4 supplies the nuisance-direction score condition, and Assumption 5 gives injectivity: if two collapsed parameters have the same fitted-index array, their difference has weighted sum of squared indices equal to zero, so the rank condition forces the difference to be zero. Therefore the treatment coordinate $\beta^\star(\delta)$ of $\theta^\star(\delta)$ has the same three-way sign as S. Since $S = \Phi/(TM)$ with $TM > 0$,

$$\beta^\star(\delta) < 0 \iff \Phi < 0, \qquad \beta^\star(\delta) = 0 \iff \Phi = 0, \qquad 0 < \beta^\star(\delta) \iff 0 < \Phi.$$

Step 6: the observed baseline proxy reproduces the untreated mean. For every supported cohort g, the first-period treatment indicator is zero: if $g = \infty$ this is the never-treated path, and if g is finite then $D_{g1} = 1$ would force the adoption date to be the first period, which the support excludes. Also $D_{\infty t} = 0$ for every t, and Assumption 2 gives $\gamma_{10} = 0$. Hence, for every $g \in \mathcal{C}$ and every t,

$$B^{obs}_{gt} = \frac{m_{g1}m_{\infty t}}{m_{\infty 1}} = \frac{\bar{b}_g e^{\gamma_{10}}\cdot\bar{b}_\infty e^{\gamma_{t0}}}{\bar{b}_\infty e^{\gamma_{10}}} = \bar{b}_g e^{\gamma_{t0}} = B_{gt},$$

where $m_{\infty 1} = \bar{b}_\infty > 0$ justifies the cancellation.

Step 7: H is nonempty and $Z > 0$. Assumption 6 supplies a supported finite cohort adopting in the second period; in that period its treatment indicator equals one, so H is nonempty. For every $(g,t) \in H$, $q_{gt} > 0$, and Step 6 gives $B^{obs}_{gt} = B_{gt} > 0$. Hence every summand of Z is nonnegative and at least one summand is strictly positive:

$$Z = \sum_{(g,t)\in H} q_{gt} B^{obs}_{gt} > 0.$$

Step 8: membership in $\mathcal{R}_T$ is equivalent to $\Phi < 0$. If the induced primitive tuple lies in $\mathcal{R}_T$, then the last clause of Definition 5 gives negativity of its treatment coordinate. In the restricted primitive coordinates the limiting criterion is the same collapsed criterion $L(\theta;\delta)$, so this coordinate equals $\beta^\star(\delta)$, and Step 5 gives $\Phi < 0$. Conversely suppose $\Phi < 0$. The horizon, support, and frontier-scope clauses of Definition 5 are hypotheses of the present theorem. The simplex clause holds because the finite-array shares sum to one for every $N \geq |\mathcal{C}|$,

$$\sum_{g\in\mathcal{C}} \frac{n_{gN}}{N} = 1,$$

every unit label lies in $\mathcal{C}$, and Assumption 1 gives $\sum_{g\in\mathcal{C}} \pi_g = 1$ after passage to the limit. The untreated-mean clause is witnessed by $g \mapsto \bar{b}_g$, since the primitive untreated coordinate is $B_{gt} = \bar{b}_g e^{\gamma_{t0}}$. The positive-effect clause is Assumption 7. The rank clause is Assumption 5, after rewriting the sum over supported cells as the double sum over $g \in \mathcal{C}$ and t. Finally Step 5 gives $\beta^\star(\delta) < 0$, which is the tuple's treatment coordinate. Hence the tuple lies in $\mathcal{R}_T$.

Step 9: the four-cohort polynomial. Let $x > 1$, $y > 1$, and consider the four-period support $\{2,3,4,\infty\}$ with $T = 4$, equal shares $\pi_g = 1/4$, unit limiting baselines, zero untreated time effects, multiplier x on every treated cell except $(2,4)$, and multiplier y at $(2,4)$. With treatment absorbing from the adoption date on, $D_{gt} = 1$ exactly when $g \leq t$ and $D_{\infty t} = 0$, so

$$4\,(h_{gt}) = \begin{pmatrix} 1 & x & x & y \\ 1 & 1 & x & x \\ 1 & 1 & 1 & x \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

with rows $g = 2,3,4,\infty$ and columns $t = 1,2,3,4$. Summing rows and columns,

$$4R_2 = 1 + 2x + y, \quad 4R_3 = 2 + 2x, \quad 4R_4 = 3 + x, \quad 4R_\infty = 4,$$

$$4C_1 = 4, \quad 4C_2 = x + 3, \quad 4C_3 = 2x + 2, \quad 4C_4 = 2x + y + 1,$$

$$4M = 10 + 5x + y, \qquad 4A = 5x + y.$$

The treated cells of row $g = 2$ are $t = 2,3,4$, those of row $g = 3$ are $t = 3,4$, and the treated cell of row $g = 4$ is $t = 4$, so

$$16 \sum_{g,t} D_{gt} R_g C_t = (1 + 2x + y)(5x + y + 6) + (2 + 2x)(4x + y + 3) + (3 + x)(2x + y + 1),$$

which expands to $20x^2 + 10xy + y^2 + 38x + 12y + 15$, while

$$16\,MA = (10 + 5x + y)(5x + y) = 25x^2 + 10xy + y^2 + 50x + 10y.$$

Subtracting gives

$$\Phi = MA - \sum_{g,t} D_{gt} R_g C_t = \frac{5x^2 + 12x - 2y - 15}{16}.$$

Step 10: the numerical identity. With $5(101/100)^2 = 51005/10000$ and $12(101/100) = 1212/100$,

$$\frac{5(101/100)^2 + 12(101/100) - 15}{2} = \frac{4441}{4000}.$$

Step 11: the value at the witness. For W_4 of Definition 6, the treated multipliers are $\exp(\log(101/100)) = 101/100$ off cell $(2,4)$ and $\exp(\log 4) = 4$ at $(2,4)$, so Step 9 applies with $x = 101/100$ and $y = 4$:

$$\Phi = \frac{5(101/100)^2 + 12(101/100) - 2 \cdot 4 - 15}{16} = -\frac{11559}{32000}.$$

Step 12: the counterfactual-share objects on treated cells. Let $(g,t) \in H$, so $D_{gt} = 1$. Step 6 gives $B_{gt}^{obs} = B_{gt}$, and $B_{gt} > 0$, so

$$\tau_{gt} = \frac{m_{gt}}{B_{gt}^{obs}} - 1 = \frac{B_{gt} \exp(\delta_{gt})}{B_{gt}} - 1 = \exp(\delta_{gt}) - 1.$$

Moreover $q_{gt} > 0$, $B_{gt}^{obs} > 0$, and $Z > 0$ by Step 7, hence

$$\omega_{gt} = \frac{q_{gt} B_{gt}^{obs}}{Z} > 0.$$

Step 13: the weights sum to one. By the definition of Z and the positivity $Z > 0$,

$$\sum_{(g,t) \in H} \omega_{gt} = \frac{1}{Z} \sum_{(g,t) \in H} q_{gt} B_{gt}^{obs} = \frac{Z}{Z} = 1.$$

Step 14: the ratio form of PTT. Expanding the definition and using $B_{gt}^{obs} > 0$ on every treated cell,

$$PTT = \sum_{(g,t) \in H} \frac{q_{gt} B_{gt}^{obs}}{Z} \left(\frac{m_{gt}}{B_{gt}^{obs}} - 1 \right) = \frac{\sum_{(g,t) \in H} q_{gt} m_{gt}}{Z} - \frac{\sum_{(g,t) \in H} q_{gt} B_{gt}^{obs}}{Z}.$$

The second fraction equals 1, so

$$PTT = \frac{\sum_{(g,t) \in H} q_{gt} m_{gt}}{\sum_{(g,t) \in H} q_{gt} B_{gt}^{obs}} - 1.$$

Step 15: the sign contrast. Suppose $\Phi < 0$. Step 5 gives $\beta^\star(\delta) < 0$. For PTT, each treated cell has $\omega_{gt} > 0$ by Step 12, and Assumption 7 gives $\delta_{gt} > 0$, hence $\exp(\delta_{gt}) > 1$ and $\tau_{gt} > 0$. The sum defining PTT has nonnegative terms and, by Step 7, at least one strictly positive term. Therefore $PTT > 0$, and

$$\Phi < 0 \quad \Longrightarrow \quad \beta^\star(\delta) < 0 < PTT.$$

□

References

Alberto Abadie. Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19, 2005. doi: 10.1111/0034-6527.00321.

James E. Anderson and Eric van Wincoop. Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192, 2003. doi: 10.1257/000282803321455214.

Donald W. K. Andrews. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, 4(3):458–467, 1988. doi: 10.1017/S0266466600013396.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ, 2009. ISBN 9780691120355.

Orley Ashenfelter and David Card. Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67(4):648–660, 1985. doi: 10.2307/1924810.

Susan Athey and Guido W. Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 2006. doi: 10.1111/j.1468-0262.2006.00668.x.

Andrew Baker, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. Difference-in-differences designs: A practitioner's guide, 2025.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6):3253–3285, 2024. doi: 10.1093/restud/rdae007.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. doi: 10.1016/j.jeconom.2020.12.001.

A. Colin Cameron and Pravin K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, 2 edition, 2013. ISBN 9781107014169. doi: 10.1017/CBO9781139013567.

Sergio Correia, Paulo Guimarães, and Tom Zylkin. Fast poisson estimation with high-dimensional fixed effects. *The Stata Journal*, 20(1):95–115, 2020. doi: 10.1177/1536867X20909691.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020. doi: 10.1257/aer.20181169.

John Gardner. Two-stage differences in differences, 2022.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021. doi: 10.1016/j.jeconom.2021.03.014.

Christian Gouriéroux, Alain Monfort, and Alain Trognon. Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica*, 52(3):701–720, 1984a. doi: 10.2307/1913472.

Christian Gouriéroux, Alain Monfort, and Alain Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700, 1984b. doi: 10.2307/1913471.

Jerry Hausman, Bronwyn H. Hall, and Zvi Griliches. Econometric models for count data with an application to the patents-r & d relationship. *Econometrica*, 52(4):909–938, 1984. doi: 10.2307/1911191.

Keith Head and Thierry Mayer. Gravity equations: Workhorse,toolkit, and cookbook. In Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, editors, *Handbook of International Economics*, volume 4, pages 131–195. Elsevier, 2014. doi: 10.1016/B978-0-444-54314-1.00003-3.

Guido W. Imbens and Jeffrey M. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86, 2009. doi: 10.1257/jel.47.1.5.

Ninon Moreau-Kastler. Proportional treatment effects in staggered settings: An approach for poisson pseudo-maximum likelihood. Working Paper 31, EU Tax Observatory, 2025. URL https://doi.org/10.2139/ssrn.5247936.

Arne J. Nagengast and Yoto V. Yotov. Staggered difference-in-differences in gravity settings: Revisiting the effects of trade agreements. *American Economic Journal: Applied Economics*, 17(1):271–296, 2025. doi: 10.1257/app.20230089.

Jonathan Roth and Pedro H. C. Sant'Anna. When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747, 2020. doi: 10.3982/ECTA19478.

Jonathan Roth, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023. doi: 10.1016/j.jeconom.2023.03.008.

Pedro H. C. Sant'Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020. doi: 10.1016/j.jeconom.2020.06.003.

J. M. C. Santos Silva and Silvana Tenreyro. The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658, 2006. doi: 10.1162/rest.88.4.641.

J. M. C. Santos Silva and Silvana Tenreyro. Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222, 2011. doi: 10.1016/j.econlet.2011.05.008.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021. doi: 10.1016/j.jeconom.2020.09.006.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982. doi: 10.2307/1912526.

Jeffrey M Wooldridge. Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26(3):C31–C66, 2023. doi: 10.1093/ectj/utad016.

Yoto V. Yotov, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. World Trade Organization and United Nations Conference on Trade and Development, Geneva, 2016. ISBN 9789287043672. URL https://www.wto.org/english/res_e/booksp_e/advancedwtounctad2016_e.pdf.