NBER WORKING PAPER SERIES

BILATERAL CONFLICT RISK AND TRADE: MILITARY WARS, TRADE WARS, AND DIPLOMATIC NOISE

Joshua Aizenman
Rodolphe Desbordes
Jamel Saadaoui

Working Paper 35077
http://www.nber.org/papers/w35077

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026, Revised May 2026

Joshua Aizenman is grateful for the support provided by the Dockson Chair in Economics and International. Relations, USC. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Bilateral Conflict Risk and Trade: Military Wars, Trade Wars, and Diplomatic Noise
Joshua Aizenman, Rodolphe Desbordes, and Jamel Saadaoui
NBER Working Paper No. 35077
April 2026, Revised May 2026
JEL No. F15, F43, F5, F50, F51

ABSTRACT

How damaging is a "trade war" compared to a "military war" or a "war of words"? Aggregate conflict indicators cannot say, because they treat missile strikes, sanctions, and diplomatic protests as equivalent. We build a monthly bilateral indicator from GDELT event data, calibrated against human-curated ground truth, that decomposes hostility into four layers: kinetic fighting ("military war"), military posture, sanctions-context tensions ("trade war"), and routine diplomacy. The decomposed panel reveals a secular shift: over the past decade, governments have steadily substituted economic coercion for military confrontation, nearly doubling the trade-weighted share of hostility channelled through sanctions contexts. In a gravity trade model, the aggregate indicator is negative, large, and statistically significant, but the decomposition reveals that only two layers drive the result. Kinetic conflict and trade-context hostility are both economically large and precisely estimated; routine diplomacy, despite dominating measured hostility, has no trade effect at all. The directed structure uncovers a retaliation channel that compounds trade losses over several months. Our measure remains a significant determinant of bilateral trade in a horse race against the closest existing alternative. Relative to a pre-escalation baseline, the geopolitical deterioration of the past decade has put roughly $334 billion of bilateral trade at risk, with the US–China pair accounting for half.

Joshua Aizenman
University of Southern California
School of International Relations
Department of Economics
and NBER
aizenman@usc.edu

Rodolphe Desbordes
SKEMA Business School
Department of Economics
rodolphe.desbordes@skema.edu

Jamel Saadaoui
University of Paris 8
jamelsaadaoui@gmail.com

1 Introduction

The US–China confrontation is routinely called a "trade war." But what does that label actually mean for trade? How does it compare with the destruction caused by a real war in Ukraine or Gaza? When Western governments impose sanctions on Russia, is it the sanctions that disrupt trade, or the broader hostile environment that accompanies them? And does it matter that China retaliates against US measures months later? These questions are central to trade policy, supply chain strategy, and the growing literature on geopolitical fragmentation, yet existing tools cannot answer them. No bilateral measure of geopolitical risk distinguishes the *type* of hostility between two countries or tracks how it evolves over time.

The need for such a measure is sharpened by the growing prominence of economic coercion. Governments are reaching for diplomatic pressures, tariffs, export controls, and financial sanctions rather than troops. But these different forms of conflict need not affect trade in the same way. Military confrontation may destroy infrastructure and sever supply chains. Sanctions-context hostility may generate policy uncertainty and trigger retaliatory escalation. Routine diplomatic friction, which fills the headlines, may leave trade untouched. If so, an aggregate indicator that treats these channels as equivalent will understate the cost of actual war, overstate the relevance of diplomatic noise, and miss the retaliation dynamics that build over time.

We construct a monthly bilateral conflict indicator from GDELT v2 event data, calibrated against the human-curated ICEWS dataset using Ridge regression. Although the indicator measures the hostile *fraction* of bilateral interaction, not its volume, calibration remains essential: GDELT systematically misclassifies routine interactions as hostile, inflating the hostile share for allied, high-trade pairs that dominate media coverage. Without correction, these measurement errors would bias the estimated conflict-trade relationship. Each observation decomposes into four additive components: kinetic fighting, military posture, trade-context hostility, and baseline diplomacy. The indicator is directed, allowing separate estimation of aggressive actions and retaliatory responses.

Three sets of findings emerge. First, the decomposed panel tracks the key geopolitical developments of the past decade. The Russia–Ukraine relationship militarises long before 2022. The US–China confrontation is overwhelmingly economic. These patterns are invisible to an aggregate indicator (Section 3).

Second, in a gravity equation, the aggregate indicator is negative, large, and statistically significant, but the decomposition reveals that only two of the four layers drive the result. Kinetic conflict ("military war") and trade-context hostility ("trade war") are both economically large and precisely estimated. Routine diplomacy, despite dominating measured hostility, has a trade effect indistinguishable from zero. The directed structure uncovers a retaliation channel that compounds trade losses over several months. Against pre-escalation trade, the geopolitical deterioration since 2015 has put roughly $334 billion of bilateral trade at risk, with the US–China pair accounting for half (Section 4).

Third, in a horse race against IntenSE (Chevalier et al., 2026), the closest existing bilateral indicator, our decomposition remains both statistically significant and economically large. Because the two indicators draw on the same underlying GDELT data, the comparison isolates the contribution of our four design choices: Goldstein weighting, supervised calibration against human-curated ground truth, a directed (asymmetric) bilateral structure,

and sanctions-context reclassification (Section 5).

The paper contributes to three literatures. In the conflict-trade and sanctions literatures, Glick and Taylor (2010) and Martin et al. (2008) establish that wars reduce trade but rely on binary measures; Michaels and Zhi (2010), Fuchs and Klann (2013), and Heilmann (2016) show that political friction affects trade using specific episodes; Hufbauer et al. (2007) catalogue sanctions, Felbermayr et al. (2020) provide a bilateral database, and Crozet and Hinz (2020) estimate Russian counter-sanctions effects. All treat conflict or sanctions as discrete events. We show that not all conflict types matter for trade, that the intensive margin of hostility within sanctions episodes dominates the extensive margin, and that the hostile *environment* surrounding sanctions, not their binary status, predicts trade disruption.

In the text-as-data literature, Baker et al. (2016) construct economic policy uncertainty from newspaper text; Caldara and Iacoviello (2022) build a geopolitical risk index; Mueller and Rauh (2018) use news text for conflict prediction. Chevalier et al. (2026) construct a bilateral tension indicator from GDELT event counts. Our indicator differs in four design choices whose consequences we examine in a gravity horse race.

Section 2 describes indicator construction. Section 3 documents the geopolitical landscape. Section 4 presents the gravity estimation. Section 5 tests robustness and runs the horse race. Section 6 concludes.

2 Indicator Construction

2.1 Data

The indicator draws on four sources. GDELT v2 (2015–2026) provides event data coded in the CAMEO taxonomy with Goldstein conflict-cooperation scores (-10 to $+10$), extracted from news in over 100 languages (Leetaru and Schrodt, 2013). We retain 209 sovereign country codes, excluding regional actors and non-sovereign territories. No event-type filtering is applied; the calibration learns which events are informative. The Integrated Crisis Early Warning System (ICEWS; Boschee et al., 2015), developed under a DARPA-funded programme at Lockheed Martin, provides human-curated event data in the same CAMEO taxonomy. Unlike GDELT's fully automated pipeline, ICEWS applies supervised learning with human-in-the-loop validation, yielding substantially lower false-positive rates for high-intensity events. We use the 2015–2019 window as ground truth for calibration. A data quality break in January 2020, when military events drop to zero globally in the public release, restricts use of ICEWS to this training window. The Global Sanctions Data Base (GSDB) v4 (Felbermayr et al., 2020) provides a bilateral sanctions database with sender, target, type (trade, financial, arms, travel), and start/end years. The Global Trade Alert (GTA) database (Evenett and Fritz, 2022) catalogues discriminatory trade policy interventions implemented since 2008; we retain only "Red" (harmful) bilateral border measures (tariffs, quotas, export controls), excluding horizontal or multilateral measures. Both databases enter only as context labels for the sanctions flag $D^{\text{sanc}}_{ij,t}$ (Section 2.4), not as conflict measures. CEPII (Head et al., 2010) and BACI (Gaulier and Zignago, 2010) provide bilateral trade data and GDP used as time-invariant weights.

2.2 The Conflict Risk Score

The score measures the hostile *fraction* of bilateral interaction. Construction proceeds in three steps: daily aggregation, monthly averaging, and exponential smoothing.

Daily risk. For directed pair $(i \to j)$ on day d, we compute:

$$r_{ij,d} = \frac{\sum_{e \in C_{ij,d}} |G_e| \cdot M_e}{\sum_{e \in C_{ij,d}} |G_e| \cdot M_e + \sum_{e \in K_{ij,d}} G_e \cdot M_e} \tag{1}$$

where C and K denote conflictual (QuadClass 3–4, $G_e < 0$) and cooperative (QuadClass 1–2, $G_e > 0$) events, G_e is the Goldstein score, and M_e is NumMentions (the number of source documents reporting the event). NumMentions serves as a media salience weight: events covered by many outlets contribute more than events reported by a single source, on the grounds that widely reported events are more likely to reflect real diplomatic or military actions rather than coding noise. Using absolute Goldstein values ensures that hostile and cooperative events of equal intensity contribute symmetrically (because cooperative events have $G_e > 0$, $|G_e| = G_e$ for those events, so the absolute value is needed only for the Goldstein conflict values). We distinguish four conflict domains: kinetic (armed violence), posture (military threats and force displays), trade (economic sanctions and coercion), and political/diplomatic (see Section 2.4 and Table A10 in the Appendix). Domain-specific versions of (1) restrict the conflict numerator to events coded in that domain, while the cooperative denominator is shared across all domains, ensuring that domain risks sum to total risk by construction. The ratio structure ensures cross-sectional comparability: pairs generating thousands of events (USA–China) and pairs generating a handful (Armenia–Azerbaijan) produce scores in $[0, 1]$ reflecting the proportion of hostility, not the volume of media coverage.

Monthly flow. The monthly score averages daily risk over all calendar days in the month:

$$r_{ij,t} = \frac{1}{|D_t|} \sum_{d \in D_t} r_{ij,d} \tag{2}$$

where $|D_t|$ is the number of days in month t (28–31). Days with no events for a pair contribute $r_{ij,d} = 0$, so months with sparse coverage produce lower flow values. This day-averaging prevents days with many events from dominating the monthly score.

Exponential stock. The flow captures the current month's hostility; the stock captures persistence. We apply exponential smoothing:

$$\bar{r}_{ij,t} = \lambda \, \bar{r}_{ij,t-1} + (1 - \lambda) \, r_{ij,t}, \qquad \lambda = 0.95 \tag{3}$$

with $\lambda = 0.95$ chosen so that the half-life is about 14 months, allowing for build-up and persistence of geopolitical tensions even if media attention diminishes.[1] The recursion is initialised using GDELT v1 data (2009–February 2015) to avoid cold-start bias. Both flow and stock enter the Ridge calibration as features (see Appendix H).

[1] The stock's half-life of 14 months bridges intermittent spikes into a continuous measure of accumulated hostility, which is better suited for tracking build-ups and for descriptive analysis than the raw flow. In practice, the choice is inconsequential for the main results: the stock variable is not used in the preferred gravity specification (Table 4), where the decomposed monthly flows enter directly. The stock serves as a robustness check (Table 6, column 4), as a moderator (column 7), and as a descriptive tool for tracking sanctions trajectories (Section 3.6).

2.3 Supervised Calibration

Without calibration, joint military exercises receive negative Goldstein scores and routine diplomatic exchanges generate excessive hostile event counts. Allied pairs such as FRA→DEU score 0.246, implying that one-quarter of Franco-German interaction is hostile.

Raw GDELT is noisy. Automated coding inflates hostile event counts for pairs that generate heavy media traffic and misclassifies routine interactions as hostile. Calibration extracts signal from this noise: it maps the noisy GDELT features onto human-curated ground truth and attenuates the measurement bias that would otherwise contaminate the conflict-trade relationship. We calibrate GDELT against ICEWS using five Ridge regression models trained on the 2015–2019 overlap: one model predicts the total risk level, four predict domain shares (kinetic, posture, political/diplomatic, trade). All five models use the same pipeline: features are standardised to zero mean and unit variance, missing values are filled with zero, and the regularisation parameter α is selected by five-fold grouped cross-validation grouped by dyad. Grouping by dyad prevents leakage: all months of a pair remain in the same fold, forcing the model to generalise across pairs rather than memorise pair-specific patterns. The 84 input features span six groups: risk flows and stocks, domain shares, log transforms, trend deviations, composition ratios, and event-level statistics from raw GDELT records. Full details and coefficient tables are in Appendix H.

The total model achieves out-of-sample $R^2 = 0.48$. The calibrated indicator is:

$$\hat{r}^{\text{cal}}_{ij,t} = \text{clip}\big(\hat{f}(X_{ij,t}),\ 0,\ 1\big) \tag{4}$$

where $\hat{f}$ is the Ridge prediction and $X_{ij,t}$ is the feature vector. Domain-level calibrated values are constructed as $\hat{r}^{\text{cal}}_{k,ij,t} = \hat{s}^{k}_{ij,t} \times \hat{r}^{\text{cal}}_{ij,t}$, where $\hat{s}^{k}$ are the four predicted domain shares normalised to sum to one. The last domain is computed as a residual to enforce exact additivity.

Calibration compresses allied pairs asymmetrically. FRA→DEU drops from 0.246 to 0.055 (-78%); USA→GBR from 0.336 to 0.132 (-61%). Adversaries compress less (ISR→PSE: -27%). Across five canonical adversary and five allied pairs, the adversary/ally mean separation ratio improves from 1.8× (raw GDELT) to 4.2× (calibrated).[2] Without calibration, levels are frequently inflated and the layers seem inconsistent with real events (Appendix E, Figure A9).

2.4 Four-Component Decomposition

The decomposition operates on the *calibrated* risk components produced by the Ridge pipeline (Section 2.3). Each GDELT event maps to one of four domains via 156 CAMEO codes: *kinetic* (CAMEO 18–20), *posture* (CAMEO 15 subset), *political/diplomatic*, and *trade*.[3] The Ridge calibrates these domain-specific scores against ICEWS, yielding $\hat{r}^{\text{cal}}_{\text{kin}}$,

[2] Adversary pairs: USA→CHN, USA→RUS, USA→IRN, ISR→PSE, RUS→UKR. Allied pairs: USA→GBR, USA→CAN, FRA→DEU, USA→JPN, DEU→NLD. Means are computed over event months (component > 0) across the full 2015–2025 panel. Raw GDELT: adversary 0.508, ally 0.286. Calibrated: adversary 0.342, ally 0.081. Allies compress by 72%, adversaries by 33%, because calibration disproportionately removes overcoded routine interactions that inflate allied pairs.

[3] The mapping is applied at the CAMEO four-digit sub-code level; Table A10 summarises the dominant assignment for each two-digit root category.

$\hat{r}^{\text{cal}}_{\text{pos}}$, $\hat{r}^{\text{cal}}_{\text{dip}}$, and $\hat{r}^{\text{cal}}_{\text{trd}}$, where each component equals the predicted domain share times total calibrated risk (Section 2.3). The decomposition then applies a binary context flag, $D^{\text{sanc}}_{ij,t} = 1$ if the pair has active trade or financial sanctions, or more than ten GTA border interventions, to produce four conflict *layers* (denoted L for layer):

$$L^{\text{kin}}_{ij,t} = \hat{r}^{\text{cal}}_{\text{kin},ij,t} \tag{5}$$

$$L^{\text{pos}}_{ij,t} = \hat{r}^{\text{cal}}_{\text{pos},ij,t} \tag{6}$$

$$L^{\text{trd}}_{ij,t} = \hat{r}^{\text{cal}}_{\text{trd},ij,t} + D^{\text{sanc}}_{ij,t} \times \hat{r}^{\text{cal}}_{\text{dip},ij,t} \tag{7}$$

$$L^{\text{base}}_{ij,t} = (1 - D^{\text{sanc}}_{ij,t}) \times \hat{r}^{\text{cal}}_{\text{dip},ij,t} \tag{8}$$

The identity $L^{\text{kin}} + L^{\text{pos}} + L^{\text{trd}} + L^{\text{base}} \equiv \hat{r}^{\text{cal}}_{ij,t}$ holds exactly. Pairwise correlations are moderate (kinetic–trade: $r = 0.19$; kinetic–posture: $r = 0.47$) and variance inflation factors are below 1.5.

The economic logic is straightforward. When sanctions are active, diplomatic hostility becomes part of the economic confrontation: demands, accusations, and threats are instruments of coercion directed at trade relationships. When sanctions are absent, the same diplomatic events enter L^{base}, which captures the residual political tension between partners. The GTA threshold of $K = 10$ interventions separates routine anti-dumping measures from genuine coercive policy.[4]

Of the four components, L^{kin} and L^{trd} are the most reliably measured. Kinetic events (armed violence) are unambiguous and well-predicted by the Ridge calibration ($R^2 = 0.23$; see Appendix H). The trade component is reinforced by external sanctions data, which resolves the trade/diplomatic boundary that the Ridge alone cannot draw. In contrast, posture events (threats, force displays) are conceptually ambiguous and poorly predicted ($R^2 = 0.08$): a naval deployment near a disputed strait could be coded as posture, diplomacy, or even kinetic depending on context. Misclassification between L^{pos} and L^{base} simply shifts mass between two components that both capture non-trade, non-kinetic hostility, without affecting the kinetic or trade channels. The gravity results should therefore be interpreted with the most confidence for kinetic and trade effects.

Table 1 illustrates the decomposition for five pairs spanning the full spectrum of conflict types. ISR→PSE is dominated by kinetic risk (0.300), with negligible trade-context hostility. USA→CHN is the mirror image: trade-context hostility (0.154) dwarfs the kinetic component (0.003). RUS→UKR combines both channels. USA→IRN is almost entirely sanctions-driven. FRA→DEU, the control case, shows only baseline diplomatic friction.

2.5 Panel Construction

The panel spans 29,236 directed pairs over 133 months (March 2015 to March 2026), a notional $29{,}236 \times 133 = 3{,}888{,}388$ pair-months. Of these, only 912,136 (23.5%) are event-

[4] Of the 3,880 pairs classified as sanctioned at $K = 10$ over 2015–2023, 66% would remain classified at $K = 20$. The $K = 10$ threshold is used to separate genuine economic coercion from minor trade disputes (e.g., anti-dumping disputes). GSDB and GTA are complementary: across the 153,628 sanctioned dyad-months in the 2015–2023 panel, 80% are flagged by GTA's bilateral Red interventions ($n^{\text{border}} > 10$) alone, 10% by GSDB (trade or financial) alone, and 10% by both. GTA captures pairs with observable bilateral tariff and anti-dumping activity, whereas GSDB captures regime-level primary sanctions that GTA's bilateral filter excludes because they are coded as horizontal multi-country measures. USA→IRN is the canonical illustration: $n^{\text{border}} = 0$ in every sample month, yet GSDB flags both trade and financial sanctions for all 106 sample months.

Table 1: Decomposition: Case Studies (2015–2023 means)

Pair	L^{kin}	L^{pos}	L^{trd}	L^{base}	Pattern
ISR→PSE	0.300	0.007	0.004	0.095	Kinetic war
USA→CHN	0.003	0.003	0.154	0.000	Trade war
RUS→UKR	0.254	0.035	0.158	0.000	Kinetic + sanctions
USA→IRN	0.023	0.001	0.204	0.000	Comprehensive sanctions
FRA→DEU	0.008	0.002	0.001	0.044	Baseline diplomacy

Notes: Mean values (2015–2023) of the four decomposed hostility components for five illustrative pairs. L^{kin}: kinetic (armed violence, CAMEO 18–20). L^{pos}: military posture (threats and force displays, CAMEO 15 subset). L^{trd}: trade-context hostility (sanctions and coercive economic measures, including diplomatic events reclassified via the sanctions flag). L^{base}: residual baseline political/diplomatic activity. By construction, the four components sum to the aggregate calibrated hostility score $\hat{r}^{\text{cal}}_{ij,t}$. "Pattern" is a qualitative label summarising the dominant channel for each pair. Values are fractions in [0,1]: a value of 0.30 means 30% of bilateral interactions fall in that component. Multiply by 100 to compare with Table A7, which reports percentage points.

months in which GDELT recorded at least one CAMEO event between i and j. Monthly flow variables enter the trade panel database directly and the absence of GDELT-based hostility is treated as $r_{ij,t} = 0$.[5] The stock variable serves as a robustness check (Table 6, column 4) and as a moderator in the structural hostility interactions (Table 6, column 7), where the pair's persistent hostility level is the relevant concept.

3 The Geopolitical Landscape

Before turning to trade effects, we examine what the decomposed panel reveals about the geopolitical landscape itself.

3.1 The Economisation of Conflict

Figure 1 presents two views of global conflict risk over 2015–2023. The top row indexes each component to 2015 = 100. Beneath a stable aggregate, the components move in opposite directions. Trade hostility rises steadily; baseline friction declines. Trade-weighted, the divergence is dramatic: trade hostility more than doubles (index 100 to 250) while baseline declines modestly (100 to 83). Kinetic and posture components are roughly stable unweighted but drift upward under trade-weighting (both rising by about a quarter), consistent with larger trading pairs becoming somewhat more militarised over the period.

[5]Descriptive figures follow one of two conventions depending on the question they address. Figures computing *unconditional* global levels or long-horizon changes (Figure 1 global trends, Figure 8 decoupling) use the trade-merged panel and treat trade-present pair-months with no recorded events as zero, because the question is 'what is the average over all pair-months?'. Figures computing *structural* comparisons, such as adversary/ally discrimination (Figures A5 and A7), between-bloc versus within-bloc hostility ratios (Figure A2), and intra-EU versus external hostility (Figure A3), use the event-only panel, so means are conditional on at least one recorded CAMEO event; captions flag this with 'event months only'. The structural question is different: 'when events happen, how hostile are they on average?' For pairs that are always active (major powers, canonical adversaries, conflict pairs), the two conventions coincide because every month has at least one event. They diverge for allied and peripheral pairs, where the conditional mean is mechanically higher than the unconditional mean. Ratio-of-sums statistics, such as the weaponisation index (Figure 5), are invariant to the choice because zero-event months contribute zero to both numerator and denominator.

The bottom row shows the same shift in compositional terms. The military share (kinetic + posture) holds steady. The economic share rises from one-fifth to nearly one-third of all bilateral hostility (unweighted), and from one-quarter to nearly one-half when weighted by trade. Baseline diplomacy declines correspondingly.

Bilateral conflicts are not becoming more military; they are increasingly occurring within a trade war context.

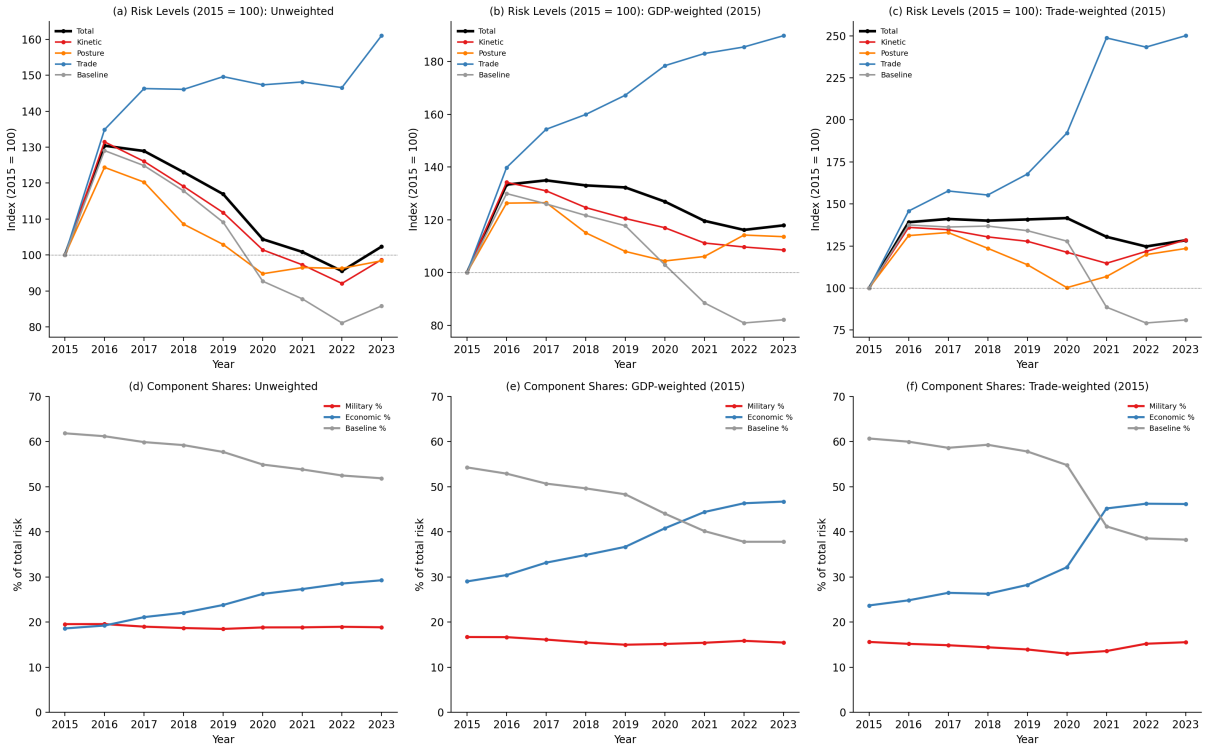

Figure 1: Global conflict risk, 2015–2023. Top: component levels indexed to 2015 = 100 (unconditional means across all pair-months). Bottom: component shares (ratio of weighted means). Left: unweighted. Centre: GDP-weighted. Right: trade-weighted. Weights are time-invariant (2015 values).

3.2 Case Studies

Figure 2 shows the decomposition for eight bilateral pairs (y-axis scales differ). ISR→PSE is kinetic-dominated: the October 2023 Gaza escalation produces a large L^{kin} spike with no sanctions context. USA→CHN is trade-dominated: the March 2018 tariff announcement appears as an L^{trd} spike. RUS→UKR combines kinetic fighting with sanctions, both surging in February 2022. FRA→DEU, the control case, shows stable low-level baseline friction.

Pairs with similar aggregate scores can have entirely different compositions. ISR→PSE is overwhelmingly kinetic. USA→CHN is overwhelmingly trade-context, with sustained hostility that compounds over time. FRA→DEU is pure baseline diplomacy.

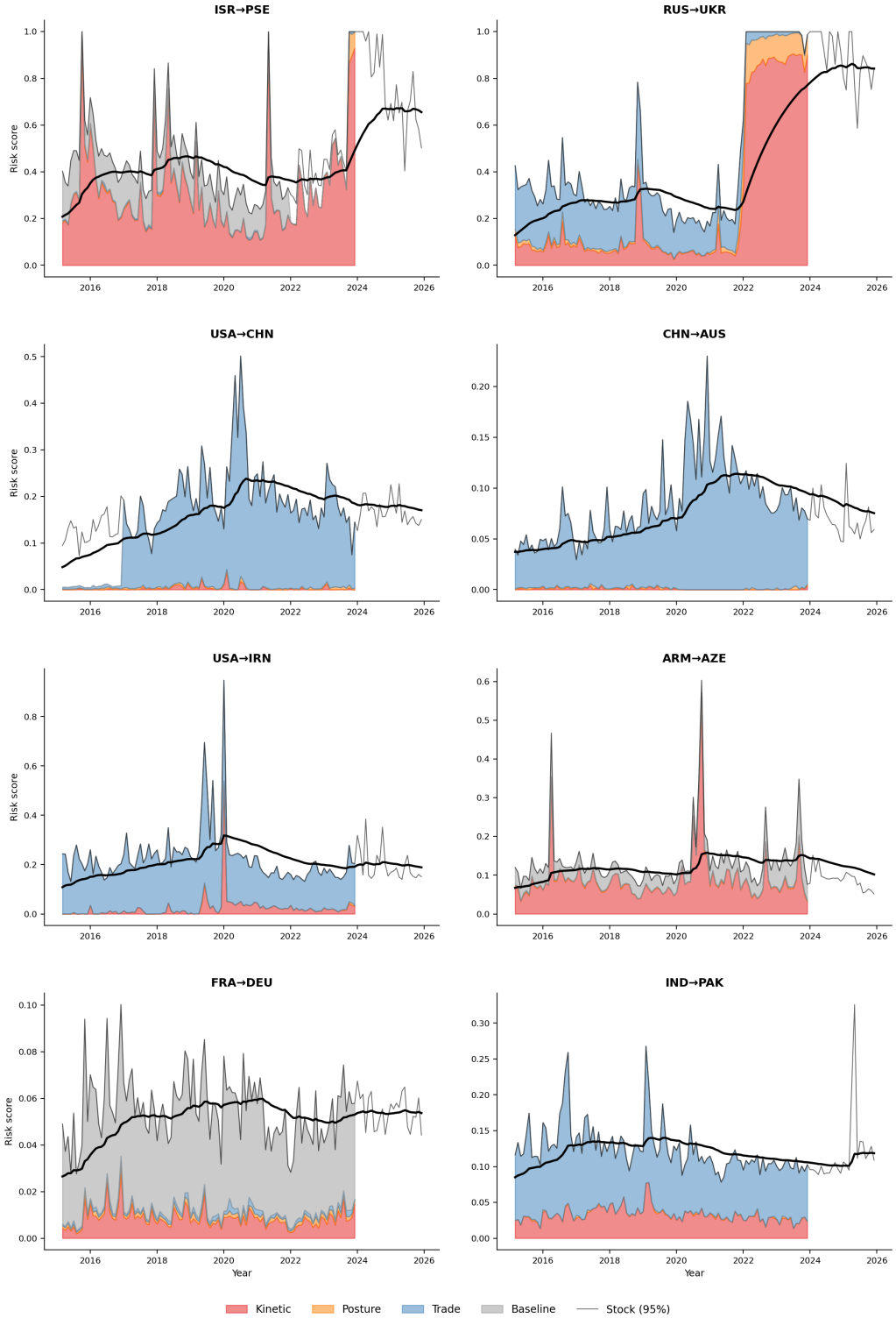

Figure 2: Four-component decomposition: eight case studies. Stacked area (through 2023): kinetic (red), posture (orange), trade (blue), baseline (grey). Thin line: monthly flow. Thick line: stock ($\lambda = 0.95$).

3.3 Great Power Rivalries

The decomposition distinguishes rivalries that aggregate indicators conflate. The US–China relationship is overwhelmingly trade-context hostility (L^{trd} accounts for over 95% of the mean bilateral risk, Table 1): it is an economic confrontation, not a military one. The US–Russia relationship is *militarising*: the kinetic share rises steadily after the Ukraine invasion, in sharp contrast. China and Russia direct hostility at geographically distinct targets, so their outgoing hostility profiles are only weakly correlated (ρ averages 0.35 over 2015–2021, peaking at 0.44 in 2020), declining further after the Ukraine invasion (ρ averages 0.24 post-2022) as Russia's profile becomes dominated by a single target (Figure 3).[6,7]

A trade analyst applying the same conflict coefficient to both the US–China and US–Russia relationships will get both wrong. The former is overwhelmingly trade-context hostility; the latter is increasingly kinetic. An aggregate indicator cannot make this distinction.

3.4 Geopolitical Alignment and Bloc Formation

We measure alignment using the profile similarity ρ defined in footnote 7. Countries whose hostility profiles are highly correlated (e.g., both hostile toward Russia and Iran, both conciliatory toward the same partners) are aligned. The alignment gap $\Delta_c = \rho(c, \text{USA}) - \rho(c, \text{CHN})$ tracks where each country sits between the two poles, ranging from strongly US-aligned (high positive) to strongly China-aligned (high negative).

Three findings emerge. First, alliance cohesion is markedly asymmetric. Western groupings exhibit high and stable profile cohesion: in 2022, AUKUS reaches 0.63, the G7 0.59, and Five Eyes 0.43. Non-Western groupings are an order of magnitude less cohesive in the same year: BRICS sits at just 0.03, the SCO at 0.11.[8] The gap reflects a structural difference in how these groupings operate. Western cohesion is amplified by coordinated sanctions: when the G7 imposes sanctions on Russia, all members' hostility profiles shift in the same direction simultaneously, mechanically raising pairwise correlation. BRICS members, by contrast, have divergent threat environments: India faces Pakistan and China, Brazil faces no major bilateral adversary, and Russia's profile is dominated by

[6] Top five targets by mean calibrated hostility (2015–2025). China: USA (0.17), Taiwan (0.12), Philippines (0.09), North Korea (0.09), Australia (0.08). Russia: Ukraine (0.53), USA (0.22), Syria (0.17), UK (0.12), Poland (0.10). Post-2022, Russia's hostility toward Ukraine rises from 0.29 to 0.93, compressing the relative weight of all other targets and driving the profile divergence.

[7] Profile similarity $\rho(i, j)$ is the Pearson correlation between two countries' outgoing hostility vectors toward common targets. For each country i in year t, construct a vector $\mathbf{v}_{i,t}$ whose k-th entry is the annual mean of the monthly calibrated hostility $r^{\text{cal}}_{i \to k,s}$ over months s in year t (the simple average, not the exponential stock of equation 3). Profile similarity is $\rho(i, j) = \text{corr}(\mathbf{v}_{i,t}, \mathbf{v}_{j,t})$, computed over the set of third countries toward which both i and j have nonzero annual mean hostility in that year (excluding i and j themselves), requiring at least five such shared targets. High ρ means two countries direct hostility toward the same targets in similar relative proportions.

[8] AUKUS: Australia, UK, USA. Five Eyes: AUKUS plus Canada and New Zealand. G7: Canada, France, Germany, Italy, Japan, UK, USA. BRICS: Brazil, Russia, India, China, South Africa. SCO: Shanghai Cooperation Organisation (China, Russia, India, Pakistan, and Central Asian members). Cohesion is the mean pairwise profile similarity ρ within each grouping in a given year; profile similarity itself is defined from annual-mean hostility vectors (footnote 7), so cohesion is year-specific and not a full-sample average.

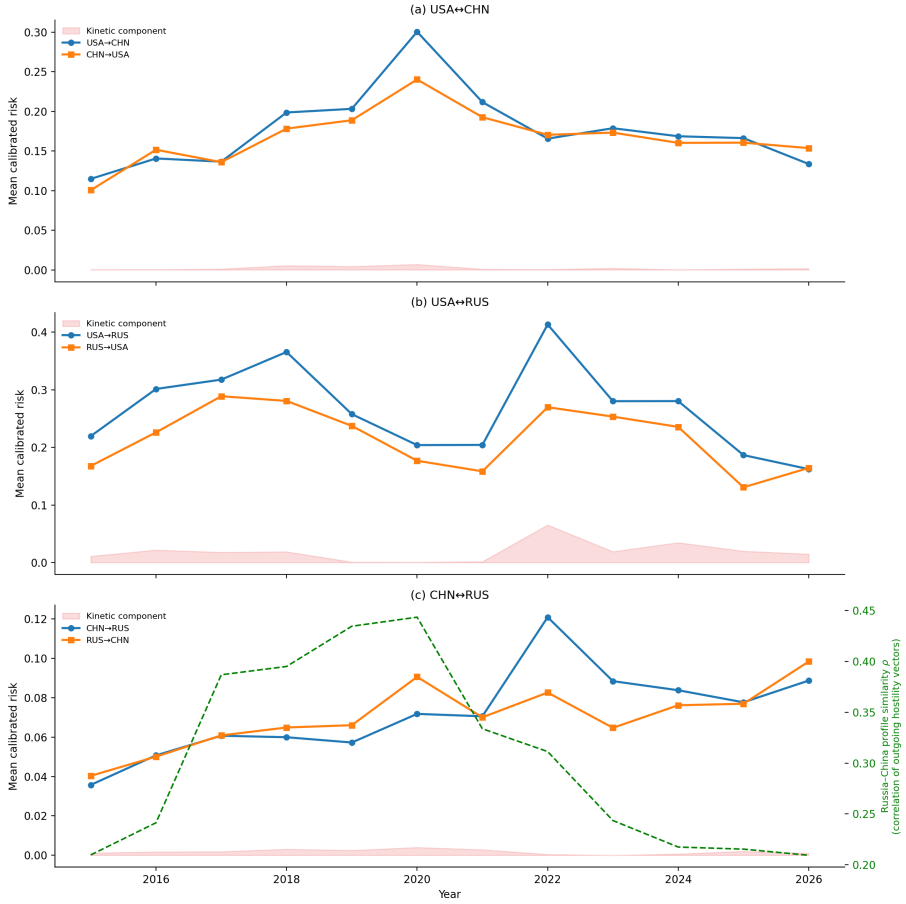

Figure 3: Great power rivalries, 2015–2025. Annual means. Panel (c) includes CHN–RUS profile similarity.

Ukraine. Their hostility profiles pull in different directions even when political declarations suggest unity.

Second, the non-aligned middle is large and heterogeneous. India sits near zero on the alignment gap throughout the sample: its hostility profile correlates moderately with both poles, driven by tensions with both China (border disputes) and Pakistan (a US security partner). Turkey swings sharply between 2015 and 2018 (from +0.44 to −0.37) before stabilising near zero after the 2019 S-400 purchase from Russia, consistent with its hedging between NATO obligations and an independent foreign policy. Saudi Arabia moves in the opposite direction from what the Yemen conflict and the Iran confrontation might suggest: from a 2017 peak of +0.20 (its most US-aligned year) it drifts toward China-alignment, reaching −0.24 in 2023 on the back of the 2023 China-brokered rapprochement with Iran and divergences from the US over oil policy. The alignment trajectories of non-aligned countries are themselves informative, not a residual to be cleaned up.

Third, the between-bloc to within-bloc hostility ratio slightly exceeds 1.0 during event months, indicating mild fragmentation rather than sharp bipolar polarisation (Appendix, Figure A2). Countries are somewhat more hostile toward opposing blocs than toward their own, but the margin is narrow. The world is not cleanly divided into two hostile camps (Figure 4).

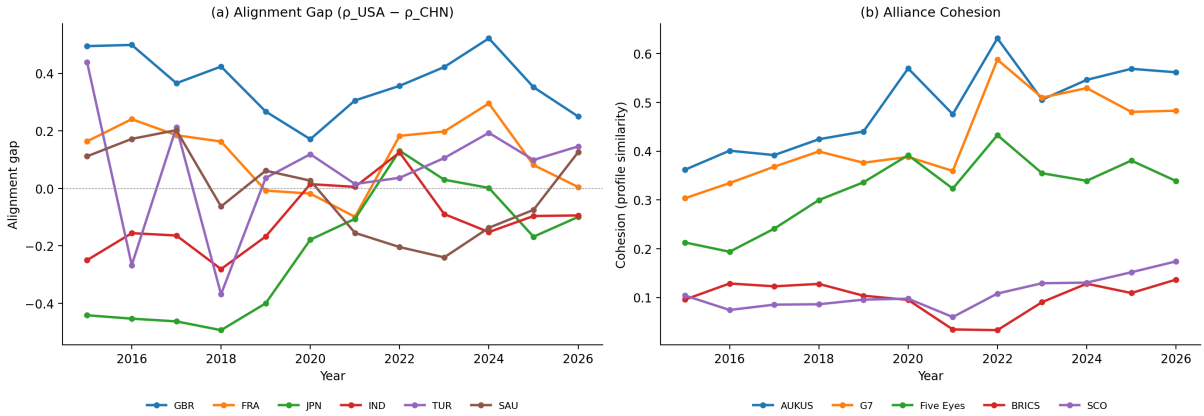

Figure 4: (a) Alignment gap: $\Delta_c = \rho(c, \text{USA}) - \rho(c, \text{CHN})$. (b) Alliance cohesion: mean pairwise profile similarity.

3.5 The Weaponisation of Economic Policy

The weaponisation index $W_i = \sum_j L_{ij}^{\text{trd}} / \sum_j (L_{ij}^{\text{kin}} + L_{ij}^{\text{pos}} + L_{ij}^{\text{trd}})$ measures how much of a country's actionable hostile activity is channelled through economic instruments (the denominator excludes L^{base}, which captures residual diplomatic tension rather than coercive action). China leads the ranking ($W_{\text{CHN}} = 0.83$), with major European democracies and India clustering behind at 0.74–0.77. The USA sits much lower ($W_{\text{USA}} \approx 0.60$) because its extensive kinetic activity abroad enlarges the denominator. Russia's weaponisation index declines slightly after 2022 as its profile militarises (0.64 to 0.59). Democracies systematically weaponise economics: democracy-to-autocracy pairs show the highest L^{trd} shares. Trade-weighted global weaponisation rose steadily over 2015–2023 (Figure 5).

The pattern is consistent with a substitution story: democracies, constrained from military action by domestic audience costs (Fearon, 1994), channel conflict through sanctions

and trade restrictions. The gravity results confirm that this channel is economically potent, not merely symbolic.

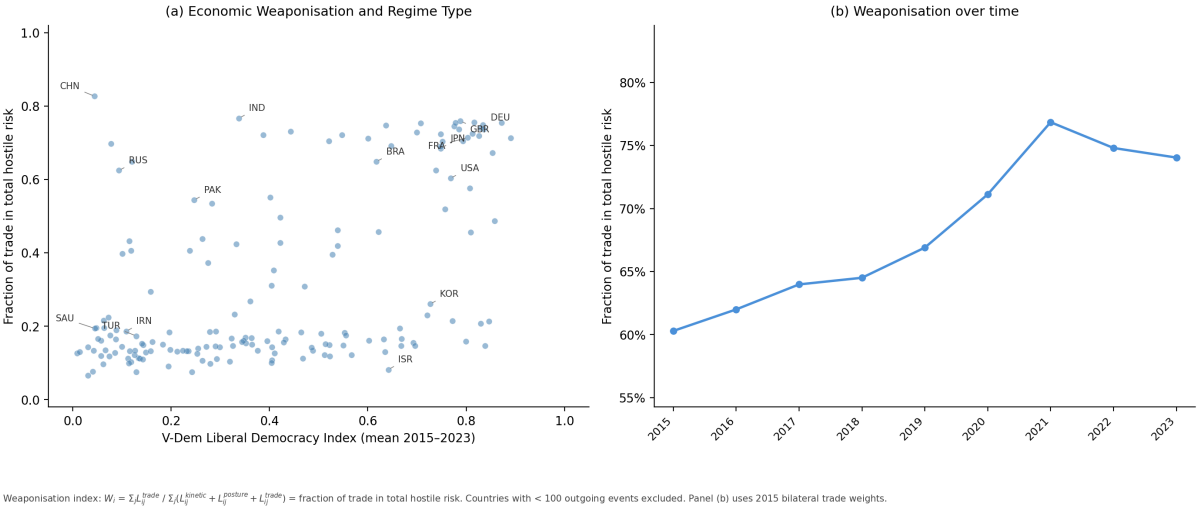

Figure 5: Weaponisation index vs V-Dem liberal democracy index, and trade-weighted time trend.

3.6 Sanctions Trajectories

Figure 6 tracks the stock variable ($\bar{r}_{ij,t}$, equation 3) around the onset of formal sanctions for four pairs. The stock is better suited than the monthly flow for this exercise because its exponential smoothing (half-life $\approx$ 14 months) filters out month-to-month noise and reveals the underlying trend in bilateral hostility. In every case, the stock was already rising before sanctions were imposed. Sanctions amplify pre-existing tensions; they do not create hostility from zero. This pattern supports the interpretation that sanctions respond to escalating conflict rather than to unobserved trade shocks.

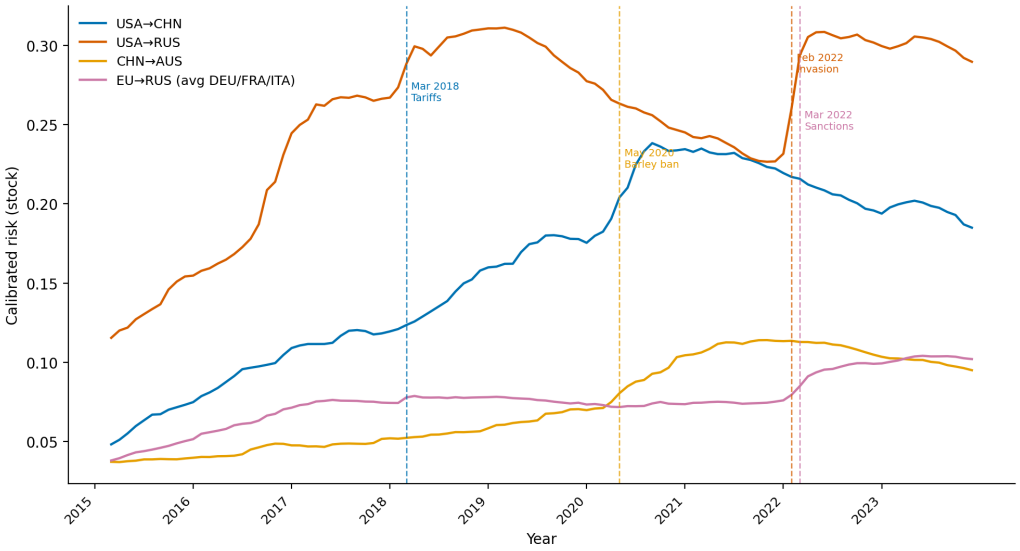

Figure 6: Stock ($\lambda = 0.95$) trajectories around sanctions onset.

3.7 Early Warning: The Ukraine Case

A growing conflict forecasting literature (Ward et al., 2010; Goldstone et al., 2010; Mueller and Rauh, 2018; Bazzi et al., 2022; Petrova and Tapsoba, 2025) predicts onset from structural variables or news sources. Our decomposed layers of bilateral risk may also help, because different components can serve as leading indicators of escalation.

Consider RUS→UKR. The posture component (L^{pos}) captures military force displays, troop movements, and exercises near borders. We define an anomaly threshold as the mean plus three standard deviations of L^{pos} over a pre-escalation baseline (2015–2020).[9] This threshold was breached in April 2021, ten months before the February 2022 invasion, when Russia began its large-scale military build-up along the Ukrainian border. Only one false positive occurred in 72 baseline months (August 2016, coinciding with a brief spike in Russian military activity near Crimea), giving a pair-specific false positive rate of 1.4%.

Crucially, the political/diplomatic component does *not* provide the same signal. For this pair, baseline diplomacy is negatively correlated with subsequent kinetic escalation: diplomatic rhetoric intensifies during periods of relative calm and quiets when military action is being prepared. An aggregate indicator, dominated by the diplomatic component, would have detected the escalation only when the kinetic component surged *after* the invasion had begun. The decomposition separates the military preparation signal from the diplomatic noise (Figure 7).

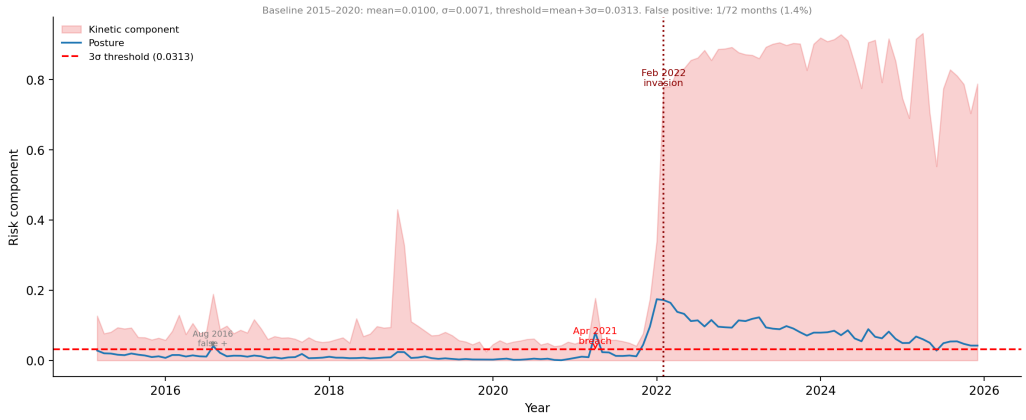

Figure 7: L^{pos} for RUS→UKR. 3σ threshold from 2015–2020 baseline.

3.8 Decoupling Risk

Among the top 500 trading pairs, most show stable or declining conflict risk. Extreme escalation is concentrated in a few pairs involving Russia (post-2022) or China (post-2018). The global trading system is predominantly resilient (Figure 8). Maritime route risk is examined in Appendix D.

[9]The 3σ threshold is a standard statistical process control rule. For RUS→UKR, the 2015–2020 baseline (72 months) has mean = 0.010 and σ = 0.007, giving a threshold of 0.031. The April 2021 breach registers $L^{\text{pos}} = 0.079$, rising to 0.172 in February 2022 (invasion month). The single false positive (August 2016) is 0.041, barely above threshold. The ISR→PSE pair illustrates a different pattern. There, the kinetic component itself spikes abruptly with the October 2023 Gaza escalation, with no posture lead time. This is consistent with the nature of the event: a surprise attack and immediate retaliation. The decomposition does not predict all escalations, but it may help distinguishing those preceded by observable military preparation from those that are not.

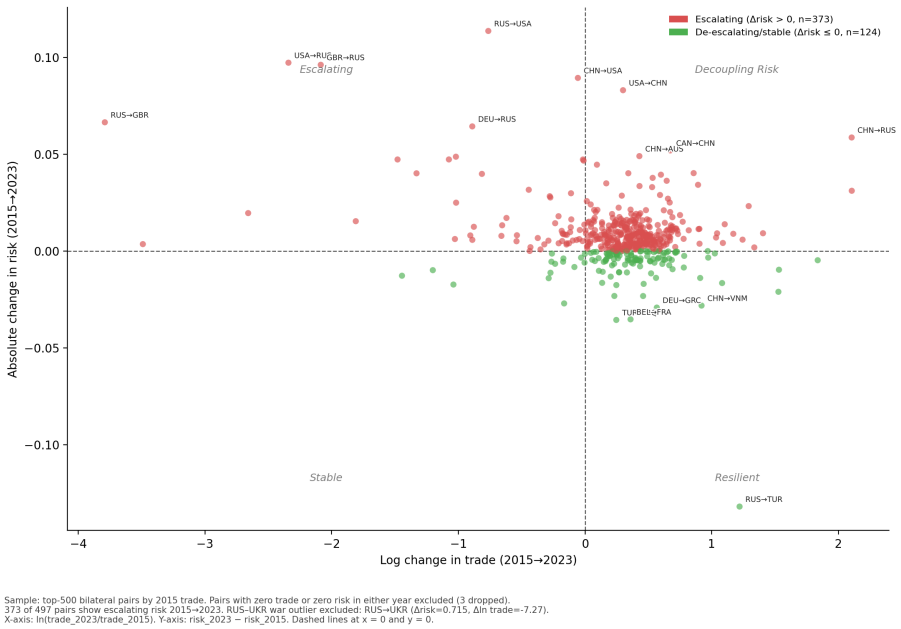

Figure 8: Decoupling risk: top 500 pairs by 2015 trade. Risk change computed as the difference of unconditional annual means of calibrated hostility between 2023 and 2015 (trade-present months; zero-event months count as zero).

4 Trade Effects: Gravity Estimation

This section estimates the trade effects of each component in a structural gravity framework. If the components have different coefficients, the decomposition captures economically meaningful distinctions. If they do not, the patterns above are descriptively interesting but irrelevant for trade.

4.1 Trade Data

The dependent variable is monthly bilateral trade from UN COMTRADE, covering 200 countries over 2015–2025. For each directed pair-month, we observe both the exporter's reported shipments (FOB) and the importer's reported receipts (CIF). Mirror reconciliation averages these two reports, expanding coverage: roughly 8% of pair-months are reported by only one side. Comtrade records positive trade flows only, so a pair with no trade in a given month is simply absent from the source rather than reported as zero. The baseline panel therefore contains no explicit zeros; pair-months where neither side reported trade drop out, which is equivalent to conditioning on the intensive margin. Table 6 column (8) addresses this by rebuilding the panel for the active pair set (pairs with at least one positive trade month in 2015–2023) and zero-filling missing trade; the preferred coefficients strengthen slightly, confirming that the baseline is conservative. For the pairs of most interest, trade declined sharply but remained positive in Comtrade (RUS→UKR reaches a monthly minimum of about $40 in 2023, i.e., forty dollars, not millions, down from $582 million in 2021), so the intensive margin captures most of the variation.[10] After

[10] Annualised, the mirror-reconciled trade data correlates at 0.975 in logs with the independently cleaned Atlas of Economic Complexity dataset (Growth Lab at Harvard, 2024), which applies the Bustos-Yildirim method to the same raw COMTRADE records. Our values are roughly 9% higher, consistent with the CIF/FOB margin introduced by averaging exporter and importer reports. This margin is nearly constant

restricting to observations with non-missing conflict data (the sanctions flag is available through 2023, limiting the effective sample), the estimation sample contains 1.5–1.9 million observations (depending on specification) across roughly 24,000 directed pairs and 200 countries. All four conflict components are heavily right-skewed: over two-thirds of pair-months have zero conflict, while the 99th percentile reaches 2–5 percentage points of bilateral interaction (Table A7 in the Appendix). Baseline diplomacy has the highest mean, consistent with Section 3. Kinetic risk is rare but extreme. Trade spans four orders of magnitude from the 25th to the 75th percentile.

4.2 Specification

We estimate:

$$x_{ij,t} = \exp\left(\sum_s \beta_s \cdot \text{Risk}_{ij,t-s} + \alpha_{ij,m} + \gamma_{i,t} + \delta_{j,t}\right) + \varepsilon_{ij,t} \tag{9}$$

where $x_{ij,t}$ is bilateral monthly trade (mirror-reconciled COMTRADE exports). The fixed effects are: pair×calendar-month ($\alpha_{ij,m}$), exporter×year-month ($\gamma_{i,t}$), and importer×year-month ($\delta_{j,t}$), following Santos Silva and Tenreyro (2006). Standard errors are clustered by pair.

Domestic conflict is absorbed by the country×time fixed effects. The conflict variable enters in the $i \to j$ direction: the exporter's hostility toward the importer. Lags enter sequentially: L_1, then $\bar{L}_{2\text{-}4}$ (average of lags 2–4), then $\bar{L}_{5\text{-}7}$ (average of lags 5–7). The cumulated effect $\hat{\beta}_1 + \hat{\beta}_{2\text{-}4} + \hat{\beta}_{5\text{-}7}$ measures the total trade disruption from a sustained unit increase in risk over seven months.

Identification. Pair×calendar-month fixed effects absorb all time-invariant bilateral characteristics and bilateral seasonality. Country×year-month fixed effects absorb all country-specific time-varying shocks, including domestic conflict, GDP fluctuations, and multilateral resistance. The identifying variation is bilateral conflict escalation or de-escalation that cannot be attributed to either country individually.

Two features of the results follow if the coefficients capture a genuine conflict-to-trade channel rather than the influence of unobservables. First, the temporal patterns should differ across components, reflecting distinct economic mechanisms: infrastructure destruction is immediate, policy uncertainty accumulates, and counter-sanctions take months to enact. A single confounding shock would produce contemporaneous correlations across all components, not this structured heterogeneity. Second, baseline diplomatic friction, which correlates with bilateral shocks just as much as other components, should have no trade effect. The results below match both predictions. The robustness analysis (Section 5) further shows that adding pair-specific linear trends or lagged trade levels leaves long-run coefficients and the temporal structure intact, which would not hold if reverse causality from trade to conflict type were driving the estimates.

within a pair (same shipping route and distance) and is absorbed by the pair×calendar-month fixed effects; only month-to-month changes in freight costs could affect identification, and these are small relative to the trade variation we exploit.

4.3 Main Results

Tables 2–4 present the estimates, building from aggregate to decomposed to preferred specification.

Table 2: Aggregate Bilateral Conflict Risk and Trade

	Calibrated			Raw
	(1) L1 only	(2) $+\bar{L}_{2–4}$	(3) $+\bar{L}_{5–7}$	(4) 7-month
L_1	-0.744^{***}	-0.289^{***}	-0.218^{***}	-0.060^{***}
	(0.146)	(0.057)	(0.055)	(0.016)
$\bar{L}_{2–4}$		-1.064^{***}	-0.630^{***}	-0.197^{***}
		(0.171)	(0.094)	(0.039)
$\bar{L}_{5–7}$			-0.813^{***}	-0.188^{***}
			(0.188)	(0.050)
Cumulated	-0.744^{***}	-1.353^{***}	-1.662^{***}	-0.444^{***}
	(0.146)	(0.215)	(0.280)	(0.096)
Pair FE	Yes	Yes	Yes	Yes
Country×Month FE	Yes	Yes	Yes	Yes
Observations	1,856,916	1,642,555	1,525,222	1,525,222

Notes: PPML-HDFE. Dependent variable: bilateral monthly trade. All specifications include pair×calendar-month, exporter×year-month, and importer×year-month fixed effects. Standard errors clustered by pair in parentheses. Columns (1)–(3) use the calibrated conflict risk measure; column (4) uses the raw Goldstein-weighted measure. L_1 is the one-month lag; $\bar{L}_{2–4}$ and $\bar{L}_{5–7}$ are averages over lags 2–4 and 5–7, respectively. Cumulated reports the sum of all included lags. Sample: 2015–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 presents the aggregate gravity results. Columns (1)–(3) build up the lag structure progressively for the calibrated indicator: a single lag yields a cumulated effect of −0.74; adding the second and third lag groups deepens it to −1.66 (column 3), indicating that conflict effects accumulate over months. Repeating the full specification with the raw (uncalibrated) indicator in column (4) yields a cumulated effect nearly four times smaller (−0.44), because the raw indicator is systematically inflated by GDELT overcoding: a unit change in the raw score represents less genuine hostility than a unit change in the calibrated score. On a standardised basis the gap narrows: a one-standard-deviation increase in the calibrated indicator (SD = 0.022) implies a trade reduction of −0.037, versus −0.034 for the raw indicator (SD = 0.078). The fixed effects absorb much of the systematic measurement error, so both indicators capture similar total variation in trade. Nevertheless, calibration remains essential: it produces a standalone indicator whose levels are interpretable across pairs, and it enables a better domain decomposition.[11]

Table 3 enters all four components simultaneously. Kinetic conflict has the largest cumulated effect (column 1: −4.19). Trade-context hostility is also statistically significant

[11]Repeating the four-component regression with raw GDELT domains yields statistically significant coefficients for all four components, including the political/diplomatic domain (Appendix E, Table A1). Without the sanctions reclassification, diplomatic events during sanctions episodes remain in the poldip domain rather than being reassigned to trade, so the raw poldip coefficient likely captures trade-relevant variation misallocated to the wrong domain.

Table 3: Four-Component Decomposition: Full Model

	(1) Kinetic	(2) Posture	(3) Trade_{ij}	(4) Baseline
L_1	-0.857^{***}	-1.374^{*}	-0.157^{**}	0.133^{*}
	(0.158)	(0.724)	(0.072)	(0.081)
$\bar{L}_{2\text{–}4}$	-1.549^{***}	-1.502	-0.667^{***}	-0.105
	(0.382)	(1.624)	(0.211)	(0.118)
$\bar{L}_{5\text{–}7}$	-1.781^{***}	-1.205	-0.747^{***}	-0.057
	(0.425)	(1.997)	(0.227)	(0.174)
Cumulated	-4.187^{***}	-4.082	-1.571^{***}	-0.029
	(0.844)	(3.553)	(0.447)	(0.327)
Joint χ^2 (all lags= 0)		3.65		8.67
p-value		0.302		0.034
Pair FE		Yes		
Country×Month FE		Yes		
Observations		1,525,222		

Notes: PPML-HDFE. Dependent variable: bilateral monthly trade. All specifications include pair×calendar-month, exporter×year-month, and importer×year-month fixed effects. Standard errors clustered by pair in parentheses. All four components enter a single regression. Joint χ^2 tests the null that all three lag groups (L_1, $\bar{L}_{2\text{–}4}$, $\bar{L}_{5\text{–}7}$) are jointly zero; the test is reported only for posture (column 2) and baseline (column 4), the two components whose cumulated effects are not individually significant. Sample: 2015–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(column 3: -1.56, $p < 0.01$). Posture is statistically insignificant (column 2), and baseline is indistinguishable from zero (column 4). Only two of the four layers have measurable trade effects, confirming that the decomposition captures economically meaningful distinctions.

The insignificance of posture and baseline is consistent with the indicator construction. Baseline captures residual diplomatic tension for unsanctioned pairs, which should not affect trade directly; its near-zero cumulated coefficient is expected, although the joint test of its three lag groups rejects at the 5% level (Table 3): a small positive first-lag effect (+0.15) is offset by slightly negative later lags. The *long-run* effect is what matters for trade policy, and it is essentially zero. Posture is harder to interpret. Three factors contribute to its insignificance: posture events are rare and concentrated in a few pairs, limiting statistical power; the posture/baseline boundary is conceptually ambiguous (Section H), so measurement error attenuates the coefficient; and in the early warning analysis (Section 3.7), posture is a leading indicator of kinetic escalation rather than an independent channel, so its trade effect may operate indirectly through subsequent kinetic events. The gravity results should not be read as evidence that military preparation is economically irrelevant.

Table 4 builds toward the preferred specification. Column (1) enters kinetic and trade_{ij} only. Column (2), our preferred specification, adds the target's trade hostility toward the sender (L_{ji}^{trd})[12]. Both kinetic and L_{ij}^{trd} attenuate slightly (kinetic from -4.67 to -4.54,

[12]The two trade-context directions are only weakly correlated within pairs over time ($\rho = 0.15$ after pair fixed effects, compared to $\rho = 0.54$ in the cross-section of pair means). The high cross-sectional

Table 4: Parsimonious Specification and Extensions

	(1) Kinetic+Trade	(2) Preferred	(3) +Route	(4) $+\bar{L}_{8\text{–}10}$
Cumulated effects				
Kinetic	-4.666^{***}	-4.544^{***}	-4.535^{***}	-5.428^{***}
	(0.631)	(0.614)	(0.604)	(0.807)
Trade_{ij}	-1.547^{***}	-1.400^{***}	-1.399^{***}	-1.407^{***}
	(0.260)	(0.251)	(0.250)	(0.254)
Trade_{ji}		-0.527^{***}	-0.529^{***}	-0.536^{***}
		(0.130)	(0.131)	(0.141)
Additional controls				
Route risk			-0.011	
			(0.030)	
Total trade-context $(ij + ji)$		-1.927^{***}		-1.943^{***}
		(0.303)		(0.315)
Lag groups	3	3	3	4
Cumulation horizon	7 mo.	7 mo.	7 mo.	10 mo.
Pair FE	Yes	Yes	Yes	Yes
Country×Year-month FE	Yes	Yes	Yes	Yes
Observations	1,525,222	1,525,222	1,525,222	1,434,020

Notes: PPML-HDFE. Dependent variable: bilateral monthly trade. All specifications include pair×calendar-month, exporter×year-month, and importer×year-month fixed effects. Standard errors clustered by pair in parentheses. All cells show cumulated effects over the indicated horizon. Column (1): kinetic + trade_{ij} only. Column (2): preferred = adds trade_{ji} (retaliation). Column (3): adds route risk (average bilateral conflict risk along the maritime route $i \to j$). Column (4): extends the lag structure to 10 months (4 lag groups: L_1, $\bar{L}_{2\text{–}4}$, $\bar{L}_{5\text{–}7}$, $\bar{L}_{8\text{–}10}$). Sample: 2015–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

trade$_{ij}$ from -1.55 to -1.40), consistent with a small negative omitted-variable bias: L^{trd}_{ji} is positively correlated with both within-pair regressors and enters with a negative coefficient, so excluding it pushes the included coefficients slightly away from zero. The retaliation channel is statistically significant ($p < 0.01$) and adds roughly 27% to the trade-context effect (cumulated total $ij + ji$: -1.93). Column (3) adds the average conflict risk along the maritime shipping route between i and j; its coefficient is close to zero and insignificant (-0.01), and the three bilateral coefficients barely move. The time fixed effects are likely to capture common maritime route shocks. Column (4) extends the lag structure to 10 months; cumulated effects deepen, confirming that conflict effects persist beyond the 7-month window.

The three statistically significant components display distinct temporal patterns (Figure 9). Kinetic effects are immediate and deepen monotonically. Trade-context hostility builds gradually over months. Retaliation is delayed: statistically insignificant at short horizons, statistically significant only at longer ones. This temporal heterogeneity is consistent with economic mechanisms. Infrastructure destruction is immediate; retaliatory sanctions take months to enact and for firms to adjust. A single confounding shock could not produce all three patterns simultaneously.

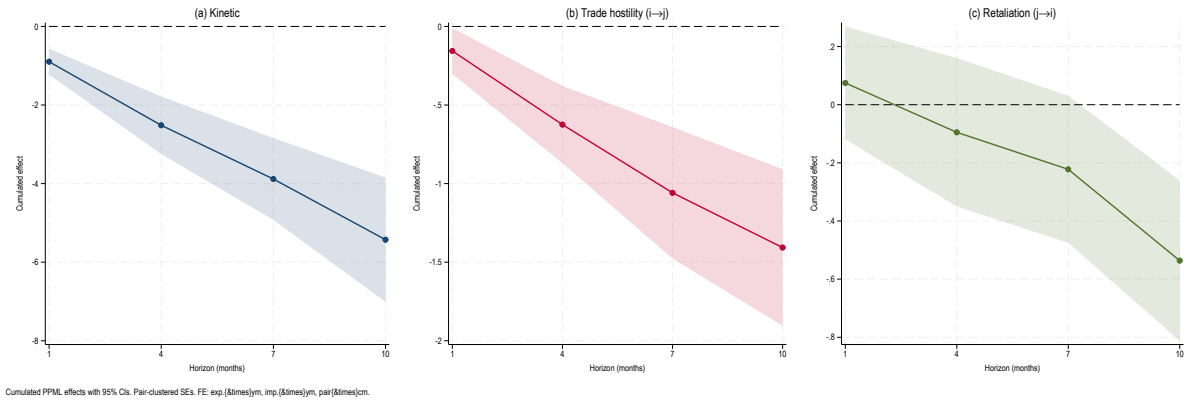

Figure 9: Trade response by lag horizon: cumulated effects at horizons 1, 4, 7, and 10 months with 95% confidence intervals. Pair-clustered standard errors. Fixed effects: exporter×year-month, importer×year-month, pair×calendar-month.

4.4 Economic Magnitude

We quantify the aggregate trade cost of conflict escalation since 2015. Let $L_{ij} = \hat{\beta}_{\text{kin}}\Delta\bar{L}^{\text{kin}}_{ij} + \hat{\beta}_{\text{trd}}\Delta\bar{L}^{\text{trd}}_{ij} + \hat{\beta}_{\text{trd},ji}\Delta\bar{L}^{\text{trd}}_{ji}$ denote the pair-level log-loss, where $\Delta\bar{L} = \bar{L}_{2023} - \bar{L}_{2015}$. The total trade at risk and the channel contributions are:

$$\text{Total}_{ij} = x_{ij,2015} \times \big(1 - \exp(L_{ij})\big) \tag{10}$$

$$\text{Channel}^{k}_{ij} = \frac{\hat{\beta}_k \Delta\bar{L}^{k}_{ij}}{L_{ij}} \times \text{Total}_{ij}, \qquad k \in \{\text{kin}, \text{trd}, \text{trd}, ji\} \tag{11}$$

correlation reflects the fact that adversary pairs are hostile in both directions on average. However, it is absorbed by country-pair fixed effects. The retaliation coefficient is identified from month-to-month asymmetric movements: months where one direction escalates faster than the other.

where $x_{ij,2015}$ is pre-escalation trade. Using 2015 weights avoids circularity: 2023 trade has already been reduced by the conflict we measure. Differencing isolates the cost of *escalation*, not the total cost of pre-existing conflict. By construction, the channel contributions sum exactly to the total within each pair, so aggregating across pairs yields channel totals that sum exactly to the global $334 billion figure.

Table 5: Trade at Risk: Decomposition by Pair (2015–2023)

	Pair	x_{2015} ($B)	$\Delta\bar{L}$ (2023−2015) Kin	$\Delta\bar{L}$ (2023−2015) Trd_{ij}	$\Delta\bar{L}$ (2023−2015) Trd_{ji}	Trade at risk ($B) Total	Trade at risk ($B) Kin	Trade at risk ($B) Trd_{ij}	Trade at risk ($B) Trd_{ji}
(1)	CHN→USA	503	−0.001	+0.168	+0.168	136.6	−2.6	101.1	38.1
(2)	USA→CHN	116	+0.002	+0.168	+0.168	32.7	0.7	23.3	8.7
(3)	USA→MEX	212	+0.001	+0.083	−0.001	24.0	1.0	23.1	−0.1
(4)	MEX→USA	303	+0.003	−0.001	+0.083	16.7	4.4	−0.6	12.8
(5)	RUS→UKR	7	+0.802	−0.222	−0.263	6.3	7.2	−0.6	−0.3
(6)	CAN→USA	321	+0.004	+0.000	+0.001	6.2	6.0	0.0	0.1
(7)	KOR→CHN	137	+0.002	+0.015	+0.012	5.1	1.4	2.8	0.9
(8)	CHN→IND	61	+0.011	+0.007	+0.027	4.4	3.0	0.5	0.9
(9)	JPN→KOR	45	+0.002	+0.044	+0.047	4.1	0.4	2.6	1.1
(10)	CHN→GBR	63	−0.001	+0.034	+0.041	4.0	−0.2	2.9	1.3
(11)	CHN→CAN	56	−0.003	+0.042	+0.046	3.8	−0.7	3.2	1.3
(12)	CHN→AUS	47	−0.000	+0.042	+0.025	3.3	0.0	2.7	0.6
(13)	USA→CAN	261	+0.002	+0.001	+0.000	3.2	2.9	0.2	0.0
(14)	ISR→USA	21	+0.016	+0.000	+0.156	3.1	1.5	0.0	1.6
(15)	CHN→DEU	103	+0.000	+0.012	+0.025	3.1	0.0	1.8	1.3
	Top 15 subtotal					257 (77%)			
Controls									
	FRA→DEU	77	+0.004	+0.001	+0.000	1.6	1.5	0.1	0.0
	DEU→NLD	77	−0.000	+0.000	+0.000	−0.1	−0.1	0.1	0.0
	All pairs					**334**	**10** (3%)	**228** (68%)	**96** (29%)

Notes: Coefficients from Table 4, column (2). Total and channel contributions are computed as in equations (10)–(11). By construction, channel values sum exactly to Total within each pair and across all pairs. A negative channel value indicates that the channel moved opposite to the net effect: for example, RUS→UKR (row 5) shows negative trade-context contributions because trade-context hostility fell as the relationship militarised, partially offsetting the dominant kinetic escalation. 2015 trade weights avoid circularity.

Table 5 decomposes this aggregate into pair-level contributions. The top 15 pairs account for three-quarters of the $334 billion total. Three patterns stand out.

First, the US–China pair dominates: CHN→USA alone accounts for $137B (row 1), with symmetric trade-hostility escalation in both directions ($\Delta L^{\text{trd}} \approx +0.17$). The kinetic contribution is negligible. Together, the two directions of the US–China relationship account for half the global total.

Second, the US–Mexico pair contributes $41B (rows 3–4), driven by rising trade-context hostility. The channel structure is asymmetric: USA→MEX is driven by sender hostility ($23B in Trd_{ij}), while MEX→USA is driven by the retaliation channel ($13B in Trd_{ji}), reflecting Mexican exports affected by US hostility.

Third, RUS→UKR (row 5) contributes $6.3B despite small baseline trade ($7B), driven entirely by kinetic escalation ($\Delta L^{\text{kin}} = +0.80$). Trade-context hostility actually *falls* as the relationship militarises, partially offsetting the kinetic destruction. This is the only pair in the top 15 where the kinetic channel dominates. Row 14 (ISR→USA) is a related outlier: its $1.6B retaliation contribution is driven almost entirely by a 2023 spike in US trade-context hostility toward Israel ($L^{\text{trd}}_{\text{USA}\to\text{ISR}}$ jumps from 0.008 in 2022 to 0.164 in 2023), reflecting US diplomatic criticism and threatened arms restrictions during the October 2023 Gaza escalation rather than implemented sanctions.

The channel decomposition across all pairs is striking: trade-context hostility accounts for 68% of the global total ($228B), the retaliation channel for 29% ($96B), and kinetic conflict for only 3% ($10B). The world's trade losses from geopolitical deterioration are overwhelmingly economic, not military. Control pairs (FRA→DEU, DEU→NLD) show near-zero predicted impacts, confirming that the model assigns economically negligible effects to pairs with stable bilateral relations.

This is a partial equilibrium calculation that ignores trade diversion. It overstates bilateral welfare losses but bounds the trade potentially destroyed by the conflict escalation of the past decade: roughly 2.4% of 2015 world bilateral trade.[13]

5 Robustness and Alternative Specifications

5.1 Robustness Specifications

Table 6 presents eight specifications. Column (1) reproduces the preferred specification from Table 4, column (2). Columns (2)–(3) test whether the intensity of trade hostility matters beyond binary sanctions status. Adding sanctions dummies (sender only in (2), bilateral in (3)) leaves them insignificant, while L^{trd} retains its statistical significance: the *intensity* of hostility matters, not sanctions status per se. Column (4) replaces the decomposed monthly flows with the exponentially smoothed aggregate stock ($\lambda = 0.95$). Its coefficient (-3.1) is roughly double the aggregate flow estimate (-1.66 in Table 2, column 3), because a unit change in the stock represents a persistent shift in the conflict environment, not a single-month spike.

Columns (5) and (6) address reverse causality: pairs that trade intensely may generate more trade disputes, while pairs with little trade may face lower costs of military escalation. If the level of trade drives the type of conflict observed, controlling for that level should attenuate the decomposition results. Column (5) adds a pair-specific linear time trend on top of pair×calendar-month fixed effects. This is a demanding specification that absorbs any secular drift within a pair, including genuine conflict escalation. The attenuation is unsurprising: for pairs like US–China or Russia–Ukraine, the escalation *is* the trend. Column (6) provides a cleaner test. Three lagged dependent variables, mirroring the conflict lag structure, control directly for the trade level at each horizon. Short-run effects are attenuated by high trade persistence, but long-run effects recover the baseline magnitudes. The temporal structure is also preserved: kinetic remains immediate, retaliation remains delayed. If reverse causality from trade to conflict type were driving the results, conditioning on lagged trade would substantially alter the long-run coefficients.

[13] Complete trade embargoes (e.g., USA→RUS post-2022) produce collapses beyond the model's marginal framework. These pairs account for less than 1% of 2015 baseline world trade.

Table 6: Robustness: Alternative Specifications

	(1) Preferred	(2) $+D_{ij}^{\text{sanc}}$	(3) $+D_{ij,ji}^{\text{sanc}}$	(4) Stock95	(5) +Pair trend	(6) LDV	(7) ×Init. host.	(8) Zero-fill
Cumulated effects (7 months)								
Kinetic	−4.544***	−4.536***	−4.556***		−3.066***	−1.428***	−6.064***	−4.852***
	(0.614)	(0.614)	(0.615)		(0.676)	(0.238)	(1.007)	(0.586)
Trade_{ij}	−1.400***	−1.305***	−1.340***		−0.530**	−0.408***	−0.881	−1.458***
	(0.251)	(0.295)	(0.284)		(0.233)	(0.083)	(0.641)	(0.251)
Trade_{ji}	−0.527***	−0.540***	−0.403**		0.158	−0.146***	−0.409	−0.554***
	(0.130)	(0.130)	(0.199)		(0.224)	(0.044)	(0.709)	(0.144)
Additional controls								
D_{ij}^{sanc}		−0.014	−0.010					
		(0.021)	(0.021)					
D_{ji}^{sanc}			−0.023					
			(0.023)					
Aggregate (stock95)				−3.110***				
				(0.537)				
$\hat{\rho}$						0.728***		
						(0.011)		
Long-run effects: SR $/(1-\hat{\rho})$								
Kinetic						−5.249***		
						(0.876)		
Trade_{ij}						−1.498***		
						(0.307)		
Trade_{ji}						−0.536***		
						(0.160)		
Interactions (× initial hostility, z-scored)								
Kinetic × stock							0.227**	
							(0.109)	
Trade_{ij} × stock							−0.144	
							(0.173)	
Trade_{ji} × stock							−0.032	
							(0.184)	
Joint p-value							[0.150]	
Pair×cal. mo. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair trend	No	No	No	No	Yes	No	No	No
Observations	1,525,222	1,525,222	1,525,222	1,532,988	1,525,222	1,525,222	1,525,222	2,698,701

Notes: PPML-HDFE. Dependent variable: bilateral monthly trade. All specifications include pair×calendar-month, exporter×year-month, and importer×year-month fixed effects. Standard errors clustered by pair in parentheses. (1) preferred. (2)–(3) add sanctions dummies. (4) exponential stock ($\lambda = 0.95$); the stock lags are populated for early-2015 months from the GDELT v1 burn-in (2009–Jan 2015), whereas the flow-based lags in the other columns are missing for those months, yielding a slightly larger sample. (5) pair linear trend. (6) three LDVs mirroring the lag structure; $\hat{\rho}$ = sum of LDV coefficients; long-run = SR/$(1-\hat{\rho})$. (7) interactions with Mar 2015 `stock95` (z-scored); level absorbed by pair FE. (8) balanced active-pair panel with zero-filled missing trade (pairs with at least one positive trade month in 2015–2023 are expanded to all months; missing trade is set to zero, and missing conflict components are set to zero). Sample: 2015–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

It does not.

Column (7) tests whether pre-existing hostility amplifies the trade effects of new conflict shocks. We interact the decomposed flow components with the pair's pre-sample stock of conflict risk (accumulated from 2009 through March 2015 using GDELT v1 data), z-scored. This burn-in value is fully predetermined with respect to the estimation sample, avoiding endogeneity of the moderator. The kinetic interaction is positive and statistically significant ($+0.23$, $p = 0.04$): pairs with a longer history of hostility are *less* sensitive to new kinetic shocks. At the average level of pre-existing hostility, the kinetic effect is -6.1; each standard-deviation increase in the burn-in stock attenuates it by $+0.23$, consistent with markets having already priced in chronic hostility. Trade-context interactions remain insignificant, and the joint test of all nine interaction terms does not reject zero ($p = 0.15$). The attenuation pattern is concentrated in the kinetic channel and is modest in magnitude: even at the 95th percentile of pre-existing hostility, the kinetic effect remains strongly negative.

Column (8) tests the extensive margin. UN Comtrade records only positive trade flows, so pair-months where trade falls to zero (most starkly UKR→RUS in 2023) silently drop out of the baseline panel. We rebuild a balanced panel for the active pair set (31,901 directed pairs with at least one positive trade month in 2015–2023), expand it to all months, set missing trade to zero, and re-estimate the preferred specification. The sample grows from 1.5 million to 2.7 million observations. Kinetic strengthens from -4.54 (column 1) to -4.85, confirming that the baseline specification is conservative: conditioning on the intensive margin understates the effect of kinetic escalation, because the most extreme cases of trade collapse leave the sample entirely.

We also explored whether the choice between trade levels and bilateral shares as the dependent variable affects the results. Appendix F shows that this choice is better understood as an intensity-weighting choice under heterogeneous pair-level responses: the two estimators identify different weighted averages of the same β_{ij}, so the gap between them is a diagnostic for heterogeneity rather than a specification problem. On the aggregate monthly panel of Table 2, the share specification delivers a coefficient of -0.86 on the calibrated indicator, a shares-to-levels ratio of 0.51 that is consistent with concentrated dollar volume driving the trade-weighted average more strongly than the share-weighted one (Appendix F.2). The same pattern carries over to HS4-annual product-level resolution using the fixed-effect structure of Chevalier et al. (2026): the calibrated indicator remains large and highly significant at -1.96, the four-layer decomposition replicates, and the shares-to-levels ratio is 0.60 (Appendix F.3).

5.2 Horse Race Against Alternative Indicators

Table 7 compares our indicator with IntenSE (Chevalier et al., 2026), the closest existing bilateral indicator. IntenSE also uses GDELT as its source but takes a different route: it classifies news articles by tone and computes, for each pair-month and each category, the share of articles falling in that category among all articles mentioning the pair. We use IntenSE's aggregate "Negative" category in columns (1)–(2) and (4), and its three negative sub-categories ("Violence", "DeclarationNegative", "ActionNegative") in columns (3) and (5). The metric is thus weighted by article counts, rather than by Goldstein event severity or NumMentions salience, and is symmetric by construction (hostility from i toward j equals hostility from j toward i), which rules out any retaliation channel by

Table 7: Horse race: alternative bilateral indicators

	(1) IntenSE aggregate	(2) Both aggregates	(3) IntenSE decomposed	(4) Both dec.+IntenSE agg.	(5) Both dec.+IntenSE dec.	(6) Ours dec. + UNGA
Panel A: Our indicator						
Calibrated hostility cumulated		−1.668***				
		(0.279)				
Kinetic cumulated				−4.553***	−4.517***	−4.828***
				(0.615)	(0.617)	(0.640)
Trade_{ij} cumulated				−1.408***	−1.408***	−1.324***
				(0.252)	(0.252)	(0.191)
Trade_{ji} cumulated				−0.529***	−0.528***	−0.579***
				(0.130)	(0.130)	(0.140)
Panel B: Alternative controls						
IntenSE cumulated	−0.000	0.011		0.018		
	(0.015)	(0.015)		(0.014)		
IntenSE Violence cumulated			−0.040*		0.004	
			(0.023)		(0.023)	
IntenSE DeclNeg cumulated			0.017		0.026	
			(0.017)		(0.016)	
IntenSE ActionNeg cumulated			−0.017		0.009	
			(0.021)		(0.020)	
UNGA agreement $_{t-12}$						0.817***
						(0.174)
IntenSE joint p		[0.191]		[0.176]	[0.380]	
Observations	1,525,222	1,525,222	1,525,222	1,525,222	1,525,222	1,495,865

Notes: PPML-HDFE. Dependent variable: bilateral monthly trade. All specifications include pair×calendar-month, exporter×year-month, and importer×year-month fixed effects. Standard errors clustered by pair in parentheses. Columns (1)–(2): aggregate indicators. Column (3): IntenSE decomposed into Violence, DeclNeg, ActionNeg. Columns (4)–(5): our decomposition vs IntenSE (aggregate or decomposed). Column (6): our decomposition with 12-month lagged UNGA voting agreement score (Bailey et al., 2017). IntenSE: Chevalier et al. (2026). Sample: 2015–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

design.

Entered alone (column 1), IntenSE has a small, statistically insignificant coefficient in our gravity specification. Adding our aggregate alongside IntenSE (column 2) yields the same pattern: our indicator is negative and significant (-1.67, $p < 0.01$), while IntenSE remains small and jointly insignificant ($p = 0.19$). Column (3) decomposes IntenSE into its three negative sub-categories (Violence, Declarations, and Actions) entered alone: none is individually significant at the 5% level. Columns (4)–(5) repeat the comparison with our decomposed specification. Our kinetic, trade_{ij}, and trade_{ji} components remain statistically significant ($p < 0.01$) whether alongside aggregate IntenSE (column 4) or decomposed IntenSE (column 5); IntenSE remains jointly insignificant in both cases ($p = 0.18$ and $p = 0.38$, respectively).

Both indicators draw on the same underlying GDELT data and decompose into sub-categories, yet differ in four design choices: (i) Goldstein-weighted event severity with NumMentions salience weighting, versus article-tone classification with article-count weighting; (ii) supervised calibration against human-curated ICEWS ground truth, versus raw tone ratios; (iii) directed (asymmetric) versus symmetric bilateral structure; and (iv) sanctions-context reclassification using external administrative data, versus a purely tone-based taxonomy. These choices produce indicators that are virtually uncorrelated within pairs, and help explain why our measure captures conflict-trade variation that the tone-based alternative does not (Figure A4). At the domain level, decomposing our calibrated indicator reveals that trade (58.6×) and kinetic (13.9×) components discriminate far more sharply between adversaries and allies than the aggregate (4.2×), confirming that the decomposition isolates substantively distinct conflict dimensions (Figure A5).[14]

Column (6) adds another control: the UNGA voting agreement score of Bailey et al. (2017), lagged one year. UNGA enters positively and statistically significantly (+0.82, $p < 0.01$); with a within-sample standard deviation of 0.15, this implies that a one-standard-deviation increase in bilateral political alignment is associated with roughly 13% higher trade, consistent with political alignment facilitating trade. Our three decomposed components retain their magnitudes: the conflict-trade relationship is not driven by underlying political alignment.[15]

5.3 Limitations

Two limitations remain. The ratio structure measures hostility *shares*, not volumes: a pair with a single hostile event scores identically to one with thousands. Monthly averaging over calendar days and exponential smoothing reduce the influence of sparse event days, and calibration corrects systematic patterns in low-activity pairs. But for pairs at the extreme tails of the event distribution, the ratio remains noisy. More fundamentally, GDELT processes media, not reality. The ratio mitigates volume bias, and calibration corrects systematic selection patterns, but residual bias from disproportionate media at-

[14] The horse race is reproduced at HS4-annual product-level resolution in Appendices F.3 and F.4 using the fixed-effect structure of Chevalier et al. (2026), with the same qualitative result: our calibrated indicator remains large and highly significant and the IntenSE coefficient is not statistically distinguishable from zero in joint specification.

[15] As another external validation, the calibrated hostility stock correlates strongly with the hand-coded Tsinghua Political Relations Index (Saadaoui, 2026) for China's twelve main bilateral partners: eight of twelve pairs exceed $r = 0.70$, with CHN–USA at $r = 0.98$ (Appendix B, Figure A6).

tention to major-power rivalries cannot be fully eliminated without human-curated ground truth.

Post-2023, GSDB and GTA data are frozen. The decomposition into L^{trd} and L^{base} is unavailable for 2024–2026; total risk and kinetic/posture components remain valid. Large language models trained on the source text of each event could in principle infer the political or economic context that CAMEO discards, producing a richer classification than the taxonomy alone permits. This is a natural avenue for future work.

6 Conclusion

This paper introduces a bilateral, directed, real-time indicator of diplomatic risk that decomposes hostility into four layers: kinetic fighting ("military war"), military posture, trade-context tensions ("trade war"), and baseline diplomacy.

The decomposition is informative on two fronts. As a descriptive tool, it tracks the defining geopolitical shifts of the past decade: the substitution of economic coercion for military confrontation, the militarisation of the Russia–Ukraine relationship well before 2022, and the overwhelmingly economic nature of the US–China confrontation. As an analytical tool in a gravity equation, the aggregate indicator is negative, large, and statistically significant, but the decomposition reveals that only two of the four layers drive the result: "military wars" and "trade wars." Routine diplomacy, despite dominating measured hostility, has a trade effect of zero. Military posture provided early warning of kinetic escalation in the two major cases in the panel (Russia–Ukraine, Israel–Palestine) but is too rare to yield a separately identifiable trade effect.

The most promising extension is to move beyond the CAMEO taxonomy altogether. CAMEO classifies events by type but discards their context: a "reduce relations" event is coded identically whether it concerns a trade dispute or a territorial claim. Large language models applied to the source text of each event could recover that context, producing a richer domain classification than event codes alone permit. This would sharpen the decomposition precisely where it is weakest, at the boundary between diplomatic friction and economic coercion, and reduce the dependence on external sanctions databases that freeze after 2023. Better event classification would also allow the decomposition to capture shifts in the global geopolitical regime directly from the data. Sectoral disaggregation, testing whether conflict composition affects strategic goods differently from commodities, is a second natural direction.

References

Bailey, M.A., Strezhnev, A., and Voeten, E. (2017). Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution*, 61(2):430–456.

Baker, S.R., Bloom, N., and Davis, S.J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636.

Breinlich, H., Novy, D., and Santos Silva, J.M.C. (2024). Trade, gravity, and aggregation. *Review of Economics and Statistics*, 106(5):1418–1426.

Bazzi, S., Blair, R.A., Blattman, C., Dube, O., Gudgeon, M., and Peck, R. (2022). The promise and pitfalls of conflict prediction: evidence from Colombia and Indonesia. *Review of Economics and Statistics*, 104(4):764–779.

Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., and Ward, M. (2015). ICEWS Coded Event Data. Harvard Dataverse, https://doi.org/10.7910/DVN/28075.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, 112(4):1194–1225.

Chevalier, N., Crozet, M., Emlinger, C., and Mirza, D. (2026). Trade under tensions: insights from media-reported bilateral events. *CEPII Working Paper* 2026-02.

Crozet, M. and Hinz, J. (2020). Friendly fire: the trade impact of the Russia sanctions and counter-sanctions. *Economic Policy*, 35(101):97–146.

Evenett, S.J. and Fritz, J. (2022). *The Global Trade Alert database handbook*. Manuscript, Global Trade Alert.

Fearon, J.D. (1994). Domestic political audiences and the escalation of international disputes. *American Political Science Review*, 88(3):577–592.

Felbermayr, G., Kirilakha, A., Syropoulos, C., Yalcin, E., and Yotov, Y.V. (2020). The Global Sanctions Data Base. *European Economic Review*, 129:103561.

Fuchs, A. and Klann, N.H. (2013). Paying a visit: the Dalai Lama effect on international trade. *Journal of International Economics*, 91(1):164–177.

Gaulier, G. and Zignago, S. (2010). BACI: International trade database at the product-level. CEPII Working Paper No. 2010–23.

Glick, R. and Taylor, A.M. (2010). Collateral damage: trade disruption and the economic impact of war. *Review of Economics and Statistics*, 92(1):102–127.

Goldstone, J.A., Bates, R.H., Epstein, D.L., Gurr, T.R., Lustik, M.B., Marshall, M.G., Ulfelder, J., and Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science*, 54(1):190–208.

Growth Lab at Harvard University (2024). International trade data (HS, 92). https://atlas.cid.harvard.edu.

Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14.

Heilmann, K. (2016). Does political conflict hurt trade? Evidence from consumer boycotts. *Journal of International Economics*, 99:179–191.

Hufbauer, G.C., Schott, J.J., Elliott, K.A., and Oegg, B. (2007). *Economic Sanctions Reconsidered*. 3rd edition, Peterson Institute for International Economics.

Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632.

Leetaru, K. and Schrodt, P.A. (2013). GDELT: Global Data on Events, Location, and Tone, 1979–2012. *ISA Annual Convention*.

Martin, P., Mayer, T., and Thoenig, M. (2008). Make trade not war? *Review of Economic Studies*, 75(3):865–900.

Michaels, G. and Zhi, X. (2010). Freedom fries. *American Economic Journal: Applied Economics*, 2(3):256–281.

Mueller, H. and Rauh, C. (2018). Reading between the lines: prediction of political violence using newspaper text. *American Political Science Review*, 112(2):358–375.

Petrova, M. and Tapsoba, A. (2025). Information and conflict: from the role of (social) media and public opinion to big data and forecasting. *Economic Policy*, 40(124):845–878.

Santos Silva, J.M.C. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4):641–658.

Saadaoui, J. (2026). Geopolitical turning points and macroeconomic volatility: a bilateral identification strategy. *Journal of Comparative Economics*, forthcoming.

Schrodt, P.A. (2012). CAMEO: Conflict and Mediation Event Observations event and actor codebook. Pennsylvania State University, Event Data Project.

Tyazhelnikov, V., Zhou, X., and Shi, X. (2024). PPML, gravity, and heterogeneous trade elasticities. Working paper, University of Sydney and University of Hong Kong.

Ward, M.D., Greenhill, B.D., and Bakke, K.M. (2010). The perils of policy by p-value: predicting civil conflicts. *Journal of Peace Research*, 47(4):363–375.

A Additional Figures

Figure A1 shows domestic conflict risk for the eight countries with the highest mean domestic risk over 2015–2025. The domestic indicator uses the same daily risk formula and exponential smoothing as the bilateral version, applied to within-country events (source and target in the same country). It is calibrated with a *separate* Ridge model trained on domestic ICEWS observations (see Appendix I). Israel, Syria, and Afghanistan have the largest kinetic components. Ukraine shows a sharp increase after February 2022. Palestine (PSE) enters the top eight, driven by the Gaza escalation in late 2023.

Figure A2 examines bloc structure. We classify countries into three blocs using the alignment gap $\Delta_c = \rho(c, \text{USA}) - \rho(c, \text{CHN})$ defined in Section 3.4: USA-aligned ($\Delta_c > 0.10$), China-aligned ($\Delta_c < -0.10$), and non-aligned (in between). The ± 0.10 threshold balances bloc sizes: at ± 0.05 the non-aligned group is too small (40 countries in 2022), forcing marginal cases into polar blocs; at ± 0.15 it dominates (117 countries), diluting the distinction. The between/within hostility ratio is robust across the $[0.05, 0.15]$ range. For countries whose bilateral hostility toward the USA or China exceeds the maximum within-NATO hostility, the corresponding profile similarity ρ is set to zero before computing Δ_c, preventing adversaries from being classified as aligned with the pole they are hostile toward.

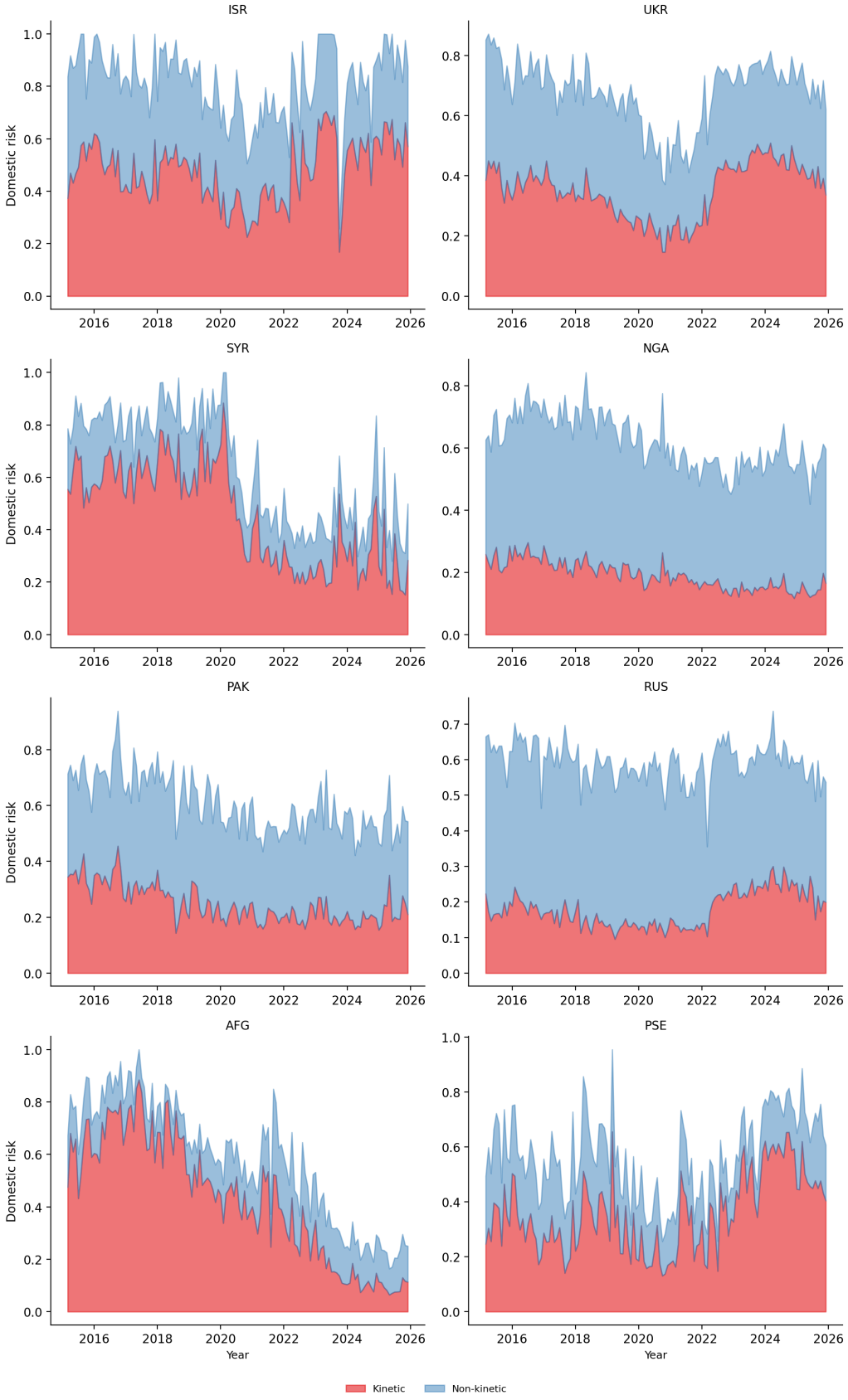

Figure A1: Domestic conflict risk: top 8 countries by mean risk (2015–2025). Red: kinetic. Blue: non-kinetic. Log-Ridge calibration.

Panel (a) plots the ratio of between-bloc to within-bloc mean hostility during event months; a ratio above 1.0 indicates that countries are more hostile toward opposing blocs than toward their own. The ratio hovers near 1.0, consistent with mild fragmentation rather than sharp bipolar polarisation. Panel (b) reports HITS hub scores from the directed conflict network, identifying the most active senders of hostility.[16] The USA and Russia dominate; Israel's hub score rises sharply after 2023.

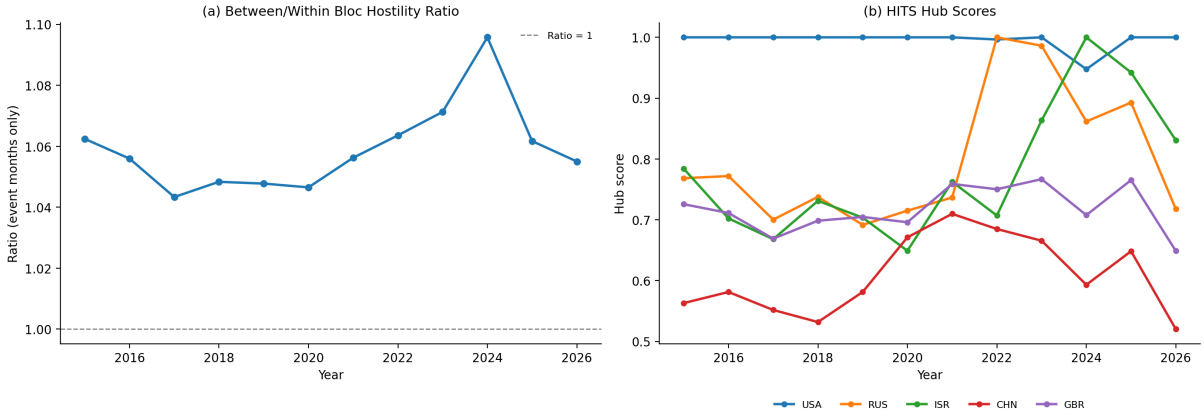

Figure A2: (a) Between/within bloc hostility ratio (event months). (b) HITS hub scores.

Figure A3 tracks EU internal cohesion (mean pairwise hostility profile similarity within the EU27) and EU–USA profile similarity (mean alignment of EU member hostility profiles with the US profile). Internal cohesion is high and stable. EU–USA alignment increases after 2022, consistent with coordinated sanctions policy against Russia converging EU threat perceptions toward US ones. Panel (b) shows that bilateral hostility *within* the EU27 averages 0.036 across 2015–2025 (event months), consistently below EU→external hostility (mean 0.041), confirming that EU members share a more cooperative baseline with each other than with the rest of the world.

Figure A4 compares our indicator with IntenSE (Chevalier et al., 2026). Panel (a) shows adversary/ally separation: our calibrated indicator achieves 4.2× discrimination versus 1.2× for IntenSE. Panel (b) reports within-pair correlations: across the full panel (including months where both indicators are silent or intermittent), the two indicators are virtually uncorrelated ($r \approx 0$), despite using the same underlying GDELT data. Panels (c–d) compare time series for selected pairs.[17]

[16] HITS (Hyperlink-Induced Topic Search) is the algorithm of Kleinberg (1999). It assigns each node in a directed graph two scores, computed jointly by power iteration. Country i's *hub* score is proportional to the sum of *authority* scores of the countries it targets, while each country's authority score is proportional to the sum of hub scores of the countries targeting it. We run HITS annually on the directed conflict graph whose edges $i \to j$ are weighted by the annual mean calibrated hostility $\langle r^{\text{cal}}_{ij,t} \rangle$. Intuitively, a country has a high hub score when it directs hostility toward countries that are themselves central targets of hostility from other active senders. For example, the USA has a high hub score because it directs hostility at China, Russia, and Iran, which are themselves among the most targeted countries in the network. A country that distributes the same total volume of hostility across many small, peripheral targets would receive a much lower hub score.

[17] The near-zero median within-pair correlation reflects the fact that most pair-months have low or zero values in at least one of the two indicators, which injects noise into the pair-level correlation. Restricting to pair-months where both indicators are strictly positive (i.e. active periods in both measures) raises the median within-pair correlation to +0.16, consistent with the positive co-movement visible in the case-study panels (c)–(d) and in major-power pairs (USA–CHN: $r = 0.50$; ISR–PSE: $r = 0.46$; RUS–UKR: $r = 0.37$).

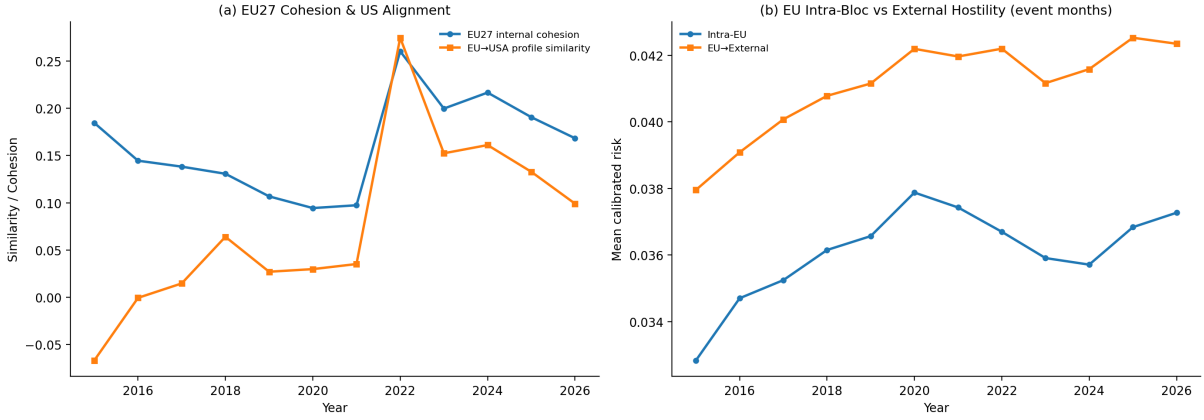

Figure A3: EU convergence. (a) EU27 internal cohesion (mean pairwise profile similarity) and EU→USA profile similarity over time. (b) Mean calibrated hostility within EU27 pairs vs from EU27 senders to external partners (event months only).

Adversary: USA→CHN, USA→RUS, USA→IRN, ISR→PSE, RUS→UKR. Allied: USA→GBR, USA→CAN, FRA→DEU, USA→JPN, DEU→NLD.

Figure A4: Comparison with IntenSE. (a) Adversary/ally separation. (b) Within-pair correlation. (c,d) Time series for selected pairs.

Figure A5 decomposes the adversary/ally discrimination by domain. The calibrated trade component achieves the highest discrimination (58.6×), followed by kinetic (13.9×) and posture (7.4×). The baseline component, which captures residual tensions not classified into the other three domains, discriminates least (2.3×).

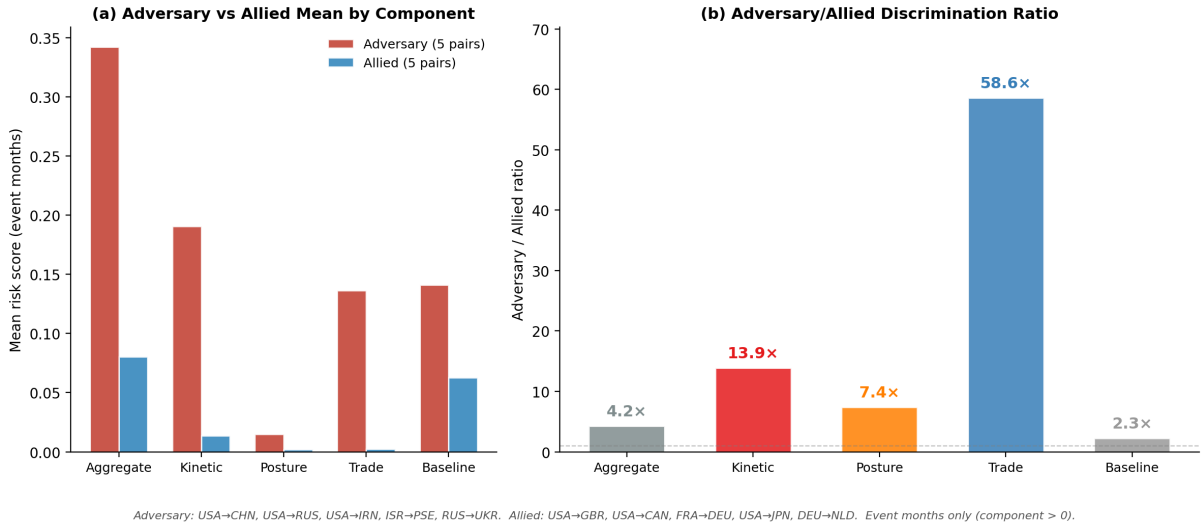

Figure A5: Domain-level adversary/ally discrimination ratios for the calibrated indicator. Five canonical adversary pairs (USA→CHN, USA→RUS, USA→IRN, ISR→PSE, RUS→UKR) versus five allied pairs (USA→GBR, USA→CAN, FRA→DEU, USA→JPN, DEU→NLD), event months only.

B Comparison with Tsinghua Political Relations Index

Figure A6 compares our calibrated hostility stock ($\bar{r}^{\text{cal}}_{ij,t}$, EWMA $\lambda = 0.95$, ∼14-month half-life) with the Tsinghua Political Relations Index (PRI), a hand-coded monthly indicator of bilateral political relations maintained by Tsinghua University for 12 China-centred dyads from 1950 to 2025. The PRI scores each relationship on a $[-10, +10]$ scale, where negative values denote hostility. Because the PRI is undirected, we average our indicator across both directions (CHN→X and X→CHN). Both series are z-scored within each pair to allow visual comparison on a common axis.

Ten of twelve pairs exhibit positive correlations between the calibrated stock and −PRI (i.e., our indicator rises when PRI deteriorates). Eight pairs exceed $r = 0.70$, with CHN–USA at $r = 0.98$, CHN–India at $r = 0.94$, CHN–UK at $r = 0.93$, and CHN–Australia at $r = 0.86$. CHN–France ($r = 0.49$) is weakest among the positive cases. Note that the PRI updates discretely: CHN–Pakistan has only two distinct PRI values across 132 months; CHN–Russia has four. Our stock variable, which accumulates monthly event flows, naturally matches the PRI's regime-characterisation construct.

Two pairs show negative correlations. CHN–Russia ($r = -0.89$): the PRI codes the Xi–Putin partnership as a plateau of maximum cooperation since 2018, while our indicator rises because GDELT records a large volume of China–Russia events post-2022, some with mildly negative Goldstein scores despite a politically cooperative relationship. This

is a construct divergence: the PRI measures political alignment, our indicator measures event intensity. For the gravity application, where the relevant variation is the intensity of concurrent hostile actions rather than underlying alignment, our construct may be the appropriate one. CHN–Pakistan ($r = -0.58$): the PRI is essentially constant (two values in 132 months), so no meaningful time-series comparison is possible.

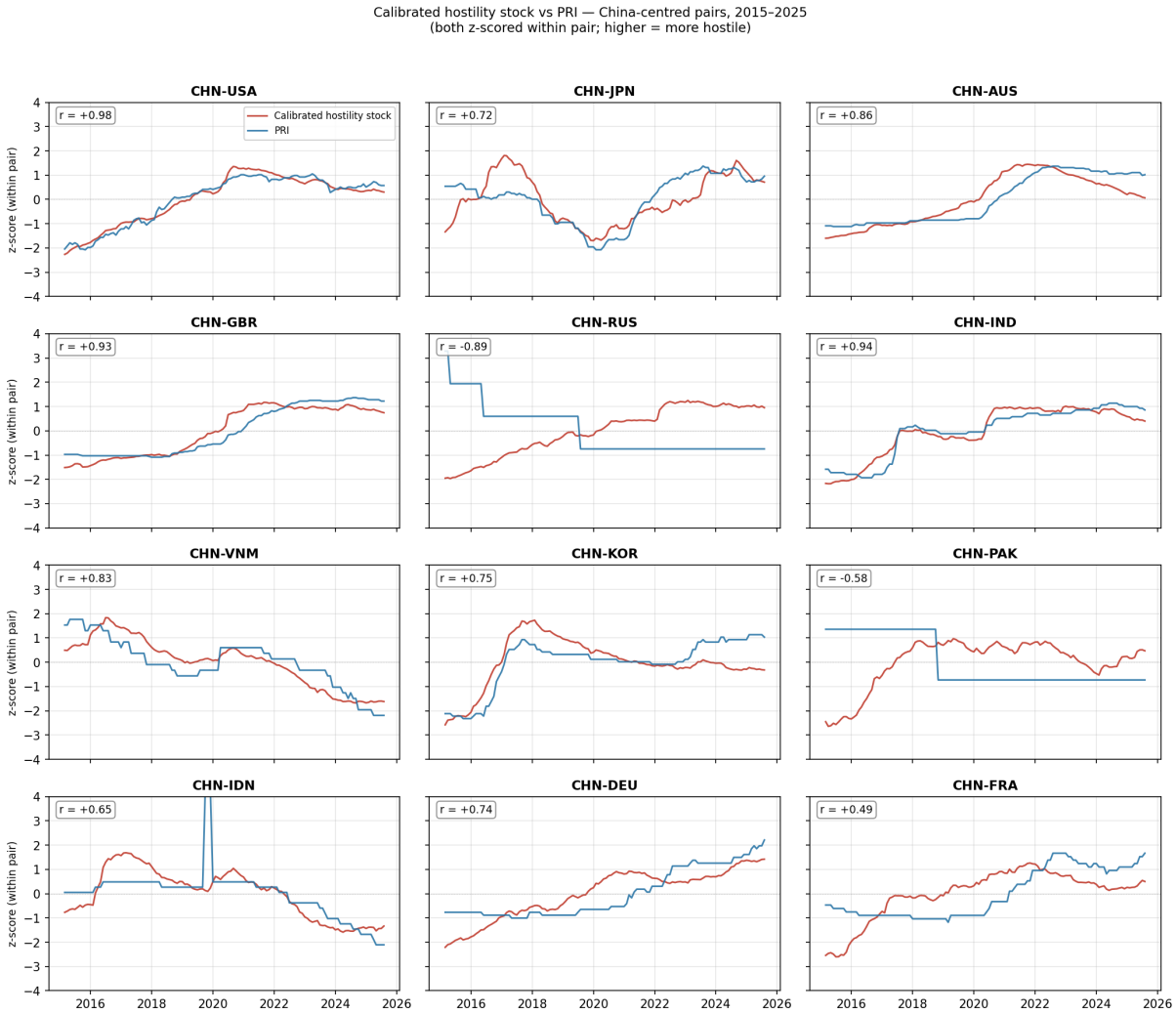

Figure A6: Calibrated hostility stock versus PRI for 12 China-centred pairs, 2015–2025. Both series z-scored within each pair; higher values denote more hostility. The calibrated hostility stock (EWMA $\lambda = 0.95$, ∼14-month half-life) is averaged over CHN→X and X→CHN directions. Pearson r annotated per panel.

C NATO-Based Adversary/Ally Classification

The main text uses five canonical adversary and five ally pairs for face validity. Here we replace this with a data-driven classification based on intra-NATO hostility. The adversary threshold equals the maximum mean bilateral hostility within NATO, excluding Turkey ($\bar{r} = 0.134$, observed for GBR→USA). Turkey is excluded due to anomalous alliance behaviour (S-400 purchase, Cyprus disputes). This yields 764 allied pairs (intra-NATO) and 23 adversary pairs (exceeding the NATO-max threshold). Figure A7 confirms

that the decomposition advantage holds under this more demanding classification: kinetic discrimination reaches 20.9× and trade 12.4×, compared with 6.2× for the aggregate.

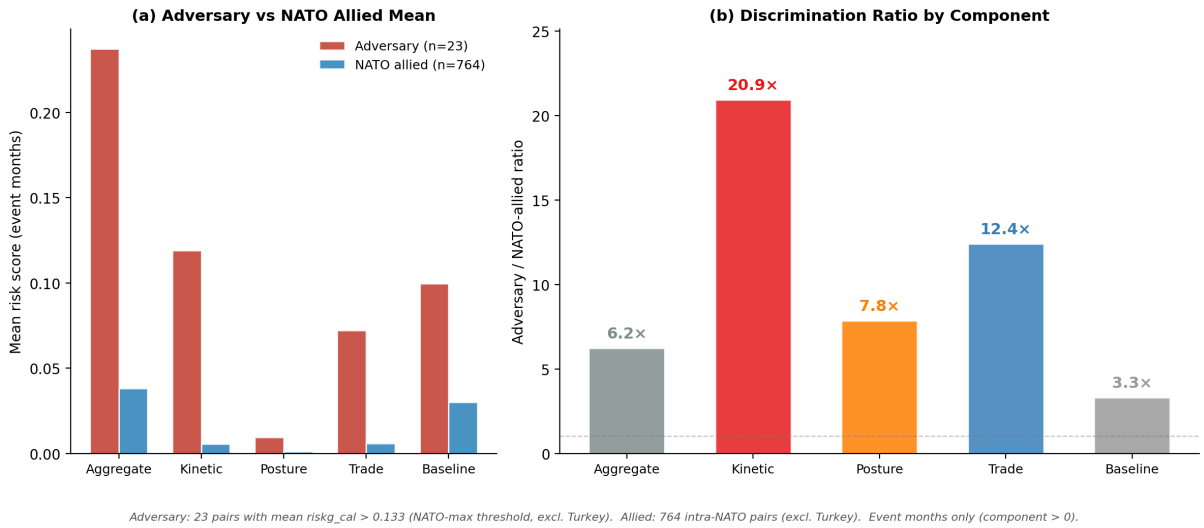

Figure A7: NATO-based adversary/ally classification. Aggregate discrimination: 6.2×; kinetic: 20.9×; trade: 12.4×.

D Maritime Route Exposure

We compute maritime route risk as the sum of bilateral conflict risk across chokepoints along each pair's shipping route. Figure A8 shows the resulting time series for selected pairs. Route risk is statistically insignificant in the gravity model conditional on direct bilateral conflict (Table 4, column 3): direct hostility, not transit risk, drives trade effects. This null result likely reflects the exporter×year-month and importer×year-month fixed effects, which absorb any chokepoint disruption that affects all pairs using a given route in a given month. The route risk variable captures only *differential* exposure across pairs sharing the same origin or destination, a narrow margin of identification.

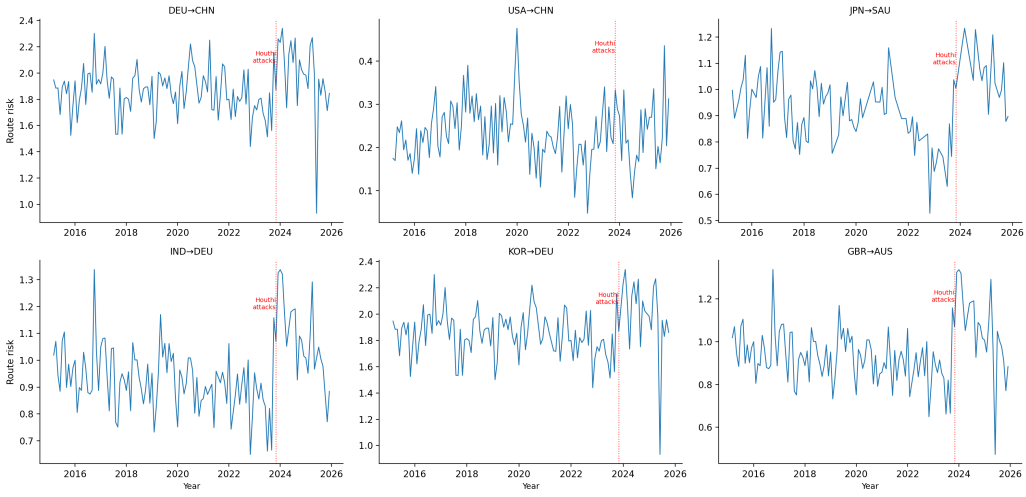

Figure A8: Maritime route risk for selected pairs. Route risk sums bilateral conflict at chokepoints (Suez, Hormuz, Malacca, etc.) along each pair's shipping lane.

E Raw (Uncalibrated) Decomposition

Figure A9 replicates the case study decomposition (Figure 2) using the raw GDELT indicator before calibration. Temporal trends and kinetic spike timing are preserved: RUS→UKR escalates post-2022, USA→CHN rises post-2018, ARM→AZE spikes during the 2020 Nagorno-Karabakh war. But levels are systematically inflated (FRA→DEU peaks at 0.4–0.5 versus 0.05 after calibration) and kinetic shares appear implausibly high, at odds with observed events. Calibration compresses both the overall level and the kinetic share, revealing the distinct compositions that the gravity estimation exploits.

The regression counterpart confirms this. Replacing the calibrated decomposition with the four raw GDELT domains (kinetic, posture, trade, political/diplomatic) in the gravity equation yields cumulated effects of −0.48 (kinetic), −1.58 (posture), −0.96 (trade), and −0.27 (political/diplomatic), all statistically significant at conventional levels. Two features stand out. First, the raw kinetic coefficient (−0.48) is an order of magnitude smaller than the calibrated one (−4.19), because the raw indicator dilutes the kinetic signal through misclassification. Second, the political/diplomatic component, which is zero in the calibrated specification, is statistically significant here ($p = 0.025$). This is expected: without the sanctions reclassification, diplomatic events during sanctions episodes remain in the poldip domain rather than being reassigned to trade, so the raw poldip component captures trade-relevant variation that should belong in L^{trd}. On a standardised basis (cumulated coefficient × within-sample standard deviation), the calibrated kinetic effect ($-4.19 \times 0.007 = -0.030$) is 58% larger than the raw counterpart ($-0.48 \times 0.040 = -0.019$); the calibrated trade effect ($-1.56 \times 0.011 = -0.017$) is 55% larger than the raw ($-0.96 \times 0.011 = -0.011$). Calibration concentrates explanatory power in the two economically meaningful domains: the standardised kinetic-plus-trade effect is −0.047 (calibrated) versus −0.030 (raw). Meanwhile, the raw poldip component captures a non-trivial standardised effect (−0.012) that vanishes after sanctions reclassification, confirming that it reflects trade-relevant variation misallocated to the wrong domain. The exercise illustrates the two distinct contributions of the pipeline: calibration compresses overcoded diplomatic noise, and the sanctions reclassification reassigns it to the correct domain.

Table A1: Raw vs. calibrated decomposition: gravity estimates

	Calibrated			Raw (uncalibrated)		
Component	$\hat{\beta}^{\text{cum}}$	SD	Std. effect	$\hat{\beta}^{\text{cum}}$	SD	Std. effect
Kinetic	−4.19***	0.007	−0.030	−0.48***	0.040	−0.019
Posture	−4.08	0.001	—	−1.58***	0.005	−0.008
Trade	−1.56***	0.011	−0.017	−0.96***	0.011	−0.011
Baseline/Poldip	−0.02	0.014	—	−0.27**	0.043	−0.012
Kinetic + Trade			−0.047			−0.030

Notes: Cumulated effects over 7 months. SD is the within-sample standard deviation of the lag-1 variable. Standardised effect = cumulated coefficient times SD. Both specifications include pair-by-calendar-month, exporter-by-year-month, and importer-by-year-month fixed effects, with pair-clustered standard errors. Calibrated column from Table 3; raw column uses the four GDELT domains without calibration or sanctions reclassification. ** $p < 0.05$, *** $p < 0.01$.

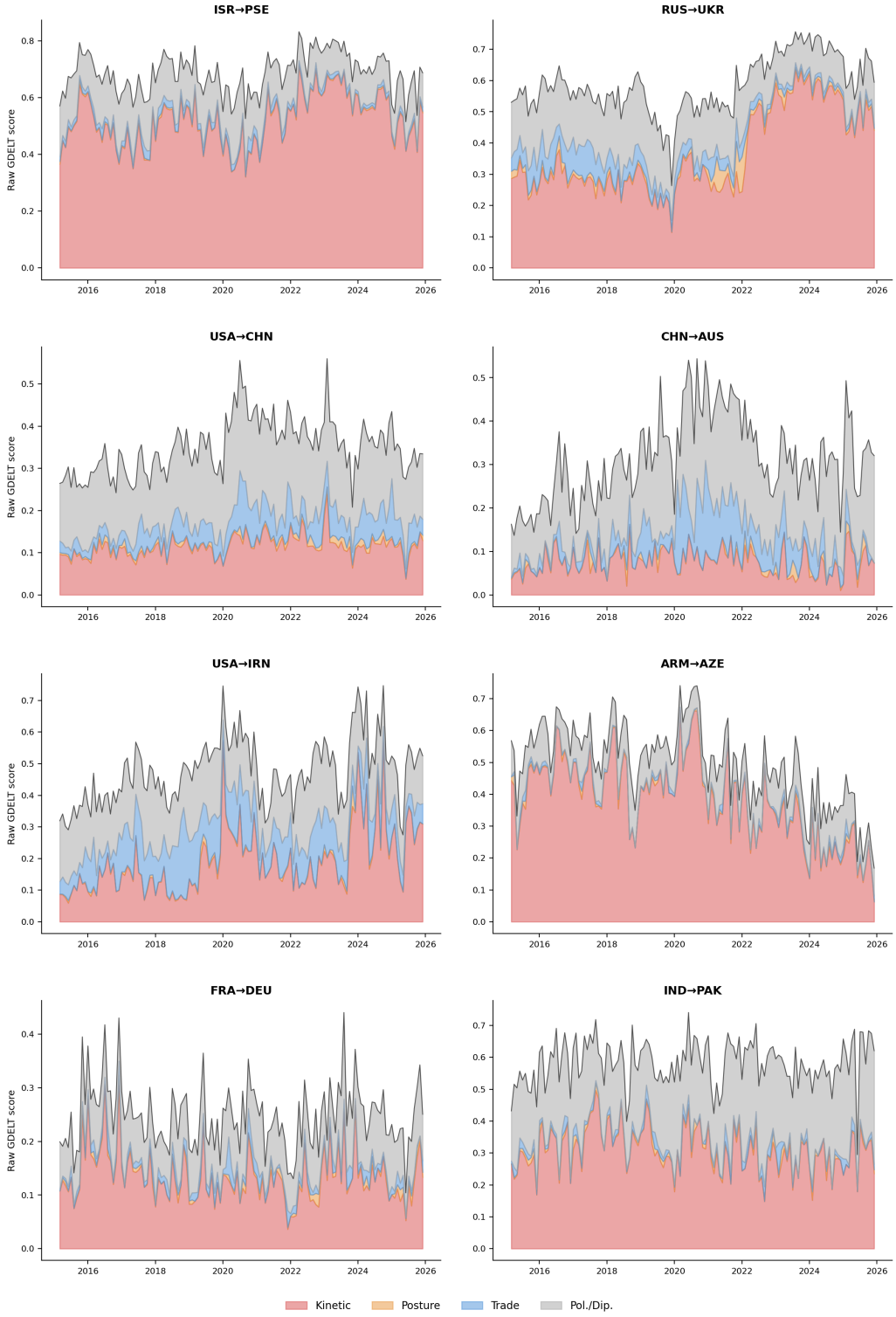

Figure A9: Raw (uncalibrated) GDELT decomposition: eight case studies. Same pairs as Figure 2. Stacked area: kinetic (red), posture (orange), trade (blue), political/diplomatic (grey). Compare with calibrated Figure 2.

F Aggregate and product-level PPML under intensity weighting

The choice between trade levels $x_{ij,t}$ and bilateral import shares $s_{ij,t} = x_{ij,t}/M_{j,t}$ as the dependent variable in a PPML gravity specification looks like a matter of scaling. Under a constant-elasticity model with a sufficiently rich fixed-effect structure, both PPML estimators are consistent for the same coefficient, so any divergence observed in practice is informative about heterogeneity in the pair-level response. Each estimator recovers a weighted average of the pair-specific responses β_{ij}, with weights determined by the dependent variable itself: levels PPML weights pair-months by expected dollar trade, while shares PPML weights them by the importer's expected share with the exporter. The dependent variable thus chooses the object of estimation. The same logic carries over to the HS4-annual product level of Appendix F.3, with $x^k_{ij,y}$ and $s^k_{ij,y} = x^k_{ij,y}/M^k_{j,y}$ replacing $x_{ij,t}$ and $s_{ij,t}$.

This appendix develops the implication at two aggregation levels. Appendix F.1 derives the intensity-weighting interpretation formally and illustrates it with a three-pair example. Appendices F.2 and F.3 apply the comparison empirically, first to the aggregate monthly Comtrade panel of Section 4 and then to an HS4-annual replication using the fixed-effects structure of Chevalier et al. (2026). In both exercises the calibrated indicator delivers a large, highly significant coefficient, and the shares-to-levels coefficient ratio matches the heterogeneity-weighting prediction. Appendix F.4 traces the pattern across three sample conventions, including the $\geq$ 3-months coverage threshold that Chevalier et al. (2026) use in constructing IntenSE (their Appendix A), imposed here as a regression-sample restriction.

F.1 Levels versus shares under heterogeneity

Consider PPML at the country-month level, matching the main-paper specification: bilateral monthly trade $x_{ij,t}$ is regressed on a scalar hostility measure $Z_{ij,t}$ with exporter-year-month, importer-year-month, and pair-calendar-month fixed effects. Let $M_{j,t} = \sum_{i'} x_{i'j,t}$ denote importer j's total imports in month t, and $s_{ij,t} = x_{ij,t}/M_{j,t}$ the bilateral import share. Under the homogeneous-response model $\mathbb{E}[x_{ij,t} \mid Z, FE] = \exp(\beta Z_{ij,t} + FE_{ij,t})$, the implied conditional mean for the share is

$$\mathbb{E}[s_{ij,t} \mid Z, FE] = \exp\big(\beta Z_{ij,t} + FE_{ij,t} - \log M_{j,t}\big),$$

so $\log M_{j,t}$ is absorbed into the importer-year-month fixed effect of the shares specification. Both PPML moment conditions $\mathbb{E}[(y - \mu)\, Z] = 0$ are then satisfied at the same β, and the two estimators are consistent for the same coefficient. The first-order conditions are not numerically identical in finite samples (the shares score is the levels score weighted by $1/M_{j,t}$), but the population target is the same. The same argument applies at the HS4-annual product level of Appendix F.3 with the importer-product-year fixed effect absorbing $\log M^k_{j,y}$.

The situation differs when the pair-level response is heterogeneous. If the data-generating process involves pair-specific β_{ij} with dispersion across pairs, no single-β exponential mean is correctly specified and each estimator converges to a different pseudo-true parameter.

Both are weighted averages of β_{ij}, but the weights differ:

$$\operatorname{plim} \hat{\beta}^{\mathrm{L}} = \sum_{ij,t} \phi^{\mathrm{L}}_{ij,t}\, \beta_{ij}, \qquad \phi^{\mathrm{L}}_{ij,t} \propto \mathbb{E}[x_{ij,t}] \cdot \operatorname{Var}(\tilde{Z}_{ij,t}),$$
$$\operatorname{plim} \hat{\beta}^{\mathrm{S}} = \sum_{ij,t} \phi^{\mathrm{S}}_{ij,t}\, \beta_{ij}, \qquad \phi^{\mathrm{S}}_{ij,t} \propto \mathbb{E}[s_{ij,t}] \cdot \operatorname{Var}(\tilde{Z}_{ij,t}),$$

where $\tilde{Z}_{ij,t}$ is the residualised regressor after partialling out the fixed effects; the weights are normalised so that $\sum \phi_{ij,t} = 1$ in each case.[18] The intensity weight ϕ^{L} is proportional to expected dollar trade on the pair-month, placing most of the identifying mass on high-volume relationships (US and China, intra-EU core, EU and Russia). The intensity weight ϕ^{S} is proportional to the expected import share, re-anchoring the estimator on pairs where a given exporter accounts for a substantial fraction of the importer's spending, regardless of the absolute dollar flow. Following Tyazhelnikov et al. (2024), $\hat{\beta}^{\mathrm{L}}$ recovers the *elasticity of the average* (trade-weighted); $\hat{\beta}^{\mathrm{S}}$ is its share-weighted analogue over the same β_{ij}.

A numerical illustration follows. Consider three representative pairs with effects ranging from -2.0 to -0.3, as reported in Table A2. With common residualised-regressor variance, the levels-weighted average gives -1.86 (US and China absorb 86% of the dollar weight, and the coefficient reflects the US-China response). The share-weighted average gives -0.70 (Bhutan and India absorb 70% of the share weight despite contributing a thousandth of dollar flow, and the coefficient reflects the Bhutan-India response). Both estimators are valid, and they identify different objects.

Table A2: Intensity weights under levels versus shares: illustrative pairs

Pair	$\mathbb{E}[x_{ij}]$ ($B/yr)	$\mathbb{E}[s_{ij}]$	β_{ij}
US and China	500	0.22	-2.0
Germany and France	80	0.12	-1.0
Bhutan and India	0.5	0.80	-0.3

A third estimator, OLS on log trade flows restricted to positive observations, identifies a third quantity: the unweighted *average elasticity* $\mathbb{E}[\beta_{ij}]$ under the standard random-coefficient assumption that β_{ij} is independent of the regressor (Tyazhelnikov et al., 2024). Two issues complicate its interpretation. Under heteroskedastic errors, Jensen's inequality implies that OLS on logs is inconsistent, with a bias that depends on how the error variance scales with Z (Santos Silva and Tenreyro, 2006); since error variance in trade data typically depends on trade volume, the bias is a systematic distortion of the estimated coefficient. The restriction to positive flows also induces selection under heterogeneity. Both issues motivate PPML as the default for gravity estimation. The intensity-weighting taxonomy clarifies that the three estimators identify three distinct objects: levels PPML the trade-weighted average, shares PPML the share-weighted average, OLS the unweighted average.

When bilateral trade is concentrated and the pair-level response is heterogeneous, the levels and shares estimators deliver systematically different coefficients, with the ratio

[18] The $\propto$ notation suppresses the common normalisation constant. The plim formula for $\hat{\beta}^{\mathrm{L}}$ is derived formally in Tyazhelnikov et al. (2024). Breinlich et al. (2024) derive the analogue for sector-level aggregation. The shares case is the direct analogue under the same argument.

depending on the joint distribution of response size and weighting quantity. Which coefficient is substantively relevant is a question about the policy object: a dollar-weighted aggregate (such as total trade at risk) maps to the levels coefficient, a representative-pair elasticity maps to the shares coefficient.

F.2 Aggregate monthly Comtrade

The aggregate monthly bilateral Comtrade panel used in Section 4 provides a first application. We apply the main-paper specification (PPML-HDFE with exporter-year-month, importer-year-month, and pair-calendar-month fixed effects; pair-clustered standard errors; three cumulated lag blocks) alternately to trade levels and bilateral shares, and we include a parallel comparison with the IntenSE indicator of Chevalier et al. (2026) to verify whether the horse race in the main text holds for each dependent variable. The sample is identical across all six cells, so differences reflect either the horse-race structure or the intensity weighting implied by the dependent variable, not sample composition.

Table A3: Aggregate-level levels versus shares: monthly Comtrade bilateral PPML

	Levels DV $x_{ij,y,m}$	Share DV $s_{ij,y,m}$	Ratio (share / levels)
Panel A: Single-indicator specifications			
r_{cal} (cum. 7-mo), alone	-1.662^{***}	-0.850^{**}	0.511
	(0.280)	(0.388)	
IntenSE (cum. 7-mo), alone	-0.00024	$+0.0290^{**}$	—
	(0.0146)	(0.0132)	
Panel B: Joint specification			
r_{cal} (cum. 7-mo), joint	-1.668^{***}	-0.864^{**}	0.518
	(0.279)	(0.386)	
IntenSE (cum. 7-mo), joint	$+0.0111$	$+0.0319^{**}$	—
	(0.0146)	(0.0133)	
Observations (all cells)	1,525,222	1,525,222	
Pseudo R^2 (levels / shares)	0.99	0.35	

Notes: PPML-HDFE on the monthly bilateral Comtrade panel used in Section 4. Dependent variable: monthly bilateral trade (Levels) or bilateral import share $s_{ij,y,m} = x_{ij,y,m}/M_{j,y,m}$ (Share). All specifications include exporter-year-month, importer-year-month, and pair-calendar-month fixed effects. Standard errors clustered by directed pair in parentheses. Coefficients are cumulated sums over three lag blocks (L_1, $\bar{L}_{2\text{–}4}$, $\bar{L}_{5\text{–}7}$), identical to the main-paper horse-race specification. The levels and share regressions are run on an identical sample of 1,525,222 pair-month observations after singleton absorption. The ratio row is omitted for IntenSE because the levels estimate is statistically indistinguishable from zero in both alone and joint specifications, making the ratio uninformative. The r_{cal} joint ratio of 0.518 is consistent with the product-level ratio of 0.60 in Table A5 and with the heterogeneity-weighting logic described in Section F.1. Sample: 2015–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3 reports the six specifications. The coefficient on the calibrated indicator is large, highly significant, and negative in both the levels and shares specifications, with little movement when IntenSE is added. The shares-to-levels ratio on the calibrated indicator is 0.51 in both the alone and joint configurations. The IntenSE coefficient is statistically indistinguishable from zero in the levels specification and small in magnitude in the shares specification.

The shares-to-levels ratio of 0.51 is below one, as the heterogeneity-weighting framework predicts when pair-specific responses correlate with trade volume: the dollar-weighted average of β_{ij} is roughly twice as negative as the share-weighted average, consistent with bilateral trade disruption from conflict operating more strongly on high-volume relationships than on dependency-heavy small-economy pairs. The main paper's trade-at-risk calculation of $334 billion uses the levels coefficient because the policy object is dollar-weighted: the counterfactual asks how much dollar trade is displaced by the observed evolution of bilateral hostility, not how a representative pair-product's share responds.

F.3 HS4-annual product-level replication

The aggregate pattern extends to disaggregated data. We replicate the analysis at HS4-annual resolution using the fixed-effects structure of Chevalier et al. (2026): exporter-product-year, importer-product-year, and pair-product fixed effects, with pair-clustered standard errors.[19] The sample is BACI HS4 annual for 2016–2023 restricted to the 98 countries appearing in the top 100 BACI traders, our risk panel, and IntenSE coverage; Hong Kong is excluded because its bilateral events are coded at the Chinese sovereign level, and Bosnia-Herzegovina because it falls outside our calibration. The 98-country subset captures 95.94% of bilateral trade within the top 100, close to the 100-country panel of Chevalier et al. (2026). After singleton absorption, 35.66 million pair-product-year cells enter the regression, identical across all reported specifications.

Table A4 reports that the main-paper coefficient patterns reappear at product-level resolution. The coefficient on the calibrated indicator is again large and highly significant, with little movement when IntenSE is added as a second regressor. The four-layer decomposition replicates the aggregate pattern: the kinetic and trade-context layers drive the response, while posture and baseline layers are not statistically distinguishable from zero. Substituting bilateral shares for trade levels in column (5) yields a smaller magnitude on the calibrated indicator, which is the intensity-weighting shift examined in the next table.

Table A5 arranges the product-level levels-versus-shares comparison in the same structure as Table A3. The two panels share an identical sample of 35.66 million observations, both converge to machine precision, and the shares-to-levels ratio on the calibrated indicator is 0.60, stable across the alone and joint specifications. The product-level ratio of 0.60 is somewhat larger than the aggregate ratio of 0.51 but exhibits the same within-sample stability across single-indicator and horse-race configurations. The stability is what Appendix F.1 predicts under heterogeneous pair-level responses: a fixed rescaling between the dollar-weighted and share-weighted averages that does not depend on whether IntenSE is also in the regression.

F.4 Robustness across sample conventions

The exercise in Tables A4 and A5 runs on the full HS4-pair-year panel: every active (i, j, k) triple with at least one positive-trade year, balanced across 2016–2023 with zero-fill on inactive years and zero-imputation of IntenSE on pair-years absent from the IntenSE *Negative* file. This zero-imputation convention is the closest analogue at our annual frequency to the main-text gravity sample of Chevalier et al. (2026). They note on

[19] Pair clustering is more conservative than the pair-quarter clustering of Chevalier et al. (2026) and allows arbitrary within-pair serial correlation across years and products.

Table A4: Product-level horse race: HS4 annual, top-100 BACI traders, 2016–2023

	(1) r_{cal} only	(2) IntenSE only	(3) Both	(4) 4-layer+IntenSE	(5) Share DV
$r_{\text{cal},t-1}$	-1.963^{***}		-1.962^{***}		-1.180^{***}
	(0.454)		(0.456)		(0.266)
IntenSE_{t-1}		-0.0064	-0.0016	-0.0036	-0.0031
		(0.0095)	(0.0096)	(0.0094)	(0.0050)
L^{kin}_{t-1}				-4.577^{***}	
				(0.713)	
L^{pos}_{t-1}				1.353	
				(3.258)	
L^{trd}_{t-1}				-1.964^{**}	
				(0.713)	
L^{bas}_{t-1}				-0.647	
				(0.464)	
DV	$x^k_{ij,y}$	$x^k_{ij,y}$	$x^k_{ij,y}$	$x^k_{ij,y}$	$s^k_{ij,y}$
Observations	35,659,617	35,659,617	35,659,617	35,659,617	35,659,617
Pseudo R^2	0.987	0.987	0.987	0.987	0.454

Notes: PPML-HDFE. Dependent variable: annual bilateral trade at the HS4 product level (columns 1–4) or bilateral import share $s^k_{ij,y} = x^k_{ij,y}/M^k_{j,y}$ (column 5). All specifications include exporter-product-year, importer-product-year, and pair-product fixed effects. Standard errors clustered by directed pair in parentheses. Sample: 98 countries at the intersection of the top-100 BACI traders, the risk panel, and IntenSE coverage. Panel balanced across 2016–2023 with zero-fill on inactive years within active (i, j, k) triples; 1,346,599 observations absorbed as singletons or separated by the fixed-effects structure. Indicators are lagged one year in annual-mean form. See Appendix F.1 for the interpretation of the levels-versus-shares specification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Levels versus shares: product-level PPML coefficients

	Levels DV $x^k_{ij,y}$	Share DV $s^k_{ij,y}$	Ratio (share / levels)
Panel A: Single-indicator specifications			
$r_{\text{cal},t-1}$, alone	-1.963^{***}	-1.182^{***}	0.60
	(0.454)	(0.267)	
IntenSE_{t-1}, alone	-0.0064	-0.00425	—
	(0.0095)	(0.0051)	
Panel B: Joint specification			
$r_{\text{cal},t-1}$, with IntenSE	-1.962^{***}	-1.180^{***}	0.60
	(0.456)	(0.266)	
IntenSE_{t-1}, with r_{cal}	-0.0016	-0.0031	—
	(0.0096)	(0.0050)	
Observations (all panels)	35,659,617	35,659,617	
Pseudo R^2	0.987	0.454	

Notes: PPML-HDFE. Dependent variable: annual bilateral trade at HS4 product level (Levels) or bilateral import share $s^k_{ij,y} = x^k_{ij,y}/M^k_{j,y}$ (Share). All specifications include exporter-product-year, importer-product-year, and pair-product fixed effects. Standard errors clustered by directed pair in parentheses. Under a homogeneous-response model the levels and shares coefficients are algebraically identical; under heterogeneous responses the ratio $\hat{\beta}^{\text{S}}/\hat{\beta}^{\text{L}}$ differs from one, with the levels estimate weighting pair-products by expected dollar trade and the shares estimate weighting by expected bilateral share (see Section F.1). The ratio of 0.60 on our calibrated indicator is stable across Panels A and B, indicating a heterogeneity-induced rescaling rather than a specification defect. Both panels share the same 35,659,617-observation sample; rows where $M^k_{j,y} = 0$ (share undefined) coincide with cells dropped for singleton/separation reasons by the levels PPML. All specifications converge in 19–20 iterations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

page 20 that no zero flows need to be created at the regression stage given the fixed-effect structure, and the arithmetic of their 52-million-observation sample implies that pair-quarters absent from the IntenSE *Negative* file enter with zero IntenSE. Our full-sample specification therefore matches their sample convention, up to the annual-versus-quarterly frequency difference.

Two alternative samples test whether this convention is material. The first restricts to pair-years where both indicators are strictly positive, so that neither carries zero mass on the observations entering identification. The second applies, on top of the full-sample merge, the $\geq$ 3-months coverage threshold that Chevalier et al. (2026) use in the construction of IntenSE itself (their Appendix A, where it "preserves 54% of observations"). This threshold determines which dyad-months appear in the IntenSE file; Chevalier et al. (2026) then zero-impute absent pair-quarters at merge in their regression, whereas we impose the threshold as a regression-sample restriction to examine whether the patterns of Tables A4 and A5 hold on the sample where IntenSE is by construction well-measured.

The $\geq$ 3-months threshold drops a meaningful share of observations. Among dyads with any IntenSE coverage in the lag year, the distribution of months-with-coverage is continuous (median 4, mean 5.2 of 12), so the threshold removes 26.3% of covered dyad-years. On our panel it retains 31.3% of pair-years unweighted but 90.1% of 2016–2023 dollar trade; the lower unweighted retention compared with the 54% reported by Chevalier et al. (2026) reflects that our denominator is the 98-country intersection rather than the IntenSE-covered subset.

Table A6 traces the coefficients across the three sample conventions. The coefficient on the calibrated indicator is close to invariant to sample selection, moving within a narrow band around -2.0 across all three samples and both alone and joint specifications. The IntenSE coefficient is not statistically distinguishable from zero in the joint specifications, with p-values of 0.87, 0.12, and 0.27 across the three samples.

Panel B reports the shares-to-levels ratio on the calibrated indicator across the three samples. The ratio rises from 0.60 on the full sample to 0.77 on the both-positive subsample and 0.79 on the $\geq$ 3-months sample; within each sample it is stable across the alone and joint specifications. The direction of movement fits the intensity-weighting framework: sample restrictions that drop zero-mass pair-years remove the most extreme observations in the weight distribution, leaving a more homogeneous support, so the levels-weighted and share-weighted averages diverge less. The ratio does not converge to one even on the most restrictive sample, indicating that some heterogeneity in pair-level responses persists among actively-covered pair-products.

G Summary Statistics

Table A7 reports descriptive statistics on the gravity estimation sample. Trade is in millions of USD; the four conflict layers and the aggregate calibrated indicator are reported in percentage points (a value of 1.00 means 1% of bilateral interaction is hostile). All conflict variables are heavily right-skewed: over two-thirds of pair-months have zero conflict, while the 99th percentile reaches several percentage points. Kinetic risk is rare but extreme; baseline diplomacy has the highest mean; trade-context hostility spans four orders of magnitude between the 25th and 75th percentiles.

Table A6: Robustness of the product-level horse race across sample conventions

	Full sample (Table A4)	Both-positive subsample	Chevalier $\geq$ 3mo filter
Panel A: Levels-DV coefficients			
$r_{\text{cal},t-1}$, alone	-1.963^{***}	-2.085^{***}	-2.098^{***}
	(0.454)	(0.488)	(0.482)
$r_{\text{cal},t-1}$, joint	-1.962^{***}	-2.057^{***}	-2.075^{***}
	(0.456)	(0.489)	(0.482)
IntenSE_{t-1}, alone	-0.0064	-0.0658^{***}	-0.0702^{**}
	(0.0095)	(0.0195)	(0.0236)
IntenSE_{t-1}, joint	-0.0016	-0.0271	-0.0232
	(0.0096)	(0.0175)	(0.0209)
IntenSE_{t-1}, joint p-value	0.87	0.12	0.27
Panel B: Heterogeneity-weighting ratio (shares / levels) on r_{cal}			
Alone	0.60	0.75	0.78
Joint	0.60	0.77	0.79
Observations	35,659,617	19,840,504	16,556,058

Notes: PPML-HDFE on the HS4-pair-year panel of Section F.3. Dependent variable: annual bilateral trade at the HS4 product level. All specifications include exporter-product-year, importer-product-year, and pair-product fixed effects. Standard errors clustered by directed pair in parentheses. Indicators lagged one year, in annual-mean form. Three sample conventions: *Full sample* keeps every active (i, j, k) triple with at least one positive-trade year, balanced across 2016–2023 with zero-fill on inactive years. *Both-positive subsample* restricts to pair-years where both $r_{\text{cal},t-1} > 0$ and $\text{IntenSE}_{t-1} > 0$. *Chevalier $\geq$ 3mo filter* retains pair-years for which the lag year $(t-1)$ records IntenSE Negative-category articles in at least three distinct months; this is the $\geq$ 3-months threshold Chevalier et al. use when constructing the IntenSE dataset (their Appendix A), applied here as a regression-sample restriction rather than at the dataset-construction stage. The filter retains 31.3% of pair-years unweighted and 90.1% of 2016–2023 dollar trade. Panel B reports the ratio of the shares-DV coefficient to the levels-DV coefficient for r_{cal} on each sample (computed from the full six-cell levels-versus-shares grid run on each sample but omitted here for brevity). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Summary statistics: gravity estimation sample.

	N	Mean	SD	P25	P75	P99
Trade (million USD)	1,856,916	81.9	690.0	0.092	11.4	1,464
	N	Mean	SD	% Zero	P75	P99
$\hat{r}^{\text{cal}}$	1,856,916	1.27	2.23	68.9	2.87	6.99
L^{kin}	1,856,916	0.21	0.73	70.0	0.22	2.43
L^{pos}	1,856,916	0.03	0.12	73.3	0.01	0.39
L^{trd}_{ij}	1,856,916	0.31	1.09	69.3	0.06	4.70
L^{base}	1,856,916	0.71	1.45	77.1	0.00	4.99
L^{trd}_{ji}	1,856,916	0.30	1.09	70.4	0.05	4.68

Notes: 1,856,916 observations, 24,118 directed pairs, 196 exporters, 196 importers, 105 year-months (before lag-based sample restrictions; the preferred estimation sample in Table 4 column (2) uses 1,525,222 observations after requiring non-missing L_1, $\bar{L}_{2\text{–}4}$, and $\bar{L}_{5\text{–}7}$ lags). Conflict variables in percentage points ($\times$100); a value of 1.00 means 1% of bilateral interaction is hostile. Trade in million USD. L^{trd}_{ji}: target's trade hostility toward sender. % Zero: share of pair-months with zero conflict.

H Calibration Details

This section documents the Ridge calibration pipeline in detail.

Training Data

The training sample consists of the 2015–2019 overlap between GDELT v2 and ICEWS. ICEWS events are filtered to *Government/Military* actor sectors only, reducing noise from non-state actors. The GDELT side is unfiltered: the calibration learns which events are informative. We treat ICEWS as the ground truth for bilateral conflict intensity and use data where both sources are simultaneously active. The resulting bilateral training set contains approximately 30,500 observations across roughly 5,500 dyads. Each observation is a directed-pair–month.

Feature Construction

The calibration faces a specific challenge: GDELT is unfiltered (any news article generates events regardless of source credibility or actor type), while ICEWS is curated (events are coded by trained analysts from vetted sources, and we further restrict to Government/Military actors). The Ridge must therefore learn to extract the signal embedded in noisy GDELT aggregates. We provide 84 features designed to give the model flexibility along three dimensions: *level* (how much conflict), *composition* (what type), and *dynamics* (transitory vs. persistent).

1. **Risk-level features** (12): The Goldstein-weighted conflict ratio and its four domain components (kinetic, posture, political–diplomatic, trade), each in both flow and stock ($\lambda = 0.95$) form. These are the natural starting point: if GDELT and ICEWS agreed perfectly, the GDELT flow alone would suffice. The stock captures persistence, since pairs with chronically elevated risk may map differently to ICEWS than those with transitory spikes.

2. **Domain shares** (4): Each domain's share of total risk (military, poldip, trade) plus kinetic's share within military. These let the model reweight domains: GDELT may over-count verbal threats, so the kinetic–posture composition tells the model how much to trust the aggregate.

3. **Log transforms** (6): $\log(1+x)$ of each flow variable. The $\log(1+\cdot)$ transform maps zero to zero, handles the mass of zeros without ad hoc trimming, and compresses the right tail. The relationship between GDELT and ICEWS risk is unlikely linear: a tenfold increase in raw GDELT hostility need not imply a tenfold increase in curated ICEWS hostility. Log features let the Ridge capture this concavity within a linear model.

4. **Deviation features** (3): Flow minus stock for total, military, and poldip risk. Separating transitory spikes from the persistent trend lets the model weight each component differently.

5. **Ratio features** (2): Log military-to-political ratio and log kinetic-to-posture ratio. These capture the *composition* of hostility independent of its level, addressing the concern that GDELT's level may be unreliable but its relative domain mix may still be informative.

6. **Event-level features** (57): Constructed directly from raw GDELT event records, bypassing our risk aggregation. These include: event counts by each CAMEO root code (01–20; see Table A10), giving the Ridge direct access to the full interaction taxonomy; distributional statistics of Goldstein scores within each pair-month (mean, std, min, max, median, p10, p90, range, IQR), because the *distribution* of event severity, not just its mean, may matter; media salience proxies (mean mentions, total mentions, max mentions, mean sources); counts of hostile, cooperative, material-hostile, verbal-hostile, and kinetic events; the hostile and cooperative fractions; the material-to-verbal hostility ratio; days active in the month; CAMEO entropy (Shannon entropy over the 20 root codes, capturing diversity of interaction types); and temporal concentration (share of events in the first half of the month). These 57 features serve as a "bypass" around our hand-crafted risk formula: if the Goldstein-weighted ratio discards useful information, the event-level features let the Ridge recover it.

The feature set is deliberately *over-complete*: many features are correlated (e.g., log risk and risk level). Ridge regularisation handles this by shrinking correlated coefficients toward each other rather than selecting one arbitrarily, making the predictions stable even when features are collinear.

Pipeline

Each of the five Ridge models follows an identical pipeline:

1. **Standardisation**: All 84 features are centred and scaled to unit variance. This ensures that Ridge regularisation penalises all features equally regardless of their original scale.
2. **Missing values**: Filled with zero (a feature absent for a pair-month contributes nothing).
3. **Cross-validation**: Five-fold grouped cross-validation, grouping by dyad. All months of a given pair remain in the same fold, preventing the model from memorising pair-specific patterns and forcing generalisation across pairs.
4. **Regularisation**: Ridge regression minimises $\sum_i (y_i - x_i'\beta)^2 + \alpha\|\beta\|^2$, where $\alpha > 0$ controls the penalty on coefficient magnitude. Larger α shrinks all coefficients toward zero, reducing variance at the cost of bias. The penalty parameter α is selected from 20 values on a log-scale grid $[10^{-2}, 10^{4}]$ by minimising leave-one-group-out MSE. A small selected α (e.g., poldip share: $\alpha = 0.38$) indicates the model needs flexibility; a large α (e.g., trade share: $\alpha = 10{,}000$) indicates heavy shrinkage is optimal because few features carry signal, consistent with the trade domain being well-identified by a handful of trade-specific inputs.

Five Models: Total Level and Domain Shares

A single Ridge model predicting calibrated risk in each domain separately would not preserve the identity $r^{\text{cal}} = \sum_k r_k^{\text{cal}}$: four independent level predictions have no reason to sum to a fifth. We therefore separate the problem into a total-level model and four share models, enforcing additivity by construction.

- The **total model** predicts the ICEWS total risk level from 84 GDELT features. It achieves out-of-sample $R^2 = 0.48$ (RMSE = 0.063), selected $\alpha = 264$. This is the workhorse: it determines *how much* conflict a pair experiences.
- Four **share models** predict each domain's fraction of total risk (kinetic/total, posture/total, poldip/total, trade/total). These determine *what type* of conflict it is. Predicted shares are clipped to $[0, 1]$ and renormalised to sum to one; the last domain is computed as a residual.

Table A8 summarises the five models. The selected α values are informative. The poldip share model has relatively low α (0.38), indicating that many features carry signal and the model needs flexibility. The kinetic, posture, and trade share models have high α (1,129, 4,833, and 10,000), indicating that the prediction collapses to a handful of domain-specific features, with everything else shrunk to near zero. This is consistent with the coefficient patterns in Table A9: the trade model loads almost exclusively on trade-specific inputs.

Table A8: Five Ridge calibration models: targets and performance.

Model	Target	α	R^2	RMSE
Total	ICEWS total risk level	264	0.48	0.063
Kinetic share	Kinetic / total	1,129	0.23	—
Posture share	Posture / total	4,833	0.08	—
Poldip share	Political-diplomatic / total	0.38	0.20	—
Trade share	Trade / total	10,000	0.04	—

Notes: All metrics are out-of-sample (5-fold grouped CV by dyad). α selected from 20 values on $[10^{-2}, 10^{4}]$ by minimising leave-one-group-out MSE. The share designated as residual (here: trade) does not affect component values: because the four normalised shares sum identically to one, computing three shares directly and the fourth as $\hat{r}^{\text{cal}} - \sum_{k \neq \text{res}} \hat{s}^k \hat{r}^{\text{cal}}$ returns the same value as the direct product $\hat{s}^{\text{res}} \hat{r}^{\text{cal}}$.

The calibrated indicator is: $\hat{r}^{\text{cal}}_{ij,t} = \text{clip}(\hat{f}(X_{ij,t}), 0, 1)$. Domain components: $\hat{r}^{\text{cal}}_{k,ij,t} = \hat{s}^k_{ij,t} \times \hat{r}^{\text{cal}}_{ij,t}$.

Share model R^2. The share models exhibit lower R^2 than the total model, particularly trade ($R^2 = 0.04$) and posture ($R^2 = 0.08$). The low R^2 means that GDELT's domain composition is a genuinely noisy proxy for ICEWS's curated domain composition: despite sharing the same CAMEO taxonomy, the two datasets disagree substantially on what fraction of hostility is trade-related for a given pair-month. We do not claim the Ridge solves this problem. Two features of the pipeline make the share noise tolerable, independently of the econometric results.

First, calibration performance is uneven across domains, and the ranking maps onto the conceptual clarity of the underlying events. The kinetic share model achieves $R^2 = 0.23$, substantially above posture (0.08) and trade (0.04). This is expected: armed violence (CAMEO 18–20) generates unambiguous Goldstein scores (-8 to -10) and distinctive media signatures, making kinetic events the easiest to classify consistently across datasets. The political/diplomatic share also performs reasonably ($R^2 = 0.20$), reflecting the large volume and stable composition of diplomatic events.

The two poorly predicted domains are trade and posture, for different reasons. The trade/diplomatic boundary is hard to draw because diplomatic events (demands, accusations) can serve either political or economic purposes depending on context; this is where

the sanctions reclassification helps (see below). Posture events (CAMEO 15 sub-codes 150–155: exhibit force, military posturing) sit in an inherently ambiguous zone between political signalling, military escalation, and diplomatic pressure. A naval deployment near a disputed strait could be coded as posture, diplomacy, or even kinetic depending on framing. The low posture R^2 reflects this genuine conceptual ambiguity rather than pure coding noise. In practice, this ambiguity is not consequential for the decomposition: posture events that are misclassified as diplomatic or vice versa simply shift mass between L^{pos} and L^{base}, both of which capture non-trade, non-kinetic hostility.

Second, the pipeline does not ask the Ridge to draw the trade/diplomatic boundary on its own. That boundary is precisely where the low trade-share R^2 bites: the Ridge cannot reliably distinguish trade hostility from diplomatic hostility. But the four-component decomposition (Section 2.4) resolves this with external data. The sanctions flag $D^{\text{sanc}}_{ij,t}$, constructed from GSDB and GTA administrative records, reclassifies all diplomatic events as trade-related whenever a pair has active trade or financial sanctions: $L^{\text{trd}} = \hat{r}^{\text{cal}}_{\text{trd}} + D^{\text{sanc}} \times \hat{r}^{\text{cal}}_{\text{dip}}$ (equation 7). For sanctioned pairs, L^{trd} is driven by $D^{\text{sanc}} \times \hat{r}^{\text{cal}}_{\text{dip}}$, where D^{sanc} comes from administrative data (independent of GDELT) and $\hat{r}^{\text{cal}}_{\text{dip}}$ is built from the poldip share ($R^2 = 0.20$) and total risk ($R^2 = 0.48$). The low trade-share R^2 has limited influence on the trade component because the sanctions reclassification substitutes external information for the Ridge's weakest prediction. For unsanctioned pairs, the trade component reduces to $\hat{r}^{\text{cal}}_{\text{trd}}$ alone, which is noisier, but these pairs also tend to have low trade-specific hostility by construction.

In short, the decomposition rests on two well-predicted domains (kinetic and diplomatic) and two external corrections. The trade/diplomatic boundary, which the Ridge cannot draw reliably, is resolved by administrative sanctions data. The posture/baseline boundary, which is conceptually ambiguous, is left unresolved but is inconsequential: both components capture non-trade, non-kinetic hostility, and misallocation between them does not affect the key distinction between armed violence, economic coercion, and residual tension.

Top Features by Standardised Coefficient

Table A9 reports the five largest Ridge coefficients (in absolute value, on the standardised scale) for each model.

Total model. The strongest predictors are event-level statistics rather than our hand-crafted risk aggregates: CAMEO diversity (the number of distinct event types in a pair-month), verbal hostile event counts, and total hostile event counts all enter positively. Total media mentions enter negatively, suggesting that high-salience pairs are *over*-reported relative to their ICEWS-measured risk. Cooperative event counts also enter negatively: more cooperation signals lower conflict, conditional on the hostile features. The dominance of event-level features over risk aggregates confirms that the Ridge extracts information orthogonal to our Goldstein-weighted ratio, validating the calibration design.

Kinetic share model. CAMEO diversity enters negatively: pairs with many different event types have a *lower* kinetic share, consistent with diverse interactions reflecting a diplomatic mix rather than pure armed violence. CAMEO 19 ("Fight") enters positively, a direct mapping from the event taxonomy to the kinetic domain. The kinetic and military stock variables enter positively, capturing persistence: pairs with a history of armed

violence continue to have high kinetic shares.

Posture share model. The posture stock (persistence of past posture events) is the top predictor, reflecting that force displays tend to be sustained over time (military build-ups, naval deployments). CAMEO 15 ("Exhibit military force") enters positively, a direct mapping from the event taxonomy. CAMEO 07 ("Provide aid", often military aid) also loads positively, consistent with arms transfers signalling posture.

Political/diplomatic share model. The signs are the mirror image of the kinetic model. Log total risk is *negative*: high-conflict pairs have lower poldip shares because kinetic events dominate. The kinetic/posture ratio and log kinetic risk enter positively: more military activity relative to diplomacy increases the residual poldip share, as the Ridge adjusts for GDELT's over-counting of verbal diplomatic events.

Trade share model. All top features are trade-specific: trade risk (flow and log), the trade stock, and GDELT's own trade share all enter positively. The trade-classified sub-codes of CAMEO 16 ("Reduce relations": embargoes, aid cuts, economic downgrades) complete the picture. This model is the most self-contained: trade events in GDELT map straightforwardly to trade events in ICEWS, which explains both its simplicity and its high regularisation ($\alpha = 10{,}000$).

CAMEO Taxonomy

The Conflict and Mediation Event Observations (CAMEO) taxonomy (Schrodt, 2012) classifies political interactions into 20 root categories, shown in Table A10. Both GDELT and ICEWS code events using this taxonomy. The Goldstein score assigns each event type a cooperation–conflict intensity on a $[-10, +10]$ scale. Our conflict ratio uses events with negative Goldstein scores (categories 10–20) as conflictual; non-negative scores (categories 01–09) are treated as cooperative or neutral and enter only the denominator.

I Domestic Conflict Risk Calibration

The impact of the domestic conflict indicator on international trade cannot be directly estimated because all country-specific effects are absorbed by exporter×year-month and importer×year-month fixed effects. The domestic indicator therefore serves as an input to the panel and to the descriptive analysis (Figure A1).

Domestic risk is defined by events where the source and target actors belong to the same country. The daily risk formula, monthly averaging, and exponential smoothing ($\lambda = 0.95$) are identical to the bilateral version (equations 1–3). Calibration uses a *separate* Ridge model trained exclusively on domestic ICEWS observations, with the same 84-feature structure and five-fold grouped cross-validation by country. The bilateral and domestic models are never cross-applied: each is trained on its own population and predicts only within that population.

The domestic total model predicts in log space ($y = \log(\text{ICEWS risk} + 10^{-4})$) rather than in levels. The log transformation accommodates the larger dynamic range of domestic risk: low-risk countries (e.g., France, Germany) have domestic ICEWS risk near zero, while high-risk countries (Syria, Yemen) have values approaching one. A level-space Ridge would compress predictions toward the mean, overstating risk for peaceful countries.

Table A9: Top 5 features by $|\beta^{\text{std}}|$ for each Ridge model.

Model	Feature	β^{std}
Total	CAMEO diversity (unique codes)	+0.037
	Verbal hostile events (count)	+0.029
	Hostile events (count)	+0.021
	Total media mentions	−0.020
	Cooperative events (count)	−0.016
Kinetic share	CAMEO diversity (unique codes)	−0.043
	Poldip risk (stock)	−0.032
	CAMEO 19: Fight	+0.031
	Kinetic risk (stock)	+0.030
	Military risk (stock)	+0.028
Posture share	Posture risk (stock)	+0.015
	CAMEO 15: Exhibit force	+0.014
	CAMEO 07: Provide aid	+0.007
	Military share of total	+0.006
	Log posture risk	+0.005
Poldip share	Log total risk	−0.662
	Log kinetic/posture ratio	+0.463
	Log kinetic risk	+0.437
	Log military/political ratio	−0.391
	Log poldip risk	+0.314
Trade share	Trade risk (flow)	+0.010
	Log trade risk	+0.010
	Trade risk (stock)	+0.006
	Trade share of total	+0.006
	CAMEO 16: Reduce relations	+0.003

Notes: β^{std} is the Ridge coefficient on standardised features (zero mean, unit variance). A one-standard-deviation increase in the feature changes the predicted outcome by β^{std}. "Flow" = current month; "stock" = exponentially weighted average ($\lambda = 0.95$). All models trained on the 2015–2019 GDELT/ICEWS overlap with 5-fold grouped cross-validation by dyad.

Table A10: CAMEO root event categories and our domain mapping.

Code	Description	Tone	Domain	Goldstein range
01	Make public statement	Mixed	—	$[-5.2, +3.4]$
02	Appeal	Coop.	—	$[+2.8, +3.0]$
03	Express intent to cooperate	Coop.	—	$[+3.0, +5.2]$
04	Consult	Coop.	—	$[+3.0, +3.4]$
05	Engage in diplomatic coop.	Coop.	—	$[+3.4, +7.4]$
06	Engage in material coop.	Coop.	—	$[+5.2, +7.4]$
07	Provide aid	Coop.	—	$[+7.0, +7.4]$
08	Yield	Coop.	—	$+5.0$
09	Investigate	Neutral	—	0.0
10	Demand	Conflict	Pol/Dip	$[-5.0, -3.0]$
11	Disapprove	Conflict	Pol/Dip	-2.2
12	Reject	Conflict	Pol/Dip	-4.0
13	Threaten	Conflict	Pol/Dip	$[-7.0, -4.4]$
14	Protest	Conflict	Pol/Dip	$[-6.5, -4.4]$
15	Exhibit force	Conflict	Posture	$[-7.2, -5.8]$
16	Reduce relations	Conflict	Trade	$[-7.0, -4.0]$
17	Coerce	Conflict	Trade	$[-7.0, -5.0]$
18	Assault	Conflict	Kinetic	$[-9.0, -8.0]$
19	Fight	Conflict	Kinetic	$[-10.0, -9.2]$
20	Unconventional mass violence	Conflict	Kinetic	-10.0

Notes: Goldstein scores from Schrodt (2012); ranges reflect variation across subcategories within each root code. Domain mapping used for the four-component decomposition. "Pol/Dip" = political–diplomatic. "Trade" includes economic sanctions and coercion. "Posture" = military threats and force displays without combat. "Kinetic" = actual use of force. Cooperative categories (01–09) enter only through the denominator of the conflict ratio; conflictual categories (10–20) enter both numerator and denominator. Category 09 ("Investigate") has Goldstein score 0; because the conflict ratio weights events by $|G_e|$ (conflict) or G_e (cooperation), neutral events contribute zero to both numerator and denominator regardless of their classification. Category 01 ("Make public statement") spans the full Goldstein range because subcategories include both praise (+3.4) and criticism (−5.2).

The log-space model prevents this by operating on a scale where the difference between 0.001 and 0.01 receives the same weight as between 0.1 and 1.0. Predicted values are exponentiated and clipped to $[0, 1]$.

Domain shares (kinetic, posture, political/diplomatic, trade) are calibrated separately for domestic events using the same share-model approach as the bilateral pipeline. The four-component decomposition is then applied identically.