ONLINE APPENDIX

When Development Divides: Chinese Projects, Relative Deprivation, and Conflict in Africa

Data Documentation and Leave-One-Country-Out Robustness

This appendix documents the principal geospatial data sources, the construction of the common grid-year panel, the cross-fitted measures of project-induced development benefit and relative deprivation, and leave-one-country-out robustness exercises that complement the empirical analysis in the main manuscript.

Part A. Data Sources and Variable Construction

The empirical analysis combines five spatially referenced data sources in a common annual framework: the Uppsala Conflict Data Program (UCDP) Georeferenced Event Dataset, harmonized DMSP–VIIRS nighttime lights, TerraClimate, the Geo-referencing Ethnic Power Relations (GeoEPR) dataset, and AidData's Geospatial Global Chinese Development Finance Dataset. Because these sources differ in their native temporal coverage and spatial resolution, we harmonize them to a common 25 km $\times$ 25 km grid over Africa and use the overlap required by each empirical specification. The source coverage reported below refers to the underlying data products; the estimation sample can be narrower because of treatment timing, outcome availability, and the covariates included in a given regression.

A.1 Spatial framework and annual harmonization

The core spatial unit is a fixed 25 km $\times$ 25 km grid cell. Point-based observations, such as conflict events and project locations, are linked to the grid using their geographic coordinates. Raster variables, including nighttime lights and climate measures, are summarized within each grid cell using the underlying raster pixels and then merged into the annual grid panel. Holding the spatial unit fixed over time prevents changes in administrative boundaries from mechanically changing the unit of analysis.

The resulting panel distinguishes true zero outcomes from missing source information. For conflict, a grid-year with no recorded event is coded as zero once that grid-year lies within the observed panel. Nighttime-light and climate observations are taken from their annual source layers rather than interpolated across missing years. Chinese projects enter the treatment construction through their geospatial locations

and recorded implementation timing; grid–project distances are then used to construct the ring and spline exposure measures used in the main analysis.

A.2 Conflict events: UCDP Georeferenced Event Dataset

Conflict outcomes are constructed from the UCDP Georeferenced Event Dataset (GED), Global Version 26.1. The GED records individual events of organized violence and provides event dates, geographic coordinates, actors, conflict identifiers, and fatality estimates. The global version covers 1989–2025 and distinguishes three principal categories of organized violence: state-based conflict, non-state conflict, and one-sided violence (Davies et al., 2026; Sundberg & Melander, 2013).

For the grid analysis, we use event coordinates to assign each georeferenced event to a 25 km grid cell and aggregate the number of events by grid and calendar year. The baseline outcome is therefore an event count rather than a fatality total. This outcome is well suited to the paper's PPML specifications because it measures the frequency of localized violent events while naturally retaining grid-years with zero events. Events that cannot be assigned to a usable geographic location are not mechanically allocated to a grid cell.

A key advantage of the GED is that its spatial and temporal coding allows conflict to be aligned closely with the local timing of development projects. At the same time, event data are subject to reporting and geocoding uncertainty, especially in remote locations or periods with limited source coverage. The leave-one-country-out exercises in Part B therefore provide an additional check that the main distributional results are not driven by the reporting environment or conflict history of any single country.

Data access: https://ucdp.uu.se/downloads/

A.3 Nighttime lights: harmonized DMSP–OLS and VIIRS

Nighttime lights are used as a spatially detailed proxy for local economic activity and development. The analysis uses the harmonized global DMSP–VIIRS nighttime-light series developed by Li and coauthors. This harmonized product is particularly useful for long-run panel analysis because the raw Defense Meteorological Satellite Program Operational Linescan System (DMSP–OLS) and Visible Infrared Imaging Radiometer Suite (VIIRS) products are not directly comparable across sensors.

For 1992–2013, the series uses DMSP–OLS stable nighttime-light observations calibrated to improve temporal comparability across satellites (Li & Zhou, 2017). From 2014 onward, VIIRS radiance observations are converted to DMSP-like digital-number values using the cross-sensor harmonization framework

developed by Li et al. (2020) Version 10 of the associated Figshare data record extends the annual harmonized series through 2024. The data are distributed as global GeoTIFF rasters at 30 arc-second resolution, approximately 1 km at the equator, with DMSP-style digital-number values on a common scale.

We overlay each annual raster with the 25 km grid and calculate the mean nighttime-light value across valid pixels in each grid-year. The principal transformed outcome is

$$\operatorname{asinh}(NTL) = \ln\left(NTL + \sqrt{NTL^2 + 1}\right).$$

The inverse-hyperbolic-sine transformation is defined at zero and behaves similarly to the natural logarithm for larger positive values, making it well suited to African grid cells that include both dark and brightly lit locations. Using a single harmonized product also avoids imposing a mechanical level shift at the DMSP–VIIRS sensor transition.

The harmonized series does not eliminate all measurement concerns. DMSP-style data can exhibit saturation in very bright areas and blooming around light sources, while very low digital-number values can be noisy. These features motivate robustness checks around the sensor transition and the treatment of low-light observations when appropriate, but they do not alter the baseline grid construction described above.

Dataset DOI: https://doi.org/10.6084/m9.figshare.9828827.v10

A.4 Climate controls: TerraClimate

Climate controls are drawn from TerraClimate, a high-spatial-resolution global dataset of monthly climate and climatic water-balance variables. The original Climatology Lab (2026) descriptor documents the 1958–2015 series at $1/24^\circ$ resolution with approximately 4 km (Abatzoglou et al., 2018). The product has subsequently been extended and revised. The Climatology Lab describes TerraClimate v1.1 as a 1950–present product and cautions users not to mix v1.0 and v1.1 files. The replication package therefore records the exact TerraClimate release used in estimation rather than treating all historical downloads as interchangeable. Variables relevant to this study include precipitation, minimum and maximum temperature, and the Palmer Drought Severity Index (PDSI). TerraClimate reports precipitation in millimeters, temperature in degrees Celsius, and PDSI as a unitless measure of relative wetness or dryness.

Monthly climate layers are first mapped to the 25 km grid and then aggregated to annual measures consistent with the estimation panel. To place temperature and precipitation fluctuations on comparable scales,

the analysis constructs standardized anomalies of the form

$$z = \frac{x - \mu}{\sigma},$$

where μ and σ are the reference mean and standard deviation used in the empirical data construction. PDSI can be used directly as a drought control: negative values indicate drier-than-normal conditions and positive values indicate wetter-than-normal conditions. These controls help separate project-associated changes in local economic activity and conflict from contemporaneous climatic shocks.

Data access: https://www.climatologylab.org/terraclimate.html

A.5 Ethnic geography: GeoEPR 2021

Ethnic-group geography is taken from GeoEPR 2021, the geospatial component of the Ethnic Power Relations (EPR) Dataset Family. GeoEPR geocodes politically relevant ethnic groups from EPR Core 2021, assigns each group a settlement pattern, and provides mapped settlement areas whenever a spatial representation is available. The 2021 release was published on June 8, 2021, and its polygons use the WGS84 coordinate reference system (Vogt et al., 2015; Wucherpfennig et al., 2011).

The analysis uses a fixed 2021 ethnic geography rather than allowing historical changes in group polygons to redefine grid membership over the outcome panel. We retain the localized settlement types *Regionally based*, *Regional & urban*, and *Urban*. Each 25 km grid cell is represented by its centroid and matched to GeoEPR polygons using a point-within-polygon join. If a centroid matches more than one retained polygon, the assignment rule is deterministic: *Regionally based* receives first priority, followed by *Regional & urban* and then *Urban*; remaining ties are resolved by ascending GeoEPR group ID. The procedure therefore assigns at most one ethnic group to each analysis grid and does not use polygon-area shares or an overlap threshold.

This fixed grid-to-group crosswalk is then used to construct the ethnic distributional measures. For a given project, only ethnic-tagged grids in the relevant project environment enter the project-by-ethnic-group-by-year construction. Keeping ethnic geography fixed is useful for interpretation because changes in ethnic own benefit and ethnic relative deprivation then arise from project exposure and predicted development gains rather than from year-to-year changes in the spatial coding of ethnicity. GeoEPR covers politically relevant ethnic groups rather than every ethnolinguistic community, so the ethnic analysis should be interpreted as applying to the politically relevant groups represented by GeoEPR.

Data access: https://icr.ethz.ch/data/epr/geoepr/

A.6 Chinese development projects: AidData Geo-GCDF Version 3.0

Chinese development-project locations and timing are drawn from AidData's Geospatial Global Chinese Development Finance Dataset, Version 3.0 (Geo-GCDF v3), linked to the underlying Global Chinese Development Finance Dataset (GCDF v3), see AidData (2023). The project-level GCDF v3 contains information on more than 20,000 projects in 165 low- and middle-income countries financed by official-sector institutions in China. Its commitment-year coverage is 2000–2021, with implementation timing recorded over 2000–2023. The geospatial version provides spatial information for 9,405 projects across 148 countries and includes precise or approximate project features as well as administrative-level locations when more exact geocoding is unavailable (Goodman et al., 2024)

The project records provide the information needed to characterize treatment exposure, including project location, implementing or financing institution, financing or assistance type, sector, implementation status and timing, and project description. We retain African projects with usable geographic information and the timing required by the empirical design. When multiple forms of spatial information are available, the construction uses the most informative project geography available to link projects to nearby grid cells and calculate distance-based exposure. The ring specification discretizes this exposure into distance bands, whereas the spline specification allows the relationship between distance and exposure to vary smoothly.

An important source-coverage clarification is that Geo-GCDF Version 3.0 should not be described as a dataset of new Chinese commitments through 2024. AidData reports commitments through 2021 and project-implementation information through 2023. Outcome data, such as conflict and nighttime lights, can extend beyond the last recorded project start and thereby provide post-treatment observations, but they do not extend the underlying AidData commitment window.

Data access: https://www.aiddata.org/data/aiddatas-geospatial-global-chinese-development-finance-dataset-version-3-0

A.7 Cross-fitted prediction of project-induced development benefit

The benefit measures used in the conflict analysis are not observed nighttime-light levels. Instead, they are out-of-fold predictions of the component of nighttime-light change associated with exposure to Chinese projects. This distinction converts the first-stage spatial difference-in-differences estimates into a predicted

project-induced development gain while reducing mechanical overfitting between the observations used to estimate the first stage and the observations to which the estimated treatment response is assigned.

The first stage is estimated on annual 25 km grid cells in project-centered stacks. The main event window is $[-5, +5]$ years around project implementation, and the 200–300 km band is the spatial reference group. The specification includes project-by-grid fixed effects and project-by-year fixed effects. Weather variables enter as controls, while the predicted development-benefit measure is constructed from the estimated treatment-response component rather than from contemporaneous weather. Standard errors in the first stage are two-way clustered by project and grid. For project p, grid g, and year t, the ring specification can be written as

$$\operatorname{asinh}(L_{pgt}) = \sum_{r} \beta_r \left[Post_{pt} \times \mathbf{1}\{d_{pg} \in r\} \right] + X'_{pgt}\theta + \alpha_{pg} + \lambda_{pt} + \varepsilon_{pgt},$$

where $Post_{pt} = \mathbf{1}\{t \geq T_p\}$ and the 200–300 km distance band is omitted. The omitted band is therefore normalized to zero.

As a smooth alternative, we replace the discrete ring indicators with a natural-cubic-spline basis in project–grid distance. In the spline specification, predicted treatment exposure is measured relative to the mean spline basis in the 200–300 km reference band, making the zero point comparable to the omitted category in the ring model.

We implement five-fold cross-fitting at the project level. Projects, rather than individual grid-year observations, are assigned to folds; all sites, grids, and years associated with the same project remain in the same fold. Folds are balanced using the number of first-stage observations. For each fold k, the first-stage model is estimated using the other four folds, or approximately 80 percent of projects, and the estimated treatment-response function is used only to predict benefits for projects in the held-out fold. Thus, no project contributes observations both to the estimation of its own treatment-response coefficients and to the prediction assigned to that project.

Let $k(p)$ denote the fold containing project p, and let the superscript $(-k(p))$ denote coefficients estimated after excluding that fold. For the spline construction, let $p^*(g)$ denote the single nearest eligible project

to grid g. The cross-fitted predicted development benefit for grid g in year t is

$$B_{gt}^{\text{Ring}} = \sum_{p} w_p \, \widehat{\beta}_{r(p,g)}^{(-k(p))} \mathbf{1}\{t \geq T_p\},$$

$$B_{gt}^{\text{Spline}} = \left[\mathbf{b}(d_{p^*(g)g}) - \overline{\mathbf{b}}_{200\text{–}300}\right]' \widehat{\boldsymbol{\beta}}^{(-k(p^*(g)))} \mathbf{1}\{t \geq T_{p^*(g)}\}.$$

Here w_p is the project or site weight used in the stacked design, T_p is the project implementation year, $r(p, g)$ is the distance ring containing grid g relative to project p, $\mathbf{b}(\cdot)$ is the spline basis evaluated at the project–grid distance, and $\overline{\mathbf{b}}_{200\text{–}300}$ is the mean spline basis over the 200–300 km reference band. In the ring model, the reference-band coefficient is zero by construction, and contributions from multiple relevant projects are accumulated across space. In the spline model, centering the basis on the reference region makes a zero prediction correspond to the same 200–300 km comparison region, and the predicted benefit relies exclusively on spatial exposure to the nearest eligible project.

A.8 Own development benefit and relative deprivation

At the grid level, own development benefit is the focal grid-year's cross-fitted predicted benefit, B_{gt}. Relative deprivation compares this own benefit with the development gains received by other nearby grids. For each focal grid g, we calculate an inverse-distance-weighted mean of the benefits received by other grids j within 75 km. The focal grid itself is excluded from the comparison set, so the neighbor benchmark is not mechanically contaminated by its own predicted benefit. Let d_{gj} denote the distance between grids g and j. The neighbor benchmark is

$$NB_{gt} = \frac{\displaystyle\sum_{\substack{j \neq g \\ d_{gj} \leq 75 \text{ km}}} d_{gj}^{-1} B_{jt}}{\displaystyle\sum_{\substack{j \neq g \\ d_{gj} \leq 75 \text{ km}}} d_{gj}^{-1}},$$

and grid-level relative deprivation is

$$RD_{gt} = \max\{NB_{gt} - B_{gt},\, 0\}.$$

This positive-part construction treats relative deprivation as a shortfall rather than as a symmetric difference. A grid is deprived only when nearby grids receive a larger predicted project-induced benefit; when the focal

grid does at least as well as its neighbors, relative deprivation is zero. An equal-weight neighbor mean can be used as a robustness alternative to the inverse-distance benchmark.

The ethnic extension changes the unit of comparison from nearby grids to localized ethnic groups within the same project environment. For project p, ethnic group e, and year t, ethnic own benefit is the simple mean of cross-fitted grid benefits among grids associated with group e:

$$EB_{pet} = \frac{1}{|G_{p,e}|} \sum_{g \in G_{p,e}} B_{gt},$$

where $G_{p,e}$ is the set of ethnic-tagged grids associated with group e in project p's local environment. Ethnic own benefit is therefore conceptually different from grid-level own benefit: it summarizes the average benefit captured by a group across its local spatial footprint rather than the benefit of a single focal cell.

The comparison benefit for "other groups" is the grid-count-weighted mean across all grids associated with ethnic groups other than e in the same project-year. Equivalently,

$$EB_{p,-e,t} = \frac{\sum_{e' \neq e} |G_{p,e'}| \, EB_{pe't}}{\sum_{e' \neq e} |G_{p,e'}|}.$$

Ethnic relative deprivation is then

$$ERD_{pet} = \max\{EB_{p,-e,t} - EB_{pet},\, 0\},$$

and is defined only when at least two ethnic groups are represented in the project environment.

The ethnic construction is formed at the project-by-ethnic-group-by-year level using ethnic-tagged grids within 300 km of the project and is then attached to the corresponding ethnic-tagged grid-year observations. In the second stage, PPML conflict models relate event counts to own benefit and relative deprivation, with grid fixed effects and country-by-year fixed effects. Ring-based and spline-based benefit constructions are carried through separately so that the conflict results do not depend on a single functional form for spatial decay.

Part B. Leave-One-Country-Out Robustness Checks

This section evaluates whether the estimated relationships between Chinese development exposure, own development benefit, relative deprivation, and conflict are disproportionately driven by any single national case. For each country in the estimation sample, we re-estimate the corresponding PPML specification after removing all observations from that country. Each point in Figures 1–4 is therefore obtained from a separate regression. Horizontal bars report 95 percent confidence intervals, the dashed vertical line marks the corresponding full-sample coefficient, and the solid vertical line marks zero. In each figure, the left panel reports the ring specification and the right panel reports the spline specification.

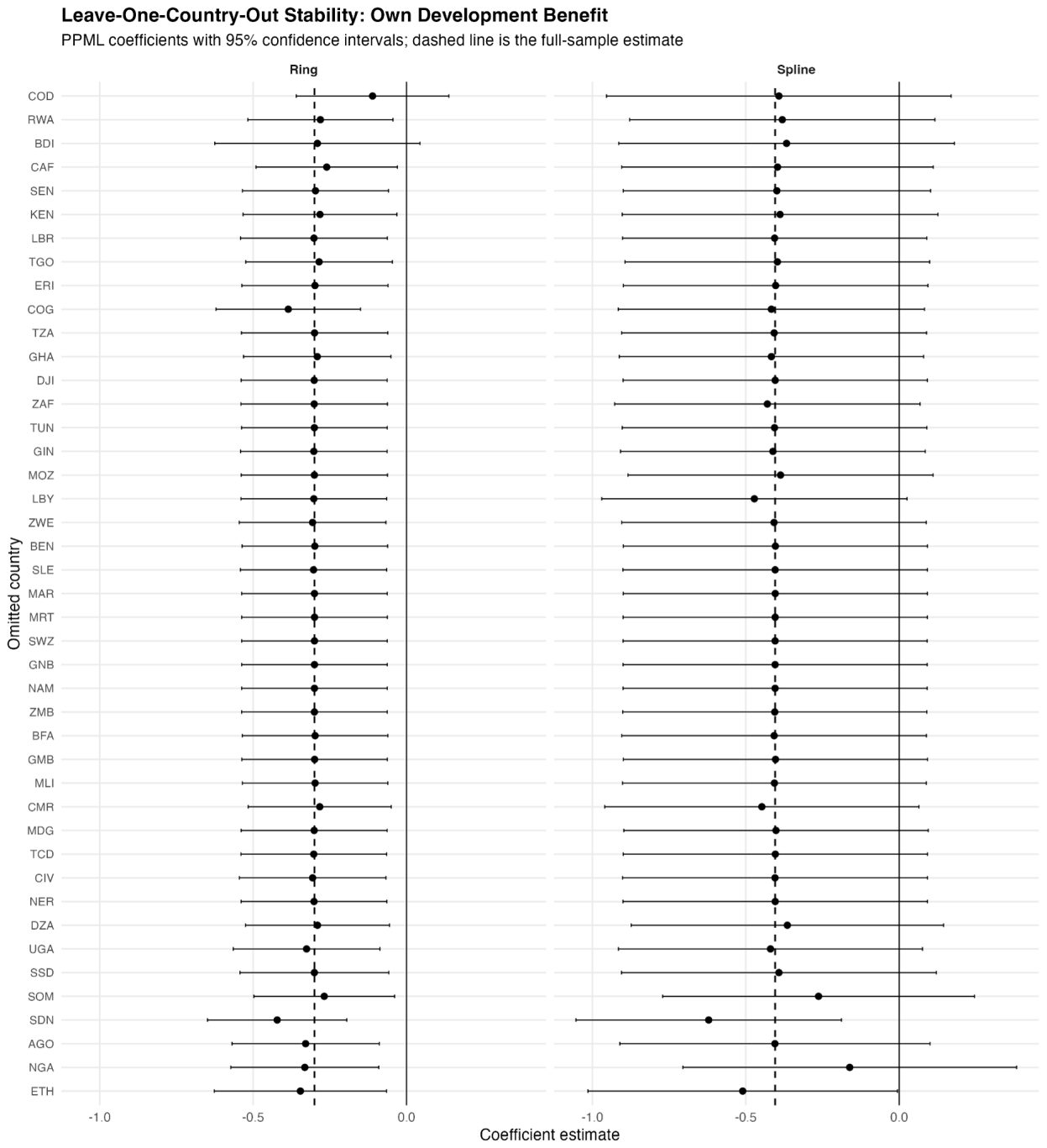
The leave-one-country-out design is more informative than a simple comparison of country weights because a country can be influential through several channels simultaneously: the number of treated grid-years, the number of observed conflict events, the timing and geographic concentration of projects, or the distribution of the constructed benefit and deprivation measures. Stability of the coefficient across these re-estimations indicates that the result is not an artifact of one unusually large or conflict-intensive country.

B.1 Grid-level distributional measures

Figures 1 and 2 report the leave-one-country-out results for grid-level own development benefit and relative deprivation, respectively. Across country omissions, the point estimates generally remain close to the corresponding full-sample estimates in both the ring and spline specifications. Some omissions generate larger movements or wider confidence intervals, as expected when the excluded country contributes substantial identifying variation, but the overall pattern is not concentrated in a single national case.

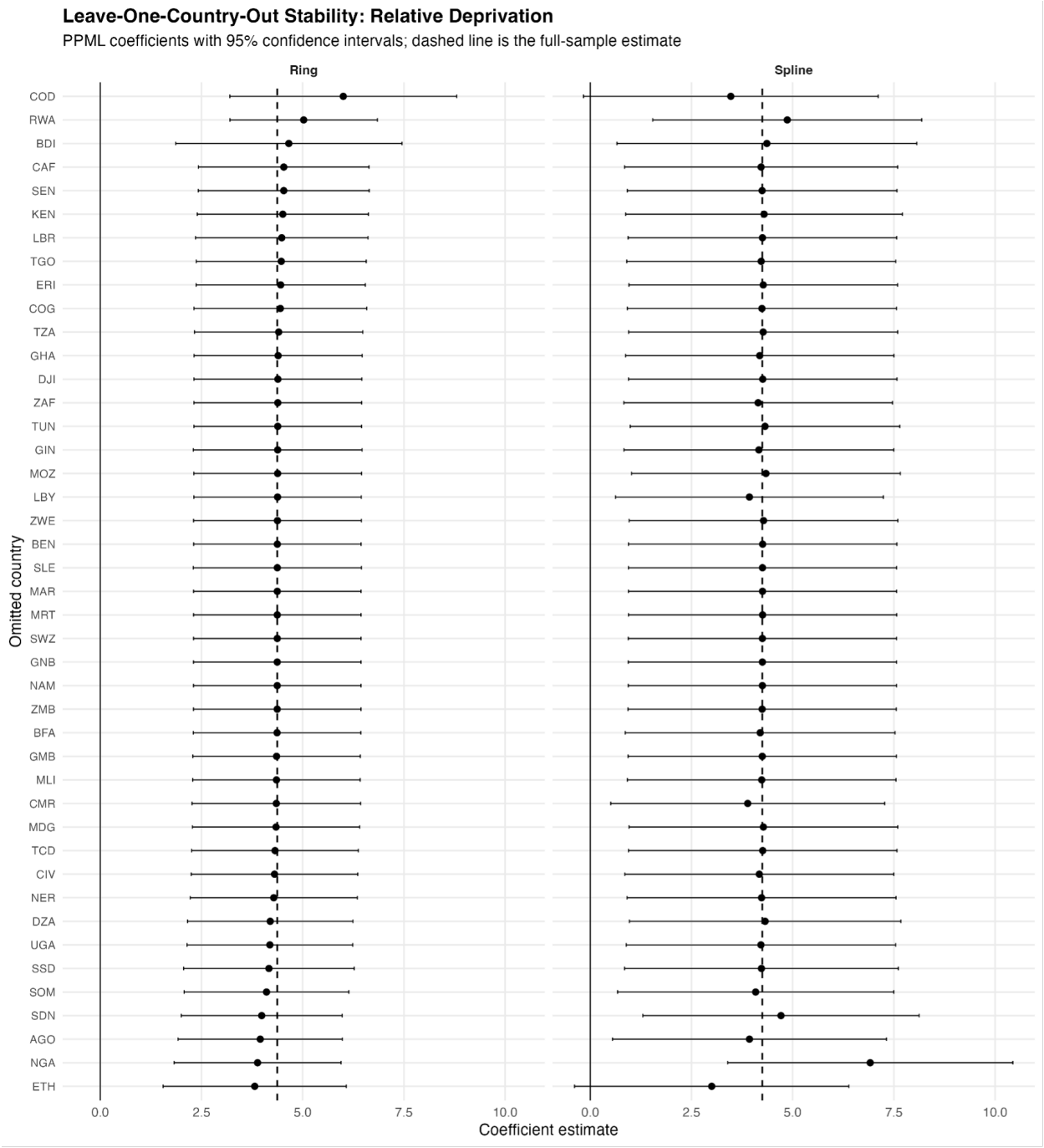
The two figures also provide a joint diagnostic of the distributional mechanism. In Figure 1, the own-benefit estimates remain centered around the negative full-sample benchmark under both spatial specifications. In Figure 2, the relative-deprivation estimates remain centered around a positive full-sample benchmark. Although the precision of individual leave-one-country-out estimates varies, omitting one country at a time does not systematically overturn the qualitative contrast between own benefit and relative deprivation. This pattern is consistent with the distinction built into the measures: own benefit captures the focal grid's predicted gain, whereas relative deprivation captures the positive shortfall relative to nearby beneficiaries.

Figure 1: Grid-level own development benefit



Notes: Each point is the PPML coefficient after omitting one country. Horizontal bars are 95% confidence intervals, the dashed vertical line marks the full-sample estimate, and the solid vertical line marks zero. Ring and spline refer to the two alternative ways of modeling spatial exposure to Chinese projects.

Figure 2: Grid-level relative deprivation



Notes: Each point is the PPML coefficient from a regression that omits the country shown on the vertical axis. Horizontal bars are 95% confidence intervals. The dashed vertical line is the corresponding full-sample estimate; the solid vertical line denotes zero. The exercise asks whether the positive association between relative deprivation and conflict depends on any single country.

B.2 Ethnic-group distributional measures

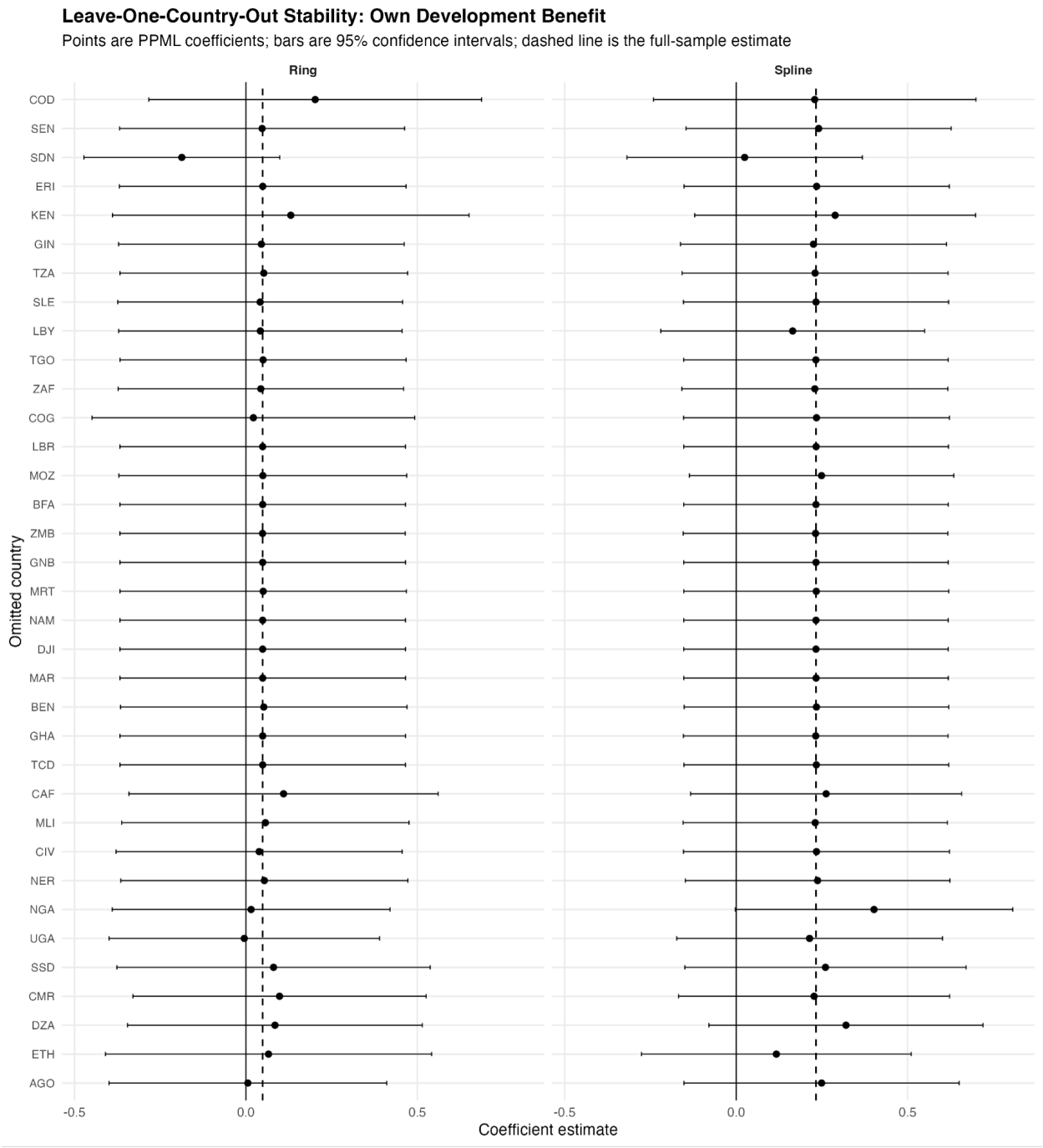
Figures 3 and 4 repeat the same diagnostic for the ethnic-group versions of own development benefit and relative deprivation. These measures allow project-associated gains and deprivation to vary with the distribution of ethnic groups around project locations rather than only with neighboring grid cells. The leave-one-country-out exercise is especially informative here because the ethnic composition of project areas and the spatial concentration of violence can differ substantially across countries.

The qualitative comparison focuses on whether the re-estimated coefficients remain near the full-sample benchmark and whether excluding a particular country reverses the substantive interpretation. As with the grid-level measures, the plots indicate broad stability across country omissions, although uncertainty around individual leave-one-country-out estimates can be considerably larger than in the full sample.

The ethnic-group own-benefit coefficient should not be expected to reproduce mechanically the sign of the grid-level own-benefit coefficient. The two variables answer different questions. Grid-level own benefit is the predicted gain of one focal cell, whereas ethnic own benefit averages the gains received by all grids associated with an ethnic group in a project environment. A positive ethnic own-benefit coefficient can therefore coexist with a negative grid-level own-benefit coefficient without constituting a coding contradiction. Figure 3 instead asks whether the group-level relationship is itself dependent on one national case; the estimates remain broadly centered on the full-sample benchmark across country omissions.

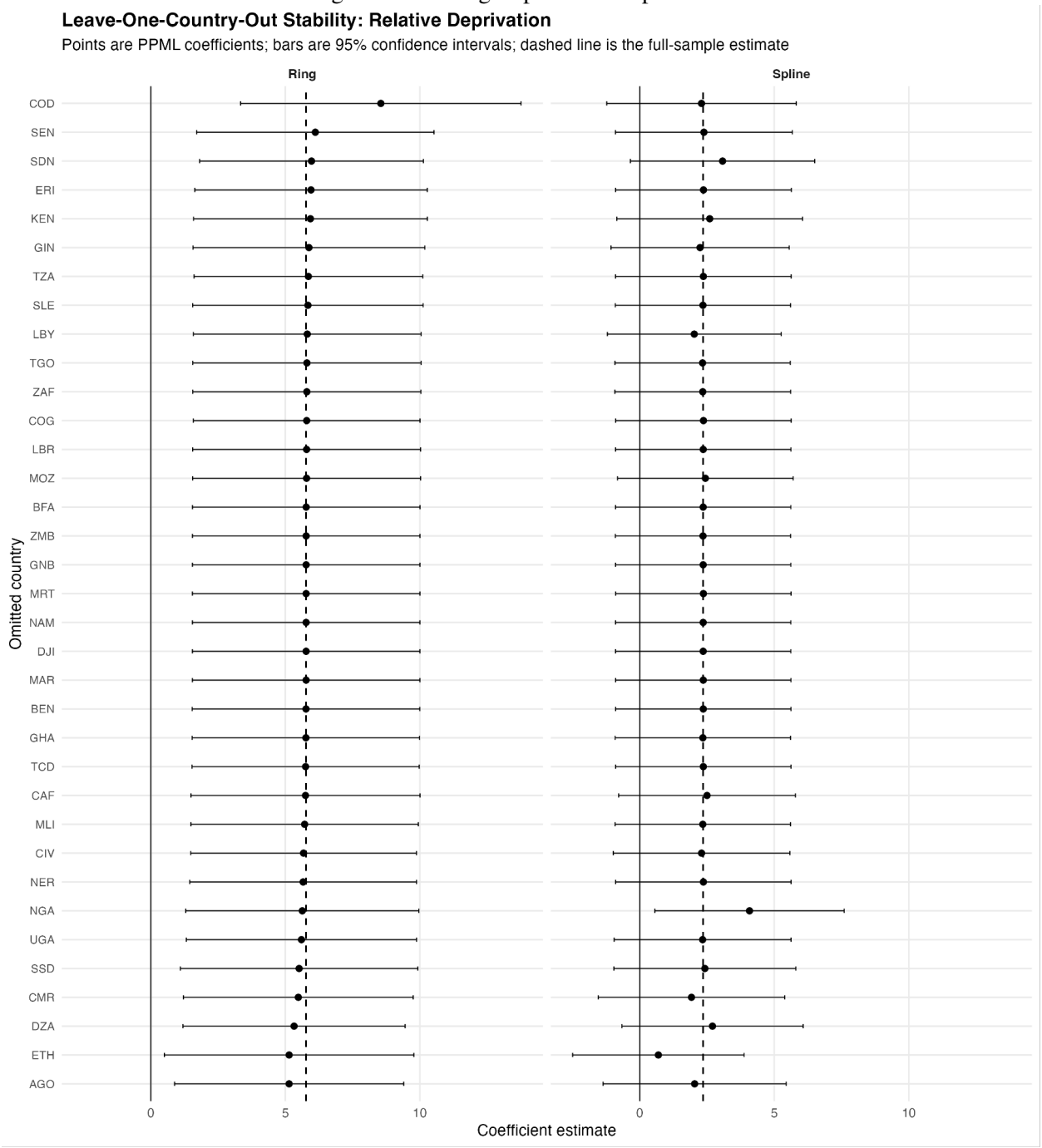
For ethnic relative deprivation, the comparison is explicitly within project-year: the focal group's average benefit is compared with the grid-count-weighted average benefit of all other represented ethnic groups, and only a positive shortfall is retained. The leave-one-country-out exercise removes the entire set of projects, ethnic groups, and conflict observations contributed by the omitted country. Figure 4 therefore provides a demanding influence diagnostic for this more aggregated construction. The positive relationship between ethnic relative deprivation and conflict is not visibly attributable to a single country, although the width of the confidence intervals varies across omissions.

Figure 3: Ethnic-group own development benefit



Notes: Each point is the PPML coefficient after omitting one country. Horizontal bars are 95% confidence intervals, the dashed vertical line marks the full-sample estimate, and the solid vertical line marks zero. Ring and spline refer to the two alternative ways of modeling spatial exposure to Chinese projects.

Figure 4: Ethnic-group relative deprivation



Notes: Each point is the PPML coefficient after omitting one country. Horizontal bars are 95% confidence intervals, the dashed vertical line marks the full-sample estimate, and the solid vertical line marks zero. The figure evaluates whether the estimated conflict response to ethnic relative deprivation is driven by a single national case.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*, 170191.

AidData. (2023). Global chinese development finance dataset, version 3.0 [Data set. AidData at William & Mary]. https://www.aiddata.org/data/aiddatas-global-chinese-development-finance-dataset-version-3-0

Climatology Lab. (2026). Terraclimate [Retrieved August 29, 2026]. https://www.climatologylab.org/terraclimate.html

Davies, S., Pettersson, T., & Öberg, M. (2026). Organized violence 1989–2025, and violent political protests. *Journal of Peace Research*, *63*(4), 705–723.

Goodman, S., Zhang, S., Malik, A. A., Parks, B. C., & Hall, J. (2024). AidData's geospatial global chinese development finance dataset. *Scientific Data*, *11*, 529.

Li, X., & Zhou, Y. (2017). A stepwise calibration of global DMSP/OLS stable nighttime light data (1992–2013). *Remote Sensing*, *9*(6), 637.

Li, X., Zhou, Y., Zhao, M., & Zhao, X. (2020). Harmonization of DMSP and VIIRS nighttime light data from 1992–2024 at the global scale [Version 10. Data set. Figshare].

Sundberg, R., & Melander, E. (2013). Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, *50*(4), 523–532.

Vogt, M., Bormann, N.-C., Rüegger, S., Cederman, L.-E., Hunziker, P., & Girardin, L. (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution*, *59*(7), 1327–1342.

Wucherpfennig, J., Weidmann, N. B., Girardin, L., Cederman, L.-E., & Wimmer, A. (2011). Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset. *Conflict Management and Peace Science*, *28*(5), 423–437.