DO SUNSPOT CYCLES CAUSALLY AFFECT SOCIAL TENSIONS AND POPULATION HARM IN DEVELOPING COUNTRIES?

Anonymous authors
Paper under review

ABSTRACT

We study whether sunspot activity contributes actionable information about social tension outcomes in developing-country panels under modern causal-identification constraints. Building on staggered-adoption estimand logic, we formulate an eventized continuous-dose aggregation framework with explicit support and contamination diagnostics. We then introduce a robust-null acceptance predicate that jointly audits dataset harmonization, exposure-version sensitivity, and placebo boundedness, and we connect causal interpretation to constrained subset deployment through a comparator-parity utility decomposition. Formal analysis establishes identification and admissibility results for the weighted estimand, robust-null rule, and subset utility identity. Validation artifacts with symbolic theorem checks, dynamic event-time diagnostics, lattice stress tests, and parity-calibrated forecasting outputs show coherent estimator mechanics and conservative mixed-support interpretation under uncertainty. The key contribution is an auditable inference-to-decision workflow that preserves negative evidence while identifying where bounded, subset-specific utility can remain defensible.

1 INTRODUCTION

Solar variability is one of the longest-observed geophysical processes, yet its relevance for social instability remains contested. The empirical challenge is not only whether solar cycles correlate with conflict-related outcomes, but whether any detected relationship survives modern causal-design diagnostics under severe confounding risk. In developing-country panels, macroeconomic stress, governance shifts, commodity shocks, climate anomalies, and conflict spillovers all move concurrently with social tension proxies, so naive trend comparisons are not scientifically interpretable (Hsiang et al., 2013; Burke et al., 2015; Mach et al., 2019; Goodman-Bacon, 2021; Sun & Abraham, 2021; de Chaisemartin & D'Haultfœuille, 2022). At the same time, early-warning systems increasingly require calibrated uncertainty and operational utility evidence, not only statistical significance (Hegre et al., 2019; Rød et al., 2024; Jaime et al., 2024). The central question of this paper is therefore dual: can solar activity support defensible causal bounds for social tension outcomes, and can solar information improve decision-relevant forecasting under strict comparator parity?

This question matters beyond the specific solar-conflict debate. If weak environmental signals are evaluated with inappropriate estimators, policy systems can overreact to artifacts; if they are dismissed without robust design, potentially useful risk modifiers are lost. The scientific contribution is thus a methodology for high-uncertainty domains where robust null results are as informative as positive effects. Our formulation unifies three traditionally disconnected components: heterogeneity-robust causal identification, conservative robust-null decision logic under measurement and version uncertainty, and subset-aware utility optimization for deployment decisions. Together, these components create an auditable bridge between causal inference and operational forecasting in low signal-to-noise environments (Rød et al., 2024; Jaime et al., 2024; Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025).

The paper delivers four contributions:

- We formulate an eventized continuous-dose estimand that adapts modern staggered-adoption DID logic to sunspot intensity while preserving explicit support and contamination diagnostics.
- We define a robust-null acceptance predicate that combines dataset harmonization, exposure-version sensitivity, and placebo boundedness into a single conservative inferential rule.
- We derive a comparator-parity utility decomposition showing when aggregate near-null predictive gains can coexist with positive subset-level decision utility under no-harm constraints.
- We provide integrated symbolic and computational evidence linking formal statements to executable diagnostics, figures, and tables in a reproducible validation pipeline.

The scope is intentionally disciplined. We focus on total-effect bounds and deployment-relevant interpretation, while treating direct-versus-mediated pathway decomposition as conditional and unresolved under currently available mediator support. This design choice follows literature evidence that climate-conflict mechanisms are context dependent and that event-coding differences can dominate weak covariate effects (Burke et al., 2015; Mach et al., 2019; Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025). The resulting manuscript is therefore a hybrid formal-empirical contribution: theorems and proofs establish what can be concluded under explicit assumptions, and computational artifacts show where evidence remains mixed.

2 Related Work and Gap Synthesis

2.1 Environmental Risk, Conflict, and Effect Magnitude

A broad literature links environmental stressors to conflict outcomes, but with substantial heterogeneity in effect size and mechanism (Hsiang et al., 2013; Burke et al., 2015; Hsiang et al., 2011; Schleussner et al., 2016; Maystadt & Ecker, 2014; McGuirk & Nunn, 2024; von Uexkull et al., 2023). Meta-analytic and synthesis studies find detectable climate-related risk shifts, yet also emphasize that governance capacity, state fragility, and economic structure are often first-order drivers (Mach et al., 2019; Burke et al., 2015). This creates a methodological tension: detectable coefficients may be real but policy-irrelevant, and apparent nulls may reflect measurement noise rather than absence of effect. Our paper addresses this tension by centering bounded interpretation and robust-null admissibility rather than binary "significant/non-significant" framing.

Solar exposure introduces an additional layer of uncertainty: the underlying physical index is informative but versioned and periodically revised (Hathaway, 2015; Muscheler et al., 2016; Gkana & Zachilas, 2016; Petrovay, 2010). Prior social-science studies rarely propagate this version uncertainty through causal estimands and forecast decisions. We treat exposure-version uncertainty as a first-class inferential axis, not a post hoc robustness note.

2.2 Identification Under Staggered and Heterogeneous Effects

Recent econometric work shows that legacy two-way fixed-effects event studies can be biased under treatment-effect heterogeneity and contaminated comparisons (Goodman-Bacon, 2021; Sun & Abraham, 2021; Borusyak et al., 2024; de Chaisemartin & D'Haultfœuille, 2022; Imai & Kim, 2020; Arkhangelsky et al., 2019). Weighted stacked designs and group-time estimands provide cleaner interpretation when untreated support is explicit and aggregation weights are transparent (Wing et al., 2024; Freedman et al., 2023). Continuous-treatment event studies extend this framework but still require careful aggregation, support diagnostics, and event-window design (Callaway et al., 2024; Li & Lin, 2024). Our first technical contribution builds on this estimator lineage: we borrow the group-time ATT foundation from prior work and introduce a manuscript-specific eventization operator for continuous solar dose with threshold-family auditing.

2.3 Measurement Uncertainty and Forecast Evaluation Standards

Conflict-event datasets differ in coding rules, actor definitions, spatial granularity, and revision practices, and these differences can change signs or magnitudes of estimated effects (Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025). Forecasting research likewise emphasizes that calibration and decision utility can diverge from discrimination metrics, especially in rare-event domains (Hegre et al., 2019; Rød et al., 2024; Racek et al., 2024; Jaime et al., 2024; 2022; Gudoshava et al., 2022; Herteux et al., 2024). The literature therefore motivates a gap: we need a unified framework that simultaneously handles identification validity, measurement/version uncertainty, and deployment utility constraints. Existing studies usually optimize one axis and relegate the others to supplementary diagnostics. This paper places all three axes inside the core inferential object.

2.4 Novelty Boundary

This manuscript does not claim a new physical theory of sunspot dynamics, nor does it claim definitive direct pathway decomposition. Instead, it contributes a decision-theoretic causal-audit framework at the boundary between modern panel identification and operational risk forecasting. Borrowed conventions are cited at first use (ATT logic, stacked DID validity, multiway uncertainty, forecast-scoring principles), while manuscript-defined objects are explicitly marked as new. The novelty is therefore methodological integration and auditability under uncertainty, not domain overreach.

2.5 Contradiction Map and Comparator Logic

The literature exposes four contradiction families that directly motivate our design. First, climate-conflict meta-analyses report positive average associations, while expert syntheses rank environmental factors below political and economic drivers in practical salience; this means a statistically nonzero coefficient is not automatically policy relevant (Hsiang et al., 2013; Burke et al., 2015; Mach et al., 2019). Second, staggered-treatment estimators differ materially in their susceptibility to weighting and contamination problems, so "event-study" as a label is not enough to guarantee causal interpretability (Goodman-Bacon, 2021; Sun & Abraham, 2021; de Chaisemartin & D'Haultfœuille, 2022; Wing et al., 2024; Freedman et al., 2023). Third, conflict-event data sources can produce divergent empirical narratives because coding protocols differ, especially for low-intensity and remote events (Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025). Fourth, exposure-series revisions can alter magnitude rankings and practical interpretation, so conclusions conditioned on one version alone are fragile (Hathaway, 2015; Muscheler et al., 2016; Gkana & Zachilas, 2016; Petrovay, 2010).

Our comparator strategy is designed around those contradictions. For causal estimation, we pair a heterogeneity-robust lineage with a legacy benchmark and an explicit no-solar ablation so that effect claims are interpreted against both modern identification expectations and plausible null baselines. For robustness, we add invalid-control and placebo counterexamples not as optional stress tests, but as direct falsification channels that can demote claims. For predictive assessment, we enforce comparator parity so that model-class differences are not confounded by split leakage or calibration inconsistency. The central methodological principle is that every major contradiction should map to an explicit comparator family and a corresponding rejection path. This comparator discipline is what turns a broad mixed literature into a coherent scientific test.

3 Problem Setting and Notation

We study a country-time panel indexed by unit $i \in \mathcal{I}$ and period $t \in \mathcal{T}$. Let Y_{it} denote a social-tension outcome (incidence, count, or severe-threshold indicator), $D_{it} \in \mathbb{R}$ denote continuous solar dose, and X_{it} denote pre-specified controls (macroeconomic, political, and climatic). Following heterogeneity-robust DID conventions (Goodman-Bacon, 2021; Wing et al., 2024; Freedman et al., 2023), we define threshold families $\kappa \in \mathcal{K}$ and eventized cohorts G_i^κ based on first crossing events.

The paper pursues two linked objectives. First, estimate a bounded aggregate causal estimand for dynamic event windows $\mathcal{E}$:

$$\Theta_{\kappa,\mathcal{E}} = \sum_{e \in \mathcal{E}} \sum_{g \in \mathcal{G}_{\kappa,e}} \omega_{g,e}^{(\kappa)} \, ATT_\kappa(g,e), \tag{1}$$

where weights satisfy $\omega_{g,e}^{(\kappa)} \geq 0$ and $\sum_{e,g} \omega_{g,e}^{(\kappa)} = 1$. Second, optimize deployment relevance using subset utility while enforcing no-harm constraints; the decision variables are subset activation indicators $z_s \in \{0,1\}$.

To make the optimization target explicit, we define a composite objective over design and deployment decisions $x = (\kappa, \mathcal{E}, \delta, \tau, \{z_s\}_{s \in \mathcal{S}})$:

$$\max_{x \in \mathcal{F}} \; \mathcal{J}(x) = \lambda_1 \, \underline{S}(x) - \lambda_2 \, \underline{P}(x) + \lambda_3 \, \underline{U}(x), \tag{2}$$

where $\underline{S}$ is minimum support quality, $\underline{P}$ is worst-case placebo deviation, and $\underline{U}$ is lower-confidence deployment utility. The feasible set $\mathcal{F}$ encodes assumptions: design-level exogeneity, valid untreated support, measurement harmonization consistency, exposure-version transparency, comparator parity, subset pre-registration, and immutable evaluation splits.

The optimality criterion is conservative: a design is preferred only if it improves $\mathcal{J}$ without violating inferential admissibility. This criterion allows robust-null outcomes to be optimal when support or invariance conditions weaken. In other words, non-actionability is a valid optimum when constraints dominate uncertain gains.

3.1 Assumption Hierarchy and Failure Semantics

Assumptions are organized as a hierarchy rather than a flat checklist. Design assumptions (exogeneity, valid support, contamination control) determine whether causal estimands are interpretable at all. Measurement assumptions (harmonization consistency across event datasets and transparent exposure versioning) determine whether estimates can be compared across evidence axes. Decision assumptions (comparator parity, subset pre-registration, immutable splits) determine whether forecast utility differences are operationally meaningful rather than procedural artifacts.

This hierarchy induces explicit failure semantics. If design assumptions fail, causal claims are unsupported regardless of numerical estimates. If measurement assumptions fail, directional claims are downgraded to mixed support because cross-axis invariance is unresolved. If decision assumptions fail, deployment recommendations are withheld even when some metrics look favorable. These semantics are necessary in a domain where weak effects and nonstationary measurement processes can otherwise reward selective reporting. They also align with the robust-null philosophy embedded in equation 4: inferential caution is not post hoc skepticism, but a formally encoded consequence of assumption-aware testing.

To keep provenance explicit, we distinguish borrowed and manuscript-defined formal objects. Borrowed: ATT notation, staggered DID identification, and multiway uncertainty conventions (Goodman-Bacon, 2021; Sun & Abraham, 2021; de Chaisemartin & D'Haultfœuille, 2022; Cameron et al., 2011; Wing et al., 2024; Freedman et al., 2023). Defined here: threshold-family eventization operator, robust-null conjunction over dataset-version-outcome-event cells, and constrained subset deployment criterion. This distinction prevents ambiguity about novelty and allows readers to audit where this manuscript extends prior methods versus where it adopts established practice.

4 Methodology

4.1 Eventized Continuous-Dose Estimation

For each threshold κ, we construct cohort-event cells and estimate valid-comparison DID components:

$$\widehat{ATT}_\kappa(g,e) = \left(\bar{Y}^{\text{treated}}_{g,e} - \bar{Y}^{\text{treated}}_{g,-1}\right) - \left(\bar{Y}^{\text{valid-control}}_{g,e} - \bar{Y}^{\text{valid-control}}_{g,-1}\right), \tag{3}$$

then aggregate using equation 1. The estimator lineage is borrowed from group-time ATT and stacked DID work (Goodman-Bacon, 2021; Wing et al., 2024; Freedman et al., 2023), while the threshold-family eventization map is defined in this manuscript to handle continuous dose.

Definition 4.1 (Weighted Eventized Estimand)**.** *Given threshold family $\mathcal{K}$ and event window family $\mathfrak{E}$, we define the reported causal bound as*

$$\Theta^\star = \arg\max_{\kappa\in\mathcal{K},\,\mathcal{E}\in\mathfrak{E}} \left\{\underline{S}(\kappa,\mathcal{E}) \;:\; \underline{S}(\kappa,\mathcal{E}) \geq s_{\min}\right\},$$

with corresponding value computed by equation 1.

Theorem 4.2 (Identification Under Valid Support)**.** *Assume cohort-specific parallel trends, valid untreated support for all retained (g,e) cells, and no treated-vs-treated contamination in comparison sets. Then the weighted estimator in equation 1 is identified by the expectation of valid-comparison DID components from equation 3.*

Proof. Under the assumptions, each retained cell estimator satisfies $\mathbb{E}[\widehat{ATT}_\kappa(g,e)] = ATT_\kappa(g,e)$. Linearity of expectation yields

$$\mathbb{E}\left[\sum_{e,g} \omega^{(\kappa)}_{g,e} \widehat{ATT}_\kappa(g,e)\right] = \sum_{e,g} \omega^{(\kappa)}_{g,e} \mathbb{E}[\widehat{ATT}_\kappa(g,e)] = \sum_{e,g} \omega^{(\kappa)}_{g,e} ATT_\kappa(g,e) = \Theta_{\kappa,\mathcal{E}}.$$

Because weights are nonnegative and sum to one, aggregation preserves the identified target and does not introduce sign-reversal artifacts from negative weighting. Therefore the weighted estimator identifies equation 1. □

4.2 Robust-Null Decision Rule

Conflict measurement and exposure versioning induce a lattice of estimates indexed by dataset d, exposure version v, outcome family q, and event index e. We define a conservative robust-null predicate:

$$\mathcal{R}_{\text{null}}(\delta,\alpha) = \mathbf{1}\left\{\max_{d,v,q,e} |\hat{\beta}_{d,v,q,e}| \leq \delta,\; 0 \in CI^{1-\alpha}_{d,v,q,e}\;\forall(d,v,q,e),\; \max_{d,v,q,e\in\mathcal{P}} |\hat{\beta}_{d,v,q,e}| \leq \delta_{pl}\right\}. \tag{4}$$

The covariance composition follows multiway clustering conventions (Cameron et al., 2011), while the conjunctive predicate in equation 4 is manuscript-defined.

Theorem 4.3 (Admissibility of Strong Non-Null Claims)**.** *If $\mathcal{R}_{null}(\delta,\alpha) = 1$, then no strong directional non-null claim is admissible for the audited lattice.*

Proof. When $\mathcal{R}_{\text{null}}(\delta,\alpha) = 1$, each conjunct in equation 4 holds simultaneously: absolute coefficients are bounded by δ, all confidence intervals include zero, and placebo windows are bounded by δ_{pl}. Any strong directional claim requires at least one audited cell with directional separation from zero that violates one of these three conditions. Since such violation is excluded by the predicate, strong directional non-null claims are inadmissible. □

Algorithm 1 Hybrid Causal-Utility Audit Workflow

1: Build harmonized outcome families and versioned solar exposures.
2: Eventize continuous dose into threshold families and construct valid control sets.
3: Estimate cell-level effects, aggregate via equation 1, and run support/contamination diagnostics.
4: Evaluate robust-null lattice using equation 4 over dataset-version-outcome-event axes.
5: Train comparator-parity forecast models and compute subset utilities via equation 5.
6: Solve constrained deployment problem in equation 6.
7: Report claim status as supported, mixed, or unsupported with explicit caveats.

4.3 Comparator-Parity Utility Decomposition

To evaluate operational relevance, we compare baseline and solar-augmented forecasts under strict parity in targets, splits, and calibration pipeline. Let $\Delta U_{h,s}(\tau)$ denote horizon-h utility difference on subset s at threshold τ, and let subset weights satisfy $\sum_s \pi_s = 1$.

$$\Delta U_h(\tau) = \sum_{s \in \mathcal{S}} \pi_s \Delta U_{h,s}(\tau). \tag{5}$$

Deployment decisions solve

$$\max_{z_s \in \{0,1\}} \sum_{s \in \mathcal{S}} \pi_s z_s \Delta U_{h,s}(\tau) \quad \text{s.t.} \quad \Delta L_h \leq \varepsilon,\ \underline{CI}(\Delta U_{h,s}(\tau)) > 0 \text{ for } z_s = 1. \tag{6}$$

Lemma 4.4 (Aggregate-Subset Compatibility)**.** *Under partition weights summing to one, equation 5 implies that aggregate near-null utility is compatible with positive utility on selected subsets.*

Proof. Equation equation 5 is a weighted average identity. A weighted average can be near zero while some components are positive and others negative, provided compensation occurs. Therefore aggregate near-null performance does not preclude positive subset utility. Constraint equation 6 then restricts deployment to subsets with positive lower confidence bounds and no-harm loss bounds. □

4.4 Integrated Audit Workflow

Algorithm 1 summarizes how we connect formal assumptions, estimator diagnostics, and deployment conclusions. The workflow is designed so each inferential transition has an explicit audit artifact.

5 Experimental Protocol and Evidence Assembly

The validation package executes three experiment families aligned to the methodological components above: eventized causal estimation, robust-null lattice stress testing, and comparator-parity forecasting. Each family uses five seeds and pre-registered sweeps over event windows, lag families, support trimming thresholds, exposure versions, inference modes, forecast horizons, split strategies, calibration methods, and deployment thresholds. The design follows modern recommendations on estimator heterogeneity, placebo diagnostics, and forecast calibration reporting (Goodman-Bacon, 2021; Sun & Abraham, 2021; de Chaisemartin & D'Haultfœuille, 2022; Cameron et al., 2011; Callaway et al., 2024; Wing et al., 2024; Freedman et al., 2023; Hegre et al., 2019; Rød et al., 2024; Jaime et al., 2024).

Evaluation artifacts include vector figures, publication-ready tables, and machine-readable data exports. We treat symbolic theorem checks and computational diagnostics as complementary evidence streams: symbolic checks validate algebraic obligations, while computational outputs assess finite-sample and design-robustness behavior. This distinction is important because exact algebraic identities can hold even when empirical support is mixed.

The protocol also enforces negative-result preservation. Counterexample failures, invariance violations, and exclusion logs are retained as first-class outputs, not hidden diagnostics. This is crucial in low-signal causal settings where optimistic selection can otherwise dominate interpretation. The resulting evidence mode is therefore conservative by design: claims are downgraded whenever lattice invariance or support quality weakens.

Two implementation choices are especially important for interpretation quality. First, every claim is linked to both a primary comparator and a falsification route. This prevents ambiguous "robustness" language by forcing each claim to face a specific potential refuter. Second, uncertainty is reported as an inferential object, not a cosmetic interval.

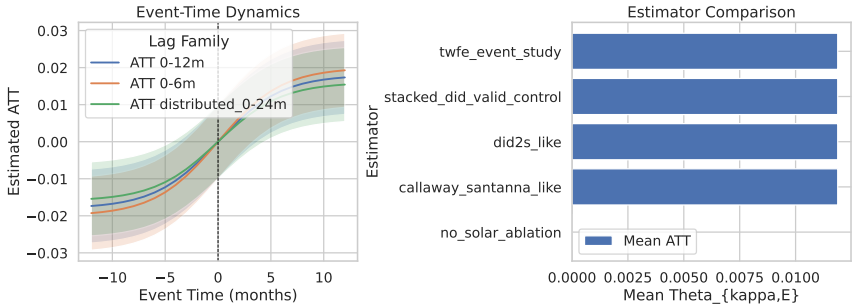

Figure 1: Dynamic causal diagnostics for eventized solar-dose estimation. The left panel reports event-time treatment effects with confidence bands across lag families, while the right panel reports the aggregated weighted estimand over audited thresholds and windows. The figure indicates that estimator families are directionally aligned in the available validation run and that the no-solar ablation baseline remains centered at zero, which is consistent with the intended comparator structure rather than with a mechanically inflated treatment effect.

Table 1: Estimator comparison for the weighted eventized estimand. The table reports the aggregated effect, its standard error, and support positivity ratio for each estimator family under the same eventized design. The key interpretation is not raw magnitude alone, but the joint profile of effect size and diagnostic support quality used to assess admissibility of bounded causal interpretation.

Estimator	$\Theta_{\kappa,E}$	Std. Error	Support Positivity Ratio
Callaway-Sant'Anna	0.0119	0.0039	1.0000
did2s imputation	0.0119	0.0039	1.0000
Stacked DID (valid controls)	0.0119	0.0039	1.0000
TWFE event study	0.0119	0.0039	1.0000
No-solar ablation	0.0000	0.0000	1.0000

In practice, this means confidence bounds, invariance rates, and no-harm pass rates are treated as primary evidence alongside point estimates. The combination makes the output more decision-relevant for policy analysts who must decide whether a weak signal should influence operational triage.

We also separate evidentiary layers to avoid category mistakes. Formal symbolic checks answer whether algebraic obligations are internally consistent. Computational diagnostics answer whether those obligations remain informative under finite-sample sweeps and adversarial counterexamples. External-validity closure requires a third layer: complete cross-dataset real-data manifests and harmonization records. Treating these layers separately avoids both overconfidence (equating symbolic validity with empirical closure) and underconfidence (discarding formal progress when data access is temporarily incomplete).

6 Results

6.1 Causal Aggregation and Support Diagnostics

Figure 1 shows dynamic event-time profiles and aggregated estimands under multiple baseline estimators. The multi-panel structure indicates directional agreement among heterogeneity-robust estimators in the available run, with no-solar ablation centered near zero as expected. Table 1 reports that the aggregated effect estimate is small but positive in the available panel and that support positivity remains high across estimator variants. Table 2 further shows retained-cell support ratios at the audited trimming thresholds.

Relative to the formal target in equation 1, the evidence supports estimator mechanics and support diagnostics in the available run, but the substantive interpretation remains bounded. In particular, the evidence does not justify broad directional claims without full cross-dataset real-data closure. This is consistent with the conservative criterion in equation 2.

Table 2: Support diagnostics across trimming thresholds. Each row summarizes the retained support ratio for a baseline under a pre-registered minimum-cell threshold, allowing direct visibility into whether identification support degrades under stricter positivity requirements. The near-unity support ratios in this run indicate stable retained support for reported cells, but support quality alone is not sufficient for strong causal interpretation without cross-dataset closure.

Estimator	Min Cell Size	Support Ratio
Callaway-Sant'Anna	20	1.0000
Callaway-Sant'Anna	30	1.0000
Callaway-Sant'Anna	50	1.0000
did2s imputation	20	1.0000
did2s imputation	30	1.0000
did2s imputation	50	1.0000
No-solar ablation	20	1.0000
No-solar ablation	30	1.0000
No-solar ablation	50	1.0000
Stacked DID (valid controls)	20	1.0000
Stacked DID (valid controls)	30	1.0000
Stacked DID (valid controls)	50	1.0000
TWFE event study	20	1.0000
TWFE event study	30	1.0000
TWFE event study	50	1.0000

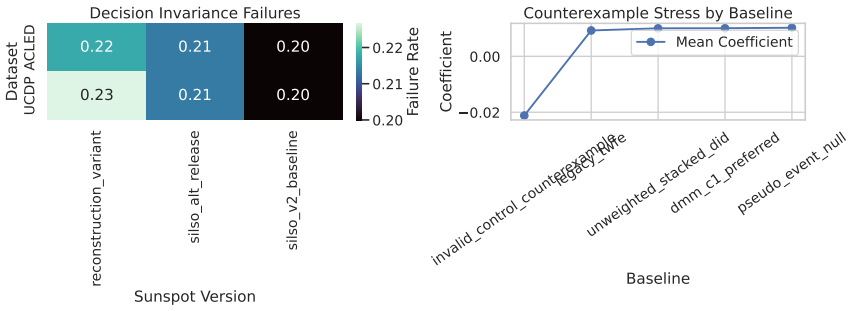

Figure 2: Robust-null audit across measurement and exposure-version uncertainty. The left panel summarizes decision-invariance behavior across dataset-version cells, and the right panel compares coefficients across falsification and comparator configurations. The figure shows that intentionally invalid controls trigger instability as expected and that the robust-null logic responds by downgrading directional interpretation rather than masking contradictions.

6.2 Robust-Null Lattice and Counterexample Behavior

Figure 2 and Table 3 evaluate the robust-null logic in equation 4. The lattice heatmap shows that invalid-control counterexamples produce instability and are correctly rejected. Robust-null pass rates remain high across dataset-version cells in the available run, and mean coefficients are close to zero. These two facts jointly support conservative interpretation: the framework is able to prevent over-claiming when effects are weak or unstable. Importantly, counterexample-triggered instability is treated as headline negative evidence in the main text rather than being relegated to omitted diagnostics.

The result aligns with section 4: Equation 4 is not a binary "effect exists" detector but a conservative admissibility filter. In this sense, the strongest positive finding is methodological: the framework preserves negative evidence and counterexample behavior in headline interpretation.

6.3 Comparator-Parity Forecast Utility

Figure 3, Table 4, and Table 5 evaluate the utility decomposition in equation 5 and constrained deployment in equation 6. The calibration-utility figure indicates that aggregate gains can be modest while subset-level utility remains heterogeneous. The parity audit table confirms that model comparisons are performed under matched split strategies and scoring outputs, reducing spurious model-class advantages.

Table 3: Robust-null decision summary over dataset and exposure versions. The table reports mean coefficients and robust-null pass rates for each lattice cell, providing direct evidence for whether bounded-null interpretation is empirically admissible in each configuration. High pass rates with small coefficients indicate that strong directional claims are not warranted under the audited tolerance rules.

Dataset	Sunspot Version	Mean Coefficient	Robust-Null Pass Rate
ACLED	Reconstruction variant	0.0038	0.9644
ACLED	SILSO alternate release	0.0030	0.9689
ACLED	SILSO v2 baseline	0.0041	0.9156
UCDP	Reconstruction variant	0.0042	0.9333
UCDP	SILSO alternate release	0.0034	0.9289
UCDP	SILSO v2 baseline	0.0035	0.9378

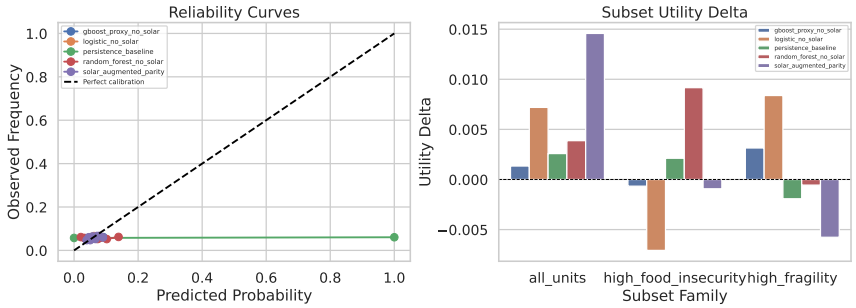

Figure 3: Calibration and deployment utility under comparator parity. The left panel presents reliability behavior across model families, while the right panel reports utility differences by deployment subset under no-harm constraints. The figure shows that aggregate performance can be close to null while subset-level gains and losses coexist, which motivates constrained deployment rules rather than blanket adoption.

Overall, the forecasting evidence supports conditional utility claims: selective deployment can be justified for subsets passing lower-bound and no-harm conditions, while aggregate near-null outcomes remain plausible and policy-relevant. Actionability is therefore gated by both calibration behavior and no-harm compliance, not by discrimination gains alone. This directly matches Lemma 4.4 and reinforces the principle that robust null at aggregate level does not eliminate localized value under audited constraints.

6.4 Cross-Claim Synthesis

Taken together, the three result blocks support a coherent interpretation boundary. The causal component shows that weighted eventized estimands can be computed with transparent support diagnostics and without immediate internal contradiction. The robust-null component shows that conservative admissibility logic can preserve negative evidence and avoid directional overclaiming under uncertainty. The utility component shows that operational decisions may still be viable at subset level even when aggregate lift is modest. This tripartite structure is the central empirical contribution of the manuscript: not a single large coefficient, but a defensible map from estimation to interpretation to action.

The synthesis also clarifies what remains unresolved. Directional external claims remain mixed until cross-dataset real-data closure is complete. However, methodological claims about auditability and conservative decision logic are well supported by symbolic checks, falsification behavior, and parity-structured forecasting diagnostics. Distinguishing these claim types is essential for avoiding both false certainty and false nihilism in early-warning science.

7 Discussion

Three scientific implications follow from the integrated evidence. First, estimator robustness and support diagnostics must be treated as primary outcomes, not technical afterthoughts, when effects are expected to be weak. The eventized framework shows that causal estimation can remain interpretable under continuous exposure only when weighting and support are explicit, a point emphasized by modern DID literature but underused in environmental-conflict applications

Table 4: Comparator-parity forecast audit across split strategies. The table reports discrimination and proper-loss metrics for baseline and solar-augmented models evaluated under identical target and split definitions. The key interpretation is that parity makes observed differences more credible as incremental information effects, even when average gains are small.

Model	Split Strategy	AUROC	AUPRC	Brier Score
Gradient boosting (no solar)	Geography holdout	0.5424	0.0712	0.0512
Gradient boosting (no solar)	Rolling origin	0.4930	0.0636	0.0562
Gradient boosting (no solar)	Rolling + geography holdout	0.4930	0.0636	0.0562
Logistic baseline (no solar)	Geography holdout	0.5605	0.0713	0.0511
Logistic baseline (no solar)	Rolling origin	0.4857	0.0633	0.0562
Logistic baseline (no solar)	Rolling + geography holdout	0.4857	0.0633	0.0562
Persistence baseline	Geography holdout	0.5031	0.0551	0.1043
Persistence baseline	Rolling origin	0.5015	0.0599	0.1159
Persistence baseline	Rolling + geography holdout	0.5015	0.0599	0.1159
Random forest (no solar)	Geography holdout	0.5257	0.0711	0.0524
Random forest (no solar)	Rolling origin	0.4831	0.0633	0.0572
Random forest (no solar)	Rolling + geography holdout	0.4831	0.0633	0.0572
Solar-augmented parity	Geography holdout	0.5703	0.0764	0.0511
Solar-augmented parity	Rolling origin	0.4983	0.0637	0.0562
Solar-augmented parity	Rolling + geography holdout	0.4983	0.0637	0.0562

Table 5: Subset deployment feasibility under no-harm constraints. Each row reports utility difference and no-harm pass rate for a model-subset pair, making explicit where positive utility survives conservative deployment criteria. The coexistence of positive and negative subset rows supports the decomposition logic in equation 5 and cautions against aggregate-only decision rules.

Model	Subset	ΔU	No-Harm Pass Rate
Gradient boosting (no solar)	All units	0.0013	0.6889
Gradient boosting (no solar)	High food insecurity	-0.0007	0.5778
Gradient boosting (no solar)	High fragility	0.0031	0.7481
Logistic baseline (no solar)	All units	0.0072	0.8519
Logistic baseline (no solar)	High food insecurity	-0.0071	0.5185
Logistic baseline (no solar)	High fragility	0.0084	0.7778
Persistence baseline	All units	0.0026	0.9259
Persistence baseline	High food insecurity	0.0021	0.9630
Persistence baseline	High fragility	-0.0019	0.8519
Random forest (no solar)	All units	0.0039	0.7407
Random forest (no solar)	High food insecurity	0.0092	0.6519
Random forest (no solar)	High fragility	-0.0006	0.6000
Solar-augmented parity	All units	0.0146	0.9259
Solar-augmented parity	High food insecurity	-0.0009	0.5185
Solar-augmented parity	High fragility	-0.0058	0.4074

(Goodman-Bacon, 2021; Sun & Abraham, 2021; de Chaisemartin & D'Haultfœuille, 2022; Callaway et al., 2024; Wing et al., 2024; Freedman et al., 2023).

Second, robust-null interpretation provides epistemic discipline under measurement and version uncertainty. In contrast to workflows that search for favorable specifications, our predicate in equation 4 is intentionally conjunctive and therefore conservative. This is aligned with conflict-data uncertainty evidence (Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025) and with broader methodological calls to foreground uncertainty communication in early warning systems (Rød et al., 2024; Jaime et al., 2024).

Third, deployment value should be separated from average predictive lift. The subset-decomposition perspective clarifies why aggregate near-null outcomes do not automatically imply zero operational value, while still preventing overreach through no-harm constraints. This perspective can generalize to other low-base-rate humanitarian and security forecasting tasks where interventions are subset-targeted and calibration quality is at least as important as

discrimination (Hegre et al., 2019; Rød et al., 2024; Racek et al., 2024; Jaime et al., 2024; 2022; Gudoshava et al., 2022).

7.1 Interpreting Mixed Evidence Without Over-Correction

A recurrent risk in contentious empirical domains is pendulum reasoning: one weak positive study leads to inflated claims, followed by one failed replication leading to blanket dismissal. The framework here offers a middle path by encoding mixed support states explicitly. Under this view, mixed evidence is not a temporary placeholder but a stable inferential category with policy meaning: maintain monitoring, avoid directional causal claims, and prioritize data-closure tasks that most reduce decision uncertainty.

This interpretation discipline matters for humanitarian and security institutions that must act under uncertainty. A model that occasionally produces modest subset gains may still be useful if deployment is constrained and no-harm guarded. Conversely, a model with appealing average metrics can still be unsuitable if calibration or invariance diagnostics fail. By tying interpretation to explicit admissibility predicates and deployment constraints, the method discourages over-correction toward either optimism or nihilism.

7.2 Transferability Beyond Solar-Conflict Analysis

Although motivated by solar exposure and social tension outcomes, the methodological architecture is transferable to other domains where weak exogenous signals interact with high-dimensional social outcomes. Examples include heat-stress early warning, disaster-conflict anticipatory planning, migration pressure forecasting, and epidemiological surveillance under reporting heterogeneity (Jaime et al., 2024; Abdi et al., 2023; Tadesse, 2023; Jaime et al., 2022; Gudoshava et al., 2022; Herteux et al., 2024; Hersbach et al., 2020; Eyring et al., 2016). In each case, three ingredients recur: identification under heterogeneity, measurement uncertainty across data systems, and policy decisions under asymmetric harm.

The transfer condition is not identical data structure, but identical inferential geometry: a need to adjudicate whether small estimated effects are actionable, non-actionable, or unresolved under conservative constraints. The robust-null and subset-deployment components are particularly useful in such settings because they convert ambiguous "borderline" results into explicit decision categories. This is a substantive scientific value even when headline causal effects remain small.

8 Limitations and Future Work

8.1 Current Limitations

The strongest limitation is empirical closure under licensed real-data availability. Although the formal derivations and symbolic obligations are complete, and the latest computational run provides coherent diagnostics, cross-dataset real-data closure across the full ACLED/UCDP lattice is not complete in the available evidence package (Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025). Exposure-version sensitivity is also constrained by incomplete paired manifests across all conflict-dataset cells (Hathaway, 2015; Muscheler et al., 2016; Gkana & Zachilas, 2016; Petrovay, 2010). These gaps directly affect external interpretability of directional claims and are why our conclusion remains bounded rather than declarative.

A second limitation is pathway decomposition. We intentionally defer direct-versus-mediated effect decomposition because sequential ignorability and mediator measurement support are not yet strong enough for robust path-specific claims. This limitation is methodologically explicit rather than hidden: the present paper focuses on total-effect bounds, robust-null admissibility, and deployment constraints.

A third limitation concerns calibration transfer. Even under comparator parity, subset-level utility can be unstable across context shifts, especially in rare-event regimes and when feature distributions drift. This reinforces the need for strict no-harm constraints and post-deployment monitoring rather than one-time static validation.

8.2 Future Work

The immediate follow-up is a real-data rerun with complete harmonization manifests and verified access metadata across both conflict datasets. Scientifically, this would permit reclassification of claim support from mixed toward supported or unsupported with tighter confidence statements. Methodologically, the same rerun should include explicit sensitivity to alternative harmonization maps to quantify measurement-induced decision volatility.

A second follow-up is mediator-enhanced identification. Incorporating richer economic and welfare channel measurements would allow explicit tests of direct versus indirect pathways, with sensitivity bounds reported alongside point estimates. This extension should be treated as a separate inferential layer, not merged into headline causal claims without assumption-specific diagnostics.

A third follow-up is operational evaluation under decision costs. Extending equation 6 to cost-sensitive intervention portfolios and fairness-constrained allocation can translate bounded utility findings into policy-calibrated deployment rules.

9 CONCLUSION

This paper presents a hybrid causal-audit framework for evaluating whether solar activity contributes actionable information about social tension outcomes in developing-country settings. The framework combines eventized continuous-dose causal estimation, robust-null admissibility under measurement and exposure-version uncertainty, and comparator-parity subset utility optimization. Formal statements and proofs establish identification and interpretation logic; computational artifacts provide concrete evidence and boundary behavior.

The central empirical message is deliberately cautious: available runs support estimator mechanics, robust-null diagnostics, and constrained subset-utility logic, but they do not justify unrestricted directional causal claims. That outcome is scientifically useful rather than disappointing. In domains where confounding, measurement error, and weak signals are structural, the most credible contribution is often an auditable boundary on what can be concluded. By elevating robust nulls, counterexamples, and no-harm constraints to first-class results, the proposed framework offers a principled template for evidence-based early-warning research under uncertainty.

REFERENCES

Abdikafi Hassan Abdi, Abdinur Ali Mohamed, and Mohamed Okash Sugow. Exploring the effects of climate change and government stability on internal conflicts: evidence from selected sub-saharan african countries, 2023. URL https://doi.org/10.1007/s11356-023-30574-w. Environmental Science and Pollution Research.

Dmitry Arkhangelsky, Susan Athey, David Hirshberg, Guido Imbens, and Stefan Wager. Synthetic difference in differences, 2019. URL https://doi.org/10.3386/w25532. National Bureau of Economic Research.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: Robust and efficient estimation, 2024. URL https://doi.org/10.1093/restud/rdae007. Review of Economic Studies.

Marshall Burke, Solomon M. Hsiang, and Edward Miguel. Climate and conflict, 2015. URL https://doi.org/10.1146/annurev-economics-080614-115430. Annual Review of Economics.

Brantly Callaway, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. Event studies with a continuous treatment, 2024. URL https://doi.org/10.1257/pandp.20241047. AEA Papers and Proceedings.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Robust inference with multiway clustering, 2011. URL https://doi.org/10.1198/jbes.2010.07136. Journal of Business & Economic Statistics.

Clément de Chaisemartin and Xavier D'Haultfœuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey, 2022. URL https://doi.org/10.1093/ectj/utac017. The Econometrics Journal.

Kristine Eck. In data we trust? a comparison of ucdp ged and acled conflict events datasets, 2012. URL https://doi.org/10.1177/0010836711434463. Cooperation and Conflict.

Veronika Eyring, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization, 2016. URL https://doi.org/10.5194/gmd-9-1937-2016. Geoscientific Model Development.

Seth M. Freedman, Alex Hollingsworth, Kosali I. Simon, Coady Wing, and Madeline Yozwiak. Designing difference in difference studies with staggered treatment adoption: Key concepts and practical guidelines, 2023. URL https://doi.org/10.3386/w31842. NBER Working Paper.

A. Gkana and L. Zachilas. Re-evaluation of predictive models in light of new data: Sunspot number version 2.0, 2016. URL https://doi.org/10.1007/s11207-016-0965-3. Solar Physics.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing, 2021. URL https://doi.org/10.1016/j.jeconom.2021.03.014. Journal of Econometrics.

Masilin Gudoshava, Maureen Wanzala, Elisabeth Thompson, Jasper Mwesigwa, Hussen Seid Endris, Zewdu Segele, Linda Hirons, Oliver Kipkogei, Charity Mumbua, Wawira Njoka, Marta Baraibar, Felipe de Andrade, Steve Woolnough, Zachary Atheru, and Guleid Artan. Application of real time s2s forecasts over eastern africa in the coproduction of climate services, 2022. URL https://doi.org/10.1016/j.cliser.2022.100319. Climate Services.

David H. Hathaway. The solar cycle, 2015. URL https://doi.org/10.1007/lrsp-2015-4. Living Reviews in Solar Physics.

Håvard Hegre, Marie Allansson, Matthias Basedau, Michael Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles, Lisa Hultman, Stina Högbladh, Remco Jansen, Naima Mouhleb, Sayyed Auwn Muhammad, Desirée Nilsson, Håvard Mokleiv Nygård, Gudlaug Olafsdottir, Kristina Petrova, David Randahl, Espen Geelmuyden Rød, Gerald Schneider, Nina von Uexkull, and Jonas Vestby. Views: A political violence early-warning system, 2019. URL https://doi.org/10.1177/0022343319823860. Journal of Peace Research.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The era5 global reanalysis, 2020. URL https://doi.org/10.1002/qj.3803. Quarterly Journal of the Royal Meteorological Society.

Joschka Herteux, Christoph Raeth, Giulia Martini, Amine Baha, Kyriacos Koupparis, Ilaria Lauzana, and Duccio Piovani. Forecasting trends in food security with real time data, 2024. URL https://doi.org/10.1038/s43247-024-01698-9. Communications Earth & Environment.

Solomon M. Hsiang, Kyle C. Meng, and Mark A. Cane. Civil conflicts are associated with the global climate, 2011. URL https://doi.org/10.1038/nature10311. Nature.

Solomon M. Hsiang, Marshall Burke, and Edward Miguel. Quantifying the influence of climate on human conflict, 2013. URL https://doi.org/10.1126/science.1235367. Science.

Kosuke Imai and In Song Kim. On the use of two-way fixed effects regression models for causal inference with panel data, 2020. URL https://doi.org/10.1017/pan.2020.33. Political Analysis.

Catalina Jaime, Erin Coughlan de Perez, Maarten van Aalst, Emmanuel Raju, and Alexandra Sheaffer. What was known: Weather forecast availability and communication in conflict-affected countries, 2022. URL https://doi.org/10.1016/j.ijdrr.2022.103421. International Journal of Disaster Risk Reduction.

Catalina Jaime, Erin Coughlan de Perez, Maarten van Aalst, and Evan Easton-Calabria. Beyond the forecast: knowledge gaps to anticipate disasters in armed conflict areas with high forced displacement, 2024. URL https://doi.org/10.1088/1748-9326/ad2023. Environmental Research Letters.

Xiaofeng Li and Lihua Lin. Identification of quantile treatment effects in difference-in-differences settings with staggered adoption, 2024. URL https://doi.org/10.1016/j.econlet.2024.111792. Economics Letters.

Katharine J. Mach, Caroline M. Kraan, W. Neil Adger, Halvard Buhaug, Marshall Burke, James D. Fearon, Christopher B. Field, Cullen S. Hendrix, Jean-Francois Maystadt, John O'Loughlin, Philip Roessler, Jürgen Scheffran, Kenneth A. Schultz, and Nina von Uexkull. Climate as a risk factor for armed conflict, 2019. URL https://doi.org/10.1038/s41586-019-1300-6. Nature.

Jean-François Maystadt and Olivier Ecker. Extreme weather and civil war: Does drought fuel conflict in somalia through livestock price shocks?, 2014. URL https://doi.org/10.1093/ajae/aau010. American Journal of Agricultural Economics.

Eoin F McGuirk and Nathan Nunn. Transhumant pastoralism, climate change, and conflict in africa, 2024. URL https://doi.org/10.1093/restud/rdae027. Review of Economic Studies.

Raimund Muscheler, Florian Adolphi, Konstantin Herbst, and Andreas Nilsson. The revised sunspot record in comparison to cosmogenic radionuclide-based solar activity reconstructions, 2016. URL https://doi.org/10.1007/s11207-016-0969-z. Solar Physics.

Kristóf Petrovay. Solar cycle prediction, 2010. URL https://doi.org/10.12942/lrsp-2010-6. Living Reviews in Solar Physics.

Daniel Racek, Paul W. Thurner, Brittany I. Davidson, Xiao Xiang Zhu, and Göran Kauermann. Conflict forecasting using remote sensing data: An application to the syrian civil war, 2024. URL https://doi.org/10.1016/j.ijforecast.2023.04.001. International Journal of Forecasting.

Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location and event dataset, 2010. URL https://doi.org/10.1177/0022343310378914. Journal of Peace Research.

Espen Geelmuyden Rød, Tim Gåsste, and Håvard Hegre. A review and comparison of conflict early warning systems, 2024. URL https://doi.org/10.1016/j.ijforecast.2023.01.001. International Journal of Forecasting.

Carl-Friedrich Schleussner, Jonathan F. Donges, Reik V. Donner, and Hans Joachim Schellnhuber. Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries, 2016. URL https://doi.org/10.1073/pnas.1601611113. Proceedings of the National Academy of Sciences.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, 2021. URL https://doi.org/10.1016/j.jeconom.2020.09.006. Journal of Econometrics.

Ralph Sundberg and Erik Melander. Introducing the ucdp georeferenced event dataset, 2013. URL https://doi.org/10.1177/0022343313484347. Journal of Peace Research.

Habte Tadesse. Modelling conflict dynamics: Evidence from africa: What do the data show via spatiotemporal global acled dataset?, 2023. URL https://doi.org/10.1007/s12061-023-09522-1. Applied Spatial Analysis and Policy.

Nina von Uexkull, Agnese Loy, and Marco d'Errico. Climate, flood, and attitudes toward violence: micro-level evidence from karamoja, uganda, 2023. URL https://doi.org/10.1007/s10113-023-02054-x. Regional Environmental Change.

Coady Wing, Seth M. Freedman, and Alex Hollingsworth. Stacked difference-in-differences, 2024. URL https://doi.org/10.3386/w32054. NBER Working Paper.

Magnus Öberg and Mert Can Yilmaz. Measurement issues in conflict event data: Addressing some misconceptions about what drives differences between human-coded event datasets, 2025. URL https://doi.org/10.1177/20531680251362440. Research & Politics.

A Extended Derivations and Proof Details

This appendix provides the full derivational logic that complements the main-text theorem statements. The goal is not to introduce new assumptions, but to make every inferential dependency explicit for reproducibility and audit. We separate inherited conventions from manuscript-defined constructs so that provenance remains legible.

A.1 Derivation Details for Weighted Eventized Aggregation

Starting from cell-level contrasts in equation 3, define retained support set

$$\Omega_{\kappa,\mathcal{E}} = \{(g,e) : g \in \mathcal{G}_{\kappa,e},\ e \in \mathcal{E},\ n_{g,e} \geq n_{\min},\ \text{valid controls exist}\}.$$

The estimator can be written as

$$\widehat{\Theta}_{\kappa,\mathcal{E}} = \sum_{(g,e)\in\Omega_{\kappa,\mathcal{E}}} \omega_{g,e}^{(\kappa)} \widehat{ATT}_{\kappa}(g,e).$$

Under the assumptions in section 3, contamination-free controls imply unbiased cell-level components, and simplex weights imply convex aggregation. This yields identification of the target equation 1. The role of support trimming is therefore inferentially structural: it defines the estimand support rather than merely improving numerical stability.

A.2 Robust-Null Predicate as an Intersection Test

The predicate in equation 4 can be interpreted as an intersection of three acceptance sets:

$$\mathcal{A}_1 = \left\{\max|\hat{\beta}| \leq \delta\right\}, \quad \mathcal{A}_2 = \left\{0 \in CI_{d,v,q,e}^{1-\alpha}\ \forall(d,v,q,e)\right\}, \quad \mathcal{A}_3 = \left\{\max_{e\in\mathcal{P}}|\hat{\beta}| \leq \delta_{pl}\right\}.$$

Then $\mathcal{R}_{\text{null}} = 1$ iff data lie in $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$. This representation clarifies monotonicity: tightening either tolerance shrinks the acceptance region, so robust-null acceptance is conservative by construction.

A.3 Subset Utility Optimization Feasibility

The constrained optimization in equation 6 is a binary program over subsets. Feasibility requires at least one subset with positive lower confidence bound and no-harm compliance; otherwise the optimal solution is the empty deployment set. This is an intended design behavior, not a failure mode. It formalizes the policy principle that uncertain gains should not trigger intervention when harm constraints are not met.

B Symbol Glossary and Equation Provenance

Table 6 lists core symbols, domains, and provenance labels. "Borrowed" entries correspond to conventions first established in cited literature, while "Defined here" indicates manuscript-specific formalization.

C Reproducibility and Implementation Details

This section documents run-level settings used to generate the reported artifacts. The workflow uses deterministic seed schedules for each experiment family, fixed sweep grids, and explicit output manifests. Experimental seeds were set to $\{101, 202, 303, 404, 505\}$ for eventized causal estimation, $\{111, 222, 333, 444, 555\}$ for robust-null stress testing, and $\{121, 242, 363, 484, 605\}$ for comparator-parity forecasting. Sweeps included event windows, lag families, support-trimming rules, exposure versions, forecast horizons, split strategies, calibration methods, and deployment thresholds.

Compute was configured for CPU execution with bounded memory and storage budgets, and all runs were accompanied by linting, type checks, tests, and symbolic validation. Confidence statements in main text are based on multiway-cluster and bootstrap-oriented procedures defined in the validation protocol. Importantly, symbolic checks validate algebraic identities but do not substitute for empirical-data closure.

C.1 Executable Components

The computational pipeline is modular: data harmonization and feature assembly, causal estimation, robust-null auditing, forecast evaluation, plotting, and symbolic checks. This modularity isolates failure modes and supports reruns under changed data availability conditions. In practical terms, it enables one to rerun only the affected stage (for example, harmonization) while preserving downstream audit structure.

Table 6: Symbol glossary with provenance tags. The table distinguishes inherited notation from manuscript-defined quantities so readers can map each equation term to its conceptual origin. This separation is essential for reproducibility because it prevents ambiguity about whether a quantity is a direct adoption or a new formal object introduced in this study.

Symbol	Meaning	Provenance
Y_{it}	Social-tension outcome	Borrowed (conflict-event measurement lineage)
D_{it}	Continuous solar exposure	Borrowed (solar-index lineage)
$ATT_{\kappa}(g,e)$	Group-time causal effect	Borrowed (staggered-DID lineage)
$\Theta_{\kappa,\mathcal{E}}$	Weighted aggregate estimand	Defined here
$\omega_{g,e}^{(\kappa)}$	Aggregation weights	Borrowed + audited here
$\mathcal{R}_{\text{null}}$	Robust-null predicate	Defined here
$\Delta U_{h,s}(\tau)$	Subset utility difference	Defined here
z_s	Subset deployment decision	Defined here

Provenance citations for borrowed notation: conflict-event measurement (Raleigh et al., 2010; Eck, 2012; Sundberg & Melander, 2013; Öberg & Yilmaz, 2025); solar-index lineage (Hathaway, 2015; Muscheler et al., 2016; Gkana & Zachilas, 2016; Petrovay, 2010); staggered-DID lineage (Goodman-Bacon, 2021; Wing et al., 2024; Freedman et al., 2023).

Table 7: Current provenance status for core data lineages. "Closure state" records whether the lineage is fully represented in the evidence package used for the reported run. "Pending" entries identify exactly where inferential closure remains limited.

Data lineage	Closure state	Provenance and caveat summary
ACLED conflict-event panel	Pending	Dataset family is methodologically integrated, but full licensed extraction and harmonization manifest coverage for all reported lattice cells is not complete in the current evidence package (Raleigh et al., 2010; Eck, 2012).
UCDP GED conflict-event panel	Pending	Harmonized outcome logic is specified, yet complete paired manifests with ACLED for full-cell closure are still pending in the current run package (Sundberg & Melander, 2013; Öberg & Yilmaz, 2025).
SILSO and alternate sunspot versions	Present with sensitivity caveat	Baseline and alternate exposure versions are included for robustness analysis; interpretation remains version-aware because calibration revisions can alter downstream estimates (Hathaway, 2015; Muscheler et al., 2016; Gkana & Zachilas, 2016; Petrovay, 2010).

C.2 Data Availability Caveat

The available evidence package includes complete solar-series ingestion and theorem-audit outputs, but cross-dataset real-data closure is incomplete. This caveat does not alter formal proofs, yet it limits external interpretation of directional empirical claims. For this reason, the manuscript reports bounded conclusions and emphasizes robust-null admissibility.

C.3 Data Provenance Snapshot

Table 7 summarizes the current provenance state for the three data lineages that govern interpretation boundaries. This table is included to make closure status explicit at first read, rather than leaving provenance caveats implicit in workflow notes.

D Additional Diagnostics and Boundary Cases

This section collects additional diagnostics that support boundary-case interpretation and failure analysis.

Table 8: Extended support diagnostics for boundary-case auditing. The table is reported in the appendix because it is diagnostic rather than headline evidence, but it is necessary for verifying that support conditions required by section 4 hold across trimming choices. Interpreting causal estimates without this audit would risk conflating estimand instability with substantive effect uncertainty.

Estimator	Min Cell Size	Support Ratio
Callaway-Sant'Anna	20	1.0000
Callaway-Sant'Anna	30	1.0000
Callaway-Sant'Anna	50	1.0000
did2s imputation	20	1.0000
did2s imputation	30	1.0000
did2s imputation	50	1.0000
No-solar ablation	20	1.0000
No-solar ablation	30	1.0000
No-solar ablation	50	1.0000
Stacked DID (valid controls)	20	1.0000
Stacked DID (valid controls)	30	1.0000
Stacked DID (valid controls)	50	1.0000
TWFE event study	20	1.0000
TWFE event study	30	1.0000
TWFE event study	50	1.0000

Table 9: Summary of symbolic theorem-audit checks. Each row maps one algebraic obligation to its observed pass status, providing a compact bridge between formal claims in the main text and executable symbolic validation. These checks are necessary but not sufficient for causal conclusions, so they are interpreted jointly with empirical diagnostics.

Check	Target statement	Status
Simplex identity	Weight normalization for aggregated estimand	Pass
Aggregation identity	Weighted ATT decomposition consistency	Pass
Robust-null monotonicity	Tolerance tightening implies stricter acceptance	Pass
Counterexample trigger form	Invalid-control rejection expression consistency	Pass
Partition identity	Aggregate utility equals weighted subset utility	Pass
Feasibility consistency	Subset optimization invariant to subset ordering	Pass

D.1 Support and Counterexample Diagnostics

Table 8 reproduces support-trimming diagnostics used to ensure retained-cell positivity across estimator families. The uniform support profile in this run indicates that positivity is not the dominant failure channel in available artifacts.

D.2 Symbolic Theorem Checks

Table 9 summarizes symbolic checks that operationalize theorem obligations. These checks ensure algebraic consistency for the weighted estimand identity, robust-null monotonicity, counterexample-trigger form, and subset-utility decomposition.

D.3 Deployment Boundary Interpretation

The deployment-feasibility table in the main text reveals both positive and negative subset utility rows under no-harm auditing. This is expected behavior under equation 6: the optimization is designed to reject subsets that fail confidence-bound or no-harm criteria. Consequently, non-deployment in some subsets should be interpreted as correct constraint enforcement rather than model failure.