(Mis)measuring Local Weather Shocks

Evidence from Drought and Migration in Ethiopia

Marius Hees[1,*] Michiel Gerritse[1] Frank van Oort[1]

[1]Erasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute

Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands.

[*]*hees@ese.eur.nl*

September 12, 2026

Abstract

Extreme weather shocks strike locally, yet their effects are often estimated from exposure aggregated across districts, regions, or countries. We show that this aggregation can systematically bias impact estimates, and we propose a simple correction. Using high-resolution rainfall and satellite-derived population data for more than 13,000 Ethiopian settlements from 2000 to 2016, we find that at fine resolution, a drought shock reduces population in vulnerable settlements by about 2.5 percent. The losses concentrate among working-age adults, consistent with out-migration, and coincide with falling food production, harvest storage, and income, along with deteriorating health outcomes. When drought shock exposure is measured at progressively coarser scales, population impacts can no longer be detected, even at resolutions commonly used in the literature. We trace this bias to three distinct causes: partly exposed areas are classified as fully treated or untreated, local vulnerability is averaged away, and area-weighted exposure misrepresents where the exposed population lives. This bias decomposition implies a simple correction: measure each shock at fine scale, weight it by local vulnerability, and only then aggregate the exposure. This correction recovers impact estimates in outcome units roughly four times larger in area than those at which the standard binary indicator fails; simulations confirm its performance. Studies relying on aggregated shock exposure may therefore miss real effects; building exposure locally before aggregating is a low-cost change that can substantially improve empirical practice.

JEL classification: C18, C21, C23, O15, Q54, R12

Keywords: drought; migration; measurement error; spatial aggregation; Ethiopia

1 Introduction

Weather shocks are central to empirical work on economic development. Studies use rainfall variation to estimate how shocks affect conflict (Miguel et al., 2004), human capital (Maccini & Yang, 2009), child marriage (Corno et al., 2020), labor markets (Jayachandran, 2006; Kaur, 2019), health (Dinkelman, 2017), and migration (Lagakos, 2020), for instance. As climate change increases the frequency and intensity of droughts and other extreme weather events, this evidence increasingly informs adaptation policy, drought early warning systems, and the targeting of social protection in low-income countries. The credibility of this evidence depends on an integral but often overlooked part of the research design: how exposure to the weather shock is constructed.

By far, the predominant way to define a weather shock (or other localized shock) is to collect rainfall levels across grids or measurement stations, aggregate them to the region, province, or other level of analysis, and declare a shock if the rainfall level meets a threshold. A common example is that outcomes such as wages, productivity, or health, are measured at an administrative level, and the weather shock data are aggregated to match the administrative level. In a systematic review of influential papers on localized shocks, we find that more than 75 percent of the coded specifications aggregate exposure to match coarser outcome data (44 of 57).[1]

This paper shows that aggregating local rainfall in larger areas prior to defining treatment meaningfully biases impact estimates. Area averages of local exposures to weather shocks do not generally represent the actual exposure of the area. A drought can hit part of a district hard while leaving the rest largely unaffected. The vulnerability of settlements to rainfall deficits can also differ even among neighboring settlements, for example because some

[1] The audit sample of 59 papers is constructed using a prespecified, fully reproducible OpenAlex pipeline; the printed tables contain 57 specification-level codings covering 55 of these papers, as two papers are coded twice because they employ two distinct analysis scales. Appendix A reports the search procedure, inclusion criteria, coding protocol, and results.

rely on rainfed agriculture while others do not. Defining binary shock indicators on the underlying localized rainfall distribution can structurally mismeasure the actual exposure and attenuate the estimates. We develop this argument around drought because rainfall shocks are central in this literature, but the logic is more general: it applies to floods, storms, wildfires, pollution, and conflict whenever exposure strikes locally but outcomes are observed for larger areas.

We document how exposure aggregation changes estimated drought impacts in Ethiopia. We construct population and rainfall data at an unusually fine scale for the entire country. We combine high-resolution gridded rainfall (Funk et al., 2015) with satellite-derived settlement boundaries and annual population estimates for more than 13,000 Ethiopian settlements between 2000 and 2016. At the finest scale, drought reduces population in rural rainfed settlements by about 2.5 percent. The decline is concentrated among working-age adults, consistent with out-migration rather than mortality. Similar settlements outside rainfed agriculture show no such decline, supporting the drought exposure channel. Urban settlements in adjacent unaffected districts gain population, consistent with nearby inflows. Georeferenced household data from the Ethiopian Rural Socioeconomic Survey show that the same droughts raise adverse shocks to food production and income, deplete grain storage, and worsen health. Ethiopia is a demanding and policy-relevant setting: most rural agriculture is rainfed, buffering through irrigation, insurance, or safety nets is limited (Dercon, 2004; Dercon et al., 2014), and drought-induced displacement is a first-order policy concern.

We then examine how the impact estimates change as the exposure measure progressively coarsens due to aggregation. We randomly merge neighboring districts, recompute the drought indicator for each merged unit, and assign that value back to the underlying settlements. Because the sample and outcomes remain unchanged, the loss of signal reflects exposure aggregation alone. Under the standard binary indicator for drought shocks, the estimated effect attenuates rapidly and becomes statistically indistinguishable from zero, once units exceed roughly 50 by 50 kilometers. Much of the literature measures exposure at a coarser resolution when using districts, regions, or countries. Similar patterns appear for food production, storage, income, and health. The estimates from common practices for aggregation would likely miss real responses to weather shocks.

A decomposition explains why aggregation biases the impact estimate. The relevant exposure is the interaction between a local shock and local vulnerability to that shock. Standard practice often aggregates the shock, while treating vulnerability as either absorbed by unit fixed effects or captured by a coarse proxy. This creates three distortions. First, thresholding an aggregated shock misclassifies partially exposed units. Second, aggregating shocks separately from vulnerability removes the within-unit covariance between shock intensity and vulnerability. Third, vulnerability proxies whose means are not preserved under aggregation put the wrong weight on exposed populations. Unit fixed effects absorb time-invariant differences in vulnerability levels, but they do not recover the interaction between time-varying local shocks and local vulnerability. Difference-in-differences designs face the same problem because the distortion is built into the treatment measure itself. A calibrated gravity simulation, in which the true migration response is known and exposure construction is the only source of distortion, reproduces both the attenuation of the standard measure and the stability of the corrected one.

The decomposition of the bias points to a simple solution: build exposure locally first and aggregate afterward. The shock component can usually be constructed at the native resolution of gridded data, even when outcomes are observed only for larger units. Constructed from the native resolution, the corrected measure weights each local shock by local vulnerability before aggregation. In Ethiopia, this means interacting drought intensity with local dependence on rainfed agriculture. In other settings, the same logic can use building quality for storms, drainage and elevation for floods, or proximity to contested territory for conflict. These vulnerability proxies are often easier to obtain than high-resolution outcome data. In the Ethiopian data, the vulnerability-weighted measure recovers the fine-scale estimate and remains statistically significant in units of up to roughly 10,600 square kilometers, about four times larger in area than the roughly 2,500 square kilometers at which the standard binary indicator fails.

Taken together, these findings yield three contributions. First, the paper provides country-wide, fine-scale evidence on the effects of drought on population and development outcomes. We show that drought reduces population in vulnerable rural settlements, that the decline is concentrated among working-age adults, and that the same shocks reduce production, storage, income, and health. Second, it explains why aggregated exposure can make such effects disappear. Third, it provides a practical correction, together with a scale-sensitivity diagnostic, that applied researchers can use with existing data.

These contributions speak to the broader research program on measurement in development economics (Beegle et al., 2024), to the tradition of cautionary results on rainfall-based research designs (Sarsons, 2015), and most directly to recent work on weather data in socioeconomic research. Michler et al. (2022) and Josephson et al. (2026) show that estimates can depend strongly on which weather product researchers use. We study a different margin of measurement. Holding the weather product fixed, we show that the spatial construction of exposure, namely the scale of aggregation, the discretization of the shock, and the treatment of local vulnerability, can itself determine whether an effect is detected. The two margins are complements: robustness across weather products does not solve the problem if exposure is then averaged and recoded at the wrong scale.

The paper also connects to work showing how aggregation and measurement error can distort estimates of impact. Research on ecological inference and the modifiable areal unit problem shows that aggregation does not generally preserve the relationship present at local levels (Robinson, 2009; Gehlke & Biehl, 1934; Openshaw, 1984); Appendix W relates this problem to our setting. The bias we describe does not derive from non-exogenous grouping or heterogeneous coefficients, but from aggregation practice per se. Work on measurement error shows that nonclassical error can bias estimates in ways that depend on how the regressor is constructed (Aigner et al., 1973; Bound et al., 2001). Specifically, Berkson (1950) shows that linearly weighted aggregation generally preserves consistency. We argue that the most common estimators for impacts of shocks do not satisfy that condition, and that the interaction of vulnerability and shock (i.e. rainfed agriculture dependence and droughts in Ethiopia), even if properly aggregated by assigning group means to members, does not generally satisfy the Berkson conditions. Hence, the common estimator in applied work likely defies the Berkson conditions for aggregation. Recent work on spatially misaligned data studies inference when outcomes and regressors are observed on different supports, often when the independent variable is coarser than the outcome (Gotway & Young, 2002; Pouliot, 2023). We discuss the opposite problem: the independent variable has a higher resolution than the outcome variable, so the issue is right-hand side aggregation, instead of interpolation (in addition to the potential interaction between vulnerability and the shock).

Substantively, the paper contributes to work on drought, coping, and mobility in low-income settings (Dercon, 2004; Kazianga & Udry, 2006; Fafchamps et al., 1998; Gray & Mueller, 2012; Mueller et al., 2020). It provides, to our knowledge, the finest-scale country-wide evidence on drought-induced population movements in Ethiopia, and links those movements to broader welfare losses. The results also connect to the basis-risk literature on index insurance, where area-level indices fail when they do not track local exposure (Dercon et al., 2014). Finally, recent evidence that micro and macro estimates of climate damages diverge sharply (Bakkensen & Barrage, 2026) is consistent with the mechanism we isolate: aggregation can remove the local shock–vulnerability variation that identifies the effect.

The paper proceeds as follows. Section 2 describes the Ethiopian setting and data. Section 3 presents the fine-scale effects of drought on population and welfare. Section 4 tests how these estimates change as exposure is coarsened. Section 5 develops the decomposition that explains why aggregated exposure fails and derives the corrected measure. Section 6 evaluates the vulnerability-weighted measure and uses a calibrated simulation to isolate the mechanism. Section 7 concludes.

2 Setting and Data

Ethiopia allows us to study aggregation directly because drought exposure and population can both be measured at unusually fine spatial scale. We use this setting to examine drought and internal migration while holding settlement-level outcomes fixed and varying only how exposure is constructed. We combine three data sources: high-resolution gridded rainfall, satellite-derived settlement boundaries with annual population estimates, and a georeferenced household survey. The main rainfall and settlement data cover the full country and are built from widely available raw inputs, which limits concerns about selective coverage and makes the approach portable to other settings.

Figure 1 illustrates the aggregation problem. The left-hand map shows 2001 drought shocks at the woreda level, Ethiopia's third-tier administrative units and the baseline exposure units in our analysis. The right-hand map overlays zones, the larger second-tier units often used in empirical work. Many zones contain both drought-affected and unaffected woredas. Defining drought at the zone level therefore labels some unaffected areas as treated and some affected areas as untreated. This is the basic misclassification problem that motivates the analysis.

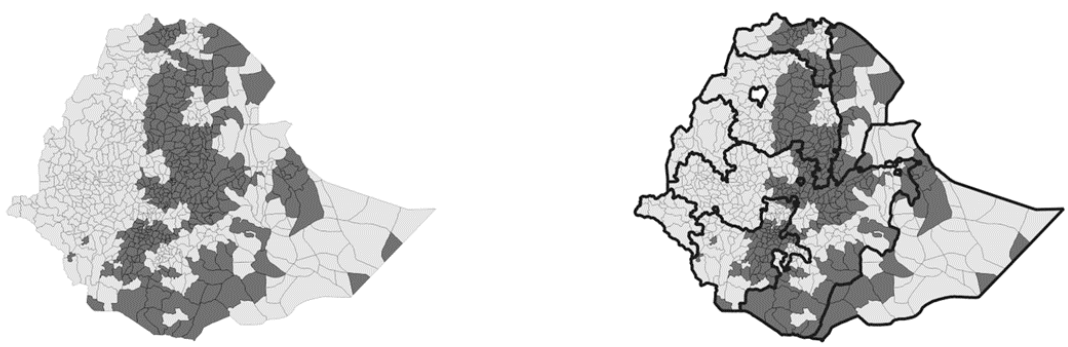


Figure 1: Drought exposure and administrative aggregation in Ethiopia, 2001. *Notes:* The left map shows third-tier administrative districts (woredas) affected by drought (dark grey) and unaffected (light grey). The right map overlays second-tier administrative boundaries (zones) in black. Many zones contain both affected and unaffected districts, illustrating the misclassification created by aggregation (Wong, 2004).

Ethiopia is a useful setting because drought risk and vulnerability both vary sharply across space. Droughts often affect some districts while nearby districts receive normal rainfall. Climate projections point to greater rainfall intensity and variability, with more frequent droughts in some areas and improved agricultural conditions in others (Conway & Schipper, 2011). Vulnerability also differs across places. We focus on rainfed regions, where agriculture depends almost entirely on direct rainfall, and distinguish rural from urban settlements.[2] This distinction matters because about 90% of rural residents are engaged in agriculture, so rainfall shocks should affect rural and urban livelihoods differently.

The link between rainfall and economic outcomes is also unusually direct. Irrigation, insurance, and formal safety nets remain limited in much of rural Ethiopia (Dercon, 2004; Dercon et al., 2014). As a result, rainfall shortfalls are more likely to translate into realized income and production shocks. This makes Ethiopia a demanding setting for testing whether exposure aggregation changes estimated drought impacts.

2.1 Rainfall and drought measurement

We measure rainfall using the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) dataset (Funk et al., 2015). CHIRPS combines satellite-based precipitation estimates with ground-station data to produce quasi-global gridded rainfall maps at high spatial resolution.[3] We extract monthly precipitation at roughly 5.5×5.5 km resolution from 1981 onward and aggregate it to annual totals. Raster tiles that cross district boundaries are weighted by the share of their area inside each district.

To define an adverse rainfall shock, we follow the majority of the economic literature. We construct district-level drought indicators using the gamma-distribution approach with a twentieth-percentile cutoff (Corno et al., 2020; Burke et al., 2015). Specifically, we estimate district-specific gamma distributions on annual rainfall and classify years below the twentieth

[2] Ethiopia's Central Statistics Agency defines urban areas as settlements with more than ten thousand inhabitants.

[3] CHIRPS combines satellite estimates with a global network of meteorological stations, integrated through geostatistical interpolation.

percentile as drought years. Alternative measures use additional information on evaporation or soil moisture, such as the Combined Drought Indicator and the Standardized Precipitation Evapotranspiration Index. For comparability, we use nominal rainfall, but note that the aggregation challenge equally applies to other locally measured shocks.

Rainfall shortfalls do not have the same economic meaning everywhere. In irrigated areas, nominal rainfall shocks can be buffered. We therefore distinguish rainfed from non-rainfed districts using the agricultural land-system map of Kassawmar et al. (2018), which identifies the parts of Ethiopia where agriculture depends almost entirely on direct rainfall, roughly 95% of agricultural water use. We overlay this rainfed polygon on district boundaries and classify a district as rainfed when at least half of its area falls inside the rainfed region. Focusing the main tests on these districts strengthens the link between rainfall deficits and local economic shocks. Appendix B documents the drought construction, Appendix C describes the rainfed classification, and Appendix D reports the spatial autocorrelation structure of the resulting drought indicators.

2.2 Settlements and population

We construct settlement boundaries for all of Ethiopia from WorldPop's Built-Settlement Growth Model. The input data are annual 100×100 meter binary grids that identify built-up cells from satellite imagery, interpolated between remotely sensed observations through 2014 and extrapolated thereafter using land-cover classifications, population data, and nighttime lights (Nieves et al., 2020). We turn these built-up cells into settlement polygons with a custom clustering algorithm. First, we place a 500-meter buffer around each built-up cell to connect nearby structures and capture low-density settlement sprawl. Second, we merge overlapping buffers into continuous settlement clusters. Third, we assign longitudinal identifiers so that settlements can be followed consistently over time, including cases in which previously separate clusters merge. This procedure yields 13,376 settlement polygons observed in an unbalanced panel from 2000 to 2016.

Annual population counts come from WorldPop's top-down, UN-adjusted people-per-pixel product at 100-meter resolution, available annually for 2000–2020 (Stevens et al., 2015). The product disaggregates census totals to grid cells using a random-forest model trained on publicly available covariates such as built-up extent, roads, land cover, and nightlights. We overlay these population rasters on the settlement polygons we construct to produce settlement-year population estimates for the 13,376 unique settlements.

Measurement error in these population estimates is unlikely to explain our results: the aggregation experiment holds the resolution of outcome data fixed and changes only how exposure is constructed.[4]

A related concern is that the population product is itself a model prediction, and some of its input covariates, such as vegetation indices and nighttime lights, can respond to weather directly. If drought lowered predicted population through these covariates rather than through actual population change, the fine-scale estimates would partly reflect the construction of the outcome. Three features of the results argue against this reading. First, meteorological drought, and any vegetation response to it, occurs in rainfed and non-rainfed districts alike, yet population declines only in rainfed districts (Section 3.2); a mechanical weather-to-covariate channel should depress predicted population in both. Second, the evidence that does not rely on WorldPop points the same way: after the same droughts, ERSS households report adverse shocks to production and income and worse health, and the ERSS-based demographic decomposition shows that the decline is concentrated among working-age adults (Appendix G), a pattern a prediction artifact would not generate. Third,

[4] Any error in the population data is therefore constant across aggregation levels and exposure measures, and cannot generate the divergence between the standard and corrected exposure measures. For the fine-scale estimates, settlement fixed effects absorb time-invariant allocation error, while residual year-to-year allocation error should mainly inflate standard errors unless it is systematically related to lagged local rainfall within settlements. If slow-moving allocation covariates smooth true annual population changes, the estimated response would be attenuated toward zero. Appendix E describes the boundary algorithm, and Appendix F reports population diagnostics.

the implied departure rate of about 1.9 percent lies within the range of survey-based estimates of drought-period migration in rural Ethiopia (Section 3.3). A residual channel in which drought reduces economic activity, and with it nighttime lights, precisely in rainfed areas cannot be excluded outright, but it would have to reproduce these demographic and survey patterns to account for the results.

2.3 Household survey (ERSS)

To examine outcomes beyond migration, we use georeferenced variables from the Ethiopian Rural Socioeconomic Survey (ERSS), collected by the Central Statistics Agency of Ethiopia. The survey covers more than 4,000 rural households across three waves within our sample period and reports detailed information on production, consumption, nutrition, income, and assets.

We focus on ten georeferenced outcomes that capture adjustment margins other than migration: self-reported productivity shocks, crop damage, adverse shocks to food production and food stocks, harvest in storage, cereal and meat intake, health problems, and adverse shocks to income and assets. These variables are chosen because their geographic precision can be aligned with the rainfall and population data. The ERSS also provides demographic information used to distinguish migration from fertility and mortality responses in Appendix G. Appendix H describes the survey design and variable definitions.

The survey data are useful but spatially sparser than the settlement panel. The settlement panel covers 13,376 settlements nested in 690 third-tier districts. ERSS households fall within fewer than 350 of those districts. The survey outcomes therefore provide complementary evidence on welfare responses, but the population panel remains the main dataset for the aggregation experiment.

3 Drought, Migration, and Welfare at Fine Scale

We first establish the fine-scale effect of drought in Ethiopia. This step serves two purposes. Substantively, it shows whether drought affects population and welfare when exposure is measured at the finest available scale. Methodologically, it gives us a benchmark against which to compare estimates based on coarser exposure measures. We begin with the settlement-level population specification, then estimate the population response, examine whether the decline reflects migration rather than mortality, and finally use the ERSS to document welfare effects beyond migration. Section 4 then asks whether these effects survive when exposure is progressively coarsened.

3.1 Empirical Specification

We estimate the effect of drought on settlement-level population using a two-way fixed effects model:

$$\log Y_{i,g,t} = \beta D_{g,t-1} + \psi_i + \gamma_t + \varepsilon_{i,g,t}, \tag{1}$$

where $\log Y_{i,g,t}$ is the log population of settlement i in district g and year t. The treatment variable $D_{g,t-1}$ equals one if district g experienced a drought in the previous year. We lag exposure by one year because the population estimates draw on imagery collected at varying points within the year, and because migration responses take time to organize and execute (Lagakos, 2020). Settlement fixed effects ψ_i absorb time-invariant differences across settlements, and year fixed effects γ_t absorb common shocks and national trends. Standard errors are clustered at the district level.[5] Settlements are classified as rural or urban using the definition of Ethiopia's Central Statistics Agency, and as located in rainfed or non-rainfed

[5]Drought is spatially correlated across districts (Appendix D), so district-level clustering may understate uncertainty if neighboring districts experience related shocks. Our central results concern how point estimates move as exposure is coarsened, rather than any single significance test, and the calibrated simulation reproduces the same attenuation pattern in the absence of sampling uncertainty. We therefore report district-clustered standard errors throughout and do not rest any conclusion on the precision of a single estimate.

districts using the classification described in Section 2.

The treatment variable is constructed from annual rainfall aggregated to the district level. For each district, we estimate a gamma distribution using annual rainfall from 1981 to 2016 and classify years below the twentieth percentile as drought years, following common practice in the rainfall-shock literature (Corno et al., 2020; Burke et al., 2015). In the estimation sample, 18.4 percent of district-years are classified as drought years, close to the twenty percent implied by construction. The difference reflects the discreteness of annual data and the fact that the drought threshold is estimated using the full 1981–2016 rainfall record, while the population panel covers 2000–2016.

The specification is deliberately parsimonious. Settlement fixed effects account for permanent differences in size, location, market access, and local geography. Year fixed effects account for nationwide population trends and common climatic shocks. Appendix I verifies that the implied two-way fixed effects weights are almost entirely positive, ruling out the negative-weight pathologies discussed by De Chaisemartin & d'Haultfoeuille (2020). Appendix J shows that adding controls for neighboring drought density, repeated shocks, and district area has negligible effects on the drought coefficient. Appendix K states the identifying assumptions and relates them to features of the data.

3.2 Fine-scale estimates

Table 1 reports estimates of Equation (1) for rural settlements. Columns 1 and 2 use the standard discrete drought indicator; columns 3 and 4 use the vulnerability-weighted continuous measure developed in Section 5.[6] Within each pair, we separate rainfed settlements, where agriculture depends directly on local rainfall, from non-rainfed settlements, which serve as a placebo group with limited economic exposure to rainfall shocks.

[6] In brief: local drought exposure is weighted by local dependence on rainfed agriculture, and the resulting product is aggregated to the district only afterward. Section 5 derives the measure and explains why this construction matters.

This table establishes two central results. First, drought reduces population in rural rainfed settlements by roughly 2.5 percent, statistically significant at the one percent level under both exposure measures. Second, in non-rainfed settlements the coefficients are small and statistically indistinguishable from zero, consistent with weaker economic exposure and supporting the drought channel.[7] At the finest scale, the two exposure measures deliver essentially the same estimate in the rainfed sample. They part ways only once exposure is aggregated, which is the comparison at the center of the paper.

Table 1: Impact of drought exposure on log population: standard versus vulnerability-weighted exposure.

Dependent variable:	Log population			
	Discrete indicator		Vulnerability-weighted	
Sample:	Rainfed	Non-rainfed	Rainfed	Non-rainfed
Drought exposure (first lag)	-0.0247^{***}	-0.0138	-0.0257^{***}	0.0445
	(0.0081)	(0.0139)	(0.0082)	(0.0641)
Settlement fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	84,705	17,920	84,705	17,920
R^2	0.271	0.349	0.271	0.349
Within R^2	7.87×10^{-5}	1.94×10^{-5}	8.4×10^{-5}	7.16×10^{-5}

Notes: Standard errors, clustered at the district level, in parentheses. Rural settlements only; rainfed and non-rainfed samples defined by the classification of Section 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.3 Migration, not mortality

A potential concern is that population decline is not driven by migration, but by excess mortality. Three results support the interpretation of migration.

First, the decline is concentrated among working-age adults rather than among the groups most exposed to mortality risk. Appendix G decomposes population changes by demographic

[7] The binary rainfed classification masks within-district heterogeneity: several non-rainfed districts contain pockets of rainfed agriculture, and the continuous vulnerability-weighted measure assigns these areas small positive weights. The column 4 estimate is imprecise (standard error 0.0641) and statistically indistinguishable from zero; we attach no interpretation to its sign.

source and shows that fertility and mortality responses cannot account for the estimated decline.

Second, urban settlements in non-drought districts adjacent to drought-affected districts gain population, consistent with short-distance inflows into nearby towns. Adjacent districts as a whole do not show differential population growth (Appendix L). These two facts are consistent: town-level gains are small relative to district totals and are diluted in district averages. We therefore read the pattern as evidence of nearby migration flows, not as a complete accounting of destinations.

Third, drought effects are smaller in border districts, where cross-district migration is costlier (Appendix J.3). If the estimated losses reflected mortality or measurement artifacts rather than movement, they should not depend on the cost of crossing district boundaries. Border districts also differ from interior districts in remoteness, market access, and agroclimate, so this evidence is corroborating rather than decisive on its own. Taken together, the three checks point to migration as the main margin of adjustment.

The magnitudes are plausible once net and gross flows are distinguished. The 2.5 percent estimate is a net settlement-level population change, combining departures and forgone arrivals in the year after a drought. Interpreted as an equilibrium effect in the gravity mapping introduced in Section 6, this net decline implies a departure rate of roughly 1.9 percent. This falls between survey-based estimates for rural Ethiopia in Gray & Mueller (2012): about 1.4 percent annual male labor migration under normal conditions and 2.6 percent under severe drought.

Destination effects are harder to detect because outflows disperse across many receiving settlements. Per-destination inflows are therefore small relative to normal growth variation. This is why nearby towns, where inflows concentrate, show detectable gains while the average adjacent settlement does not. Small and spatially dispersed inflows are exactly the type of response that becomes difficult to detect once outcomes or exposure are aggregated.

3.4 Beyond migration: production, income, and health

The population response captures only one margin of adjustment to drought. Before turning to aggregation, we show that the same drought measure is associated with broader welfare losses. We use georeferenced household outcomes from the ERSS and focus on four representative variables that span production, storage, income, and health. The full set of ten outcomes is described in Appendix H and reported in Appendix M.

Table 2 reports fine-scale two-way fixed effects estimates. Drought exposure increases the probability of adverse shocks to food production and income, reduces harvest in storage, and raises the incidence of recent health problems. These results show that the drought indicator captures real economic damage beyond the migration margin.

Table 2: Effect of drought on production and welfare indicators.

Dependent variable:	Adverse shock to food production	Harvest in storage (kg)	Adverse shock to income	Health problems, past two months
Drought shock (first lag)	0.1478***	−69.37***	0.1456***	0.0273**
	(0.0333)	(24.10)	(0.0323)	(0.0121)
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	10,849	8,162	10,849	15,426
R^2	0.327	0.249	0.332	0.031
Within R^2	0.023	0.003	0.022	0.0003

Notes: Standard errors, clustered at the district level, in parentheses. Outcomes from the ERSS, described in Section 2. The full set of ten outcomes is reported in Appendix M. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 The Consequences of Aggregation for Impact Estimates

Do the fine-scale drought effect estimates survive the exposure construction commonly used in applied work? We re-examine the estimates, holding the settlement-level outcomes and the estimation sample fixed, while progressively coarsening only the exposure measure. Because nothing else changes, any movement in the estimates reflects how exposure is constructed rather than changes in outcomes, sample composition, or model specification.

4.1 Randomized area aggregation

To trace how exposure aggregation affects the estimates, we apply a randomized merger algorithm. Starting from the full set of districts, we repeatedly merge two neighboring districts at random, recompute the drought indicator for each resulting polygon, and assign that value to all settlements inside it. The indicator is recomputed from the rainfall data at every step. We aggregate the constituent rainfall series to the merged polygon, estimate the polygon-specific gamma distribution, and apply the twentieth-percentile cutoff. Exposure at each aggregation level is therefore constructed in the same way as the baseline district-level indicator.

This procedure ensures that changes in the coefficient estimate are not driven by a mechanical recoding of existing district indicators. Instead, it reflects the mixing of affected and unaffected areas when rainfall is averaged over larger spatial units. We then re-estimate the baseline two-way fixed effects regression on the original settlement-year panel using the progressively coarsened treatment variable. The number of observations remains fixed throughout, so any loss of precision reflects changes in the regressor rather than a smaller sample. Because each merger path depends on the evolving set of neighboring units, we repeat the procedure across 65 independent random aggregation paths and record the mean coefficient and median p-value at each aggregation level.

4.2 Aggregation and the population estimates

Figure 2 presents the results for the standard discrete drought indicator. The coefficient attenuates rapidly from its fine-scale value of about -0.025 toward zero. It becomes statistically indistinguishable from zero once average unit size exceeds roughly 50×50 km. At coarser levels, some estimates even turn positive. A researcher observing the same settlements but measuring drought exposure at these larger spatial units would therefore fail to detect the population response documented in Section 3.

The pattern is exactly what one would expect if aggregation mixes exposed and unexposed places. Larger polygons increasingly contain both drought-affected and unaffected districts. The assigned drought measure therefore becomes less informative about the exposure of any given settlement. Since the outcome data remain unchanged, the attenuation comes from the treatment measure itself.[8]

The practical implication is important. In the audit of Appendix A, 32 of the 37 codings with clearly classifiable spatial units, or 86 percent (56 percent of all 57 codings), measure exposure at the district level or coarser; the remaining 20 codings use firm-, community-, grid-cell-, or village-level units for which the district comparison is not defined. Many use regional or national units: spatial scales at or above the point where the standard indicator fails in our setting.[9] A researcher observing Ethiopian drought and population data at those scales would conclude that drought has no detectable effect on population. That would be a false negative created by exposure construction, not by the absence of a real effect.

The audit does not imply that every estimate in the literature is biased. It shows that

[8] A complementary exercise in Appendix N shows that aggregating the outcome variable produces similar attenuation through a distinct mechanism: within-unit migration is absorbed, compressing observable variation in population changes. Appendix P shows formally that aggregating outcomes does not remove the exposure wedge created by separately aggregated shock and vulnerability proxies.

[9] The printed audit tables contain 57 specification-level codings covering 55 of the 59 sampled papers; two papers are coded twice because they employ two distinct analysis scales. Counting all 57 codings, 56 percent measure exposure at district level or coarser; the remainder are finer-than-district, grid-cell or village, or non-areal, firm- or community-level, designs.

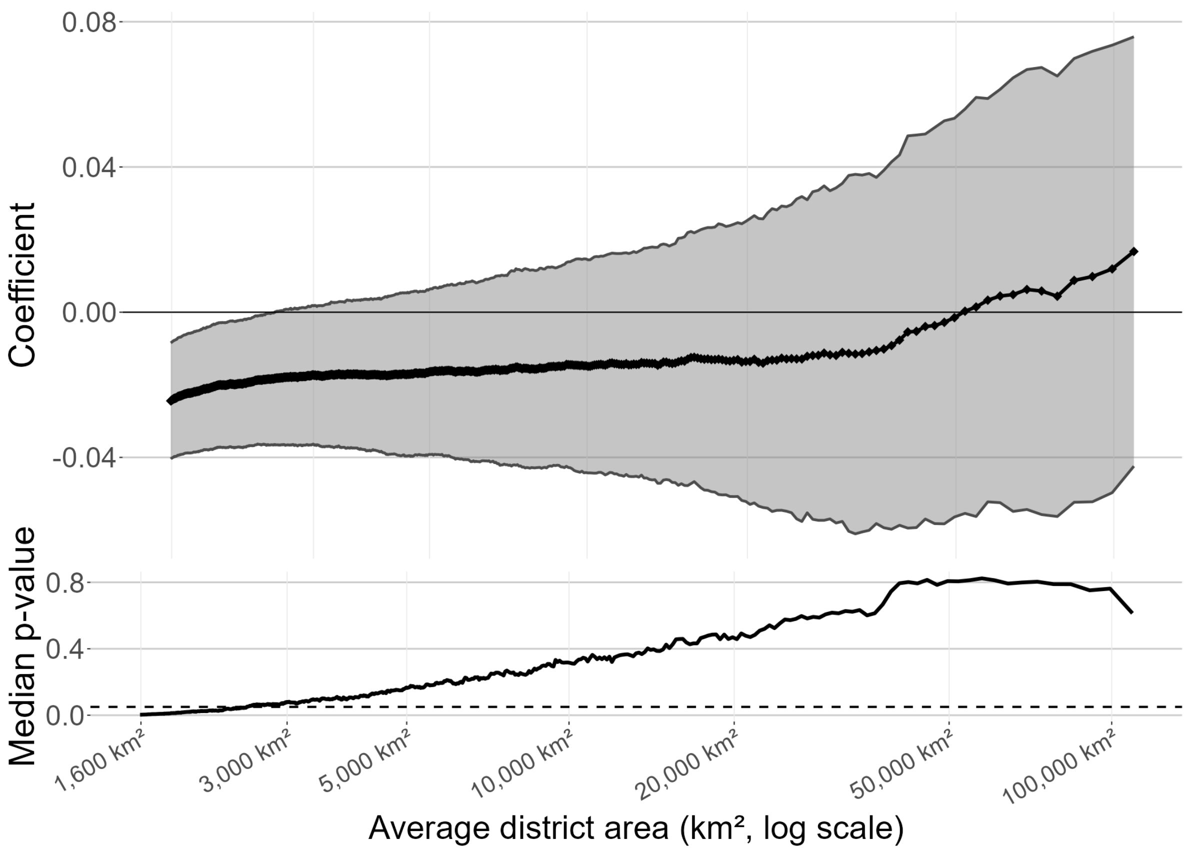


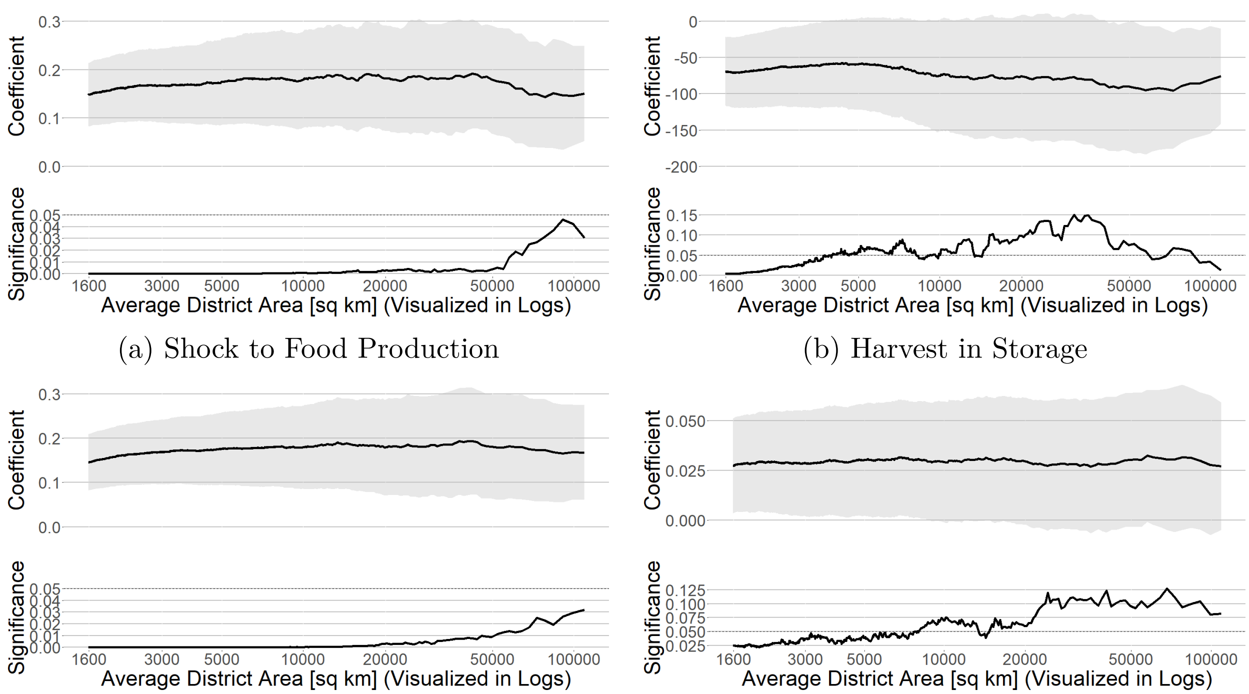
Figure 2: Aggregation sensitivity of the drought effect on log population: standard discrete indicator. *Notes:* Rural rainfed settlements. Upper panel: mean coefficient across 65 random aggregation paths; the shaded band is the mean coefficient $\pm$ 1.96 clustered standard errors, averaged across paths. Lower panel: median p-value across paths, with the 0.05 threshold dashed.

the relevant practice is widespread. The problem should be most severe when exposure is averaged over large units, converted into a discrete treatment indicator, and applied to populations whose vulnerability varies within those units. The next section shows that this pattern is not specific to population: the same attenuation appears for household measures of production, income, and health.

4.3 Aggregation and the survey outcomes

We apply the same randomized merger algorithm to the ERSS outcomes. The exercise is identical to the population analysis: households and survey waves are held fixed, while only the spatial scale of drought exposure changes.

Figure 3 shows the results. Across all four outcomes, the pattern mirrors the population estimates. Coefficients begin close to their fine-scale values, then move toward zero as exposure is measured over larger units. Precision also falls as aggregation increases. The point at which the signal disappears differs across outcomes. Food production remains informative over a wider range of spatial scales, while harvest in storage loses precision more quickly. This variation is expected because the outcomes differ in baseline variance, sample size, and spatial structure.



(a) Shock to Food Production

(b) Harvest in Storage

(c) Shock to Income

(d) Health Problems

Figure 3: Aggregation sensitivity of drought effects on four ERSS outcomes. *Notes:* Each panel applies the randomized merger algorithm to a different dependent variable from the ERSS, holding households and survey waves fixed while coarsening only exposure.

The ERSS results should be read as supporting evidence rather than as the main aggregation test. The survey covers about 4,000 households across fewer than 350 districts and three waves, so some aggregation steps operate on a relatively sparse spatial grid. Even so, the same qualitative pattern appears across production, storage, income, and health.

Coarsening exposure makes the estimated drought effects weaker and less stable.

Together, the population and survey results show that aggregation can erase real effects. The outcome data, sample, and specification remain fixed, yet the estimated effects fade when the drought measure is constructed at coarser spatial scales. The next section explains why this happens and why the problem is not solved by fixed effects or by standard difference-in-differences logic.

5 Bias Decomposition: Why Aggregated Exposure Fails

This section analyzes how the OLS estimator changes with aggregation to understand why aggregation leads the estimated population effects to disappear. The key issue is that the relevant exposure is local: it depends on both the intensity of the shock and the vulnerability of the place hit by the shock. When researchers aggregate shocks first and define treatment afterward, they can break this link. The decomposition identifies three distinct channels: misclassification from coarsening, loss of within-unit covariance between shocks and vulnerability, and incorrect weighting of vulnerable populations. Together, these channels explain why the estimates deteriorate as exposure is aggregated and motivate the correction developed in Section 6.

5.1 Setup and aggregation

Let locations i be nested within larger observational units g. The local shock is $S_{i,t}$ and local vulnerability is V_i.[10] The fine-scale data generating process is

$$\log Y_{i,t} = \beta V_i S_{i,t} + \varepsilon_{i,t}, \tag{2}$$

[10] Allowing vulnerability to vary over time, $V_{i,t}$, does not change the logic, only the notation. We treat V_i as time invariant for simplicity. With location fixed effects, V_i in levels is absorbed, but β remains identified from within-location variation in $V_i S_{i,t}$ as long as $S_{i,t}$ varies over time.

where $E[\varepsilon_{i,t} \mid V_i S_{i,t}] = 0$.

The object that matters for the outcome is the local interaction $V_i S_{i,t}$. In the Ethiopian application, this is the interaction between local drought exposure and dependence on rain-fed agriculture. In other settings, it could be the interaction between flood intensity and drainage, storm exposure and building quality, or conflict exposure and proximity to contested territory.

In many empirical applications, however, outcomes are observed at a coarser level than shocks. Researchers therefore aggregate the shock and vulnerability separately,

$$S_{g,t} = F(\{S_{i,t}\}_{i \in g}), \qquad V_g = G(\{V_i\}_{i \in g}),$$

where F and G are aggregation rules. In practice, these rules often involve averaging, thresholding, binning, or other nonlinear recodings. The empirical regression then uses the coarsened proxy $V_g S_{g,t}$ instead of the fine-scale interaction $V_i S_{i,t}$:

$$\log Y_{i,t} = \beta V_g S_{g,t} + u_{i,t}. \tag{3}$$

The problem is that aggregation and interaction do not generally commute. In general,

$$E[V_i S_{i,t} \mid g] \neq V_g S_{g,t},$$

whenever shocks and vulnerability vary within units and are aggregated separately. A unit can therefore have the same average rainfall shock as another unit, but very different effective exposure if its most vulnerable places are the ones hit hardest.

To see the wedge directly, define the within-unit discrepancies

$$\varepsilon_{S_{i,t}} \equiv S_{i,t} - S_{g,t}, \qquad \varepsilon_{V_i} \equiv V_i - V_g.$$

Substituting these into the regression error yields

$$u_{i,t} = \beta\big(V_g\varepsilon_{S_{i,t}} + \varepsilon_{V_i}S_{g,t} + \varepsilon_{V_i}\varepsilon_{S_{i,t}}\big) + \varepsilon_{i,t}. \tag{4}$$

Equation (4) shows that coarsening exposure creates a wedge between actual local exposure and the proxy used in the regression. This wedge is the source of aggregation-induced bias. Appendix AA derives the decomposition step by step and states the moment conventions used below.

5.2 Bias decomposition

The wedge in Equation (4) contains three terms. Each term reflects a different way in which aggregation can make the assigned exposure differ from the exposure that matters locally. Because these terms are functions of the coarsened regressor $V_gS_{g,t}$, they enter the OLS estimand.

Appendix O gives the full derivation, including the remainder term. Under the maintained orthogonality condition $\mathrm{Cov}(V_gS_{g,t}, \varepsilon_{i,t}) = 0$, pooled OLS from regressing $\log Y_{i,t}$ on $V_gS_{g,t}$ satisfies

$$\hat{\beta} = \beta\Big[1 + B_1 + B_2 + B_3 + \Gamma\Big], \tag{5}$$

where, writing $c_{g,t} \equiv E[\varepsilon_{V_i}\varepsilon_{S_{i,t}} \mid g, t]$ for the average within-unit product of the two discrepancies in unit-year (g, t),

$$\begin{aligned}
B_1 &= \frac{E[V_g]^2\ \mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) + E[S_{g,t}]^2\ \mathrm{Cov}(V_g, \varepsilon_{V_i})}{\mathrm{Var}(V_gS_{g,t})},\\
B_2 &= \frac{\mathrm{Cov}\left(V_gS_{g,t},\ c_{g,t}\right)}{\mathrm{Var}(V_gS_{g,t})},\\
B_3 &= \frac{E[V_g]E[\varepsilon_{V_i}]\ \mathrm{Var}(S_{g,t}) + E[S_{g,t}]E[\varepsilon_{S_{i,t}}]\ \mathrm{Var}(V_g)}{\mathrm{Var}(V_gS_{g,t})},
\end{aligned}$$

and Γ collects cross-covariance and third-moment terms that vanish under mean-preserving aggregation with weights aligned to the estimation sample.

Proposition 1 (Aggregation-induced bias decomposition) *Under* (2) *and* (3), *and assuming* $\text{Cov}(V_g S_{g,t}, \varepsilon_{i,t}) = 0$, *the pooled OLS estimand from regressing* $\log Y_{i,t}$ *on* $V_g S_{g,t}$ *admits the exact decomposition*

$$\hat{\beta} = \beta[1 + B_1 + B_2 + B_3 + \Gamma],$$

where the three components correspond to misclassification from coarsening, within-unit sorting between shock intensity and vulnerability, and mean shifts from incorrect weighting, and the remainder Γ *collects cross-covariance and third-moment terms. Under mean-preserving aggregation with weights aligned to the estimation sample,* B_1, B_3, *and* Γ *vanish, and the entire aggregation bias reduces to the sorting component* B_2. *The exact expressions are derived in Appendix O.*

This is an accounting decomposition of a single covariance wedge, not a set of mutually exclusive structural mechanisms. The same fine-scale primitives can enter more than one term. Still, each component isolates a common way in which applied exposure measures fail.

Component B_1: misclassification from coarsening. The first component arises when an aggregate proxy covaries with the difference between assigned and local exposure. This is especially likely when shocks are averaged over a large area and then converted into a binary indicator. A district classified as drought-exposed can contain settlements with mild or no drought, while a district classified as unexposed can contain severely affected settlements. The assigned treatment is then systematically related to the measurement error it creates.

This channel explains why binary indicators can attenuate quickly as the unit grows. In difference-in-differences language, the problem is not only whether treated and untreated units would have followed parallel trends. The treatment indicator itself must also classify local exposure correctly, or at least misclassify it in a way that is unrelated to assigned treatment. Coarsened binary indicators often fail this requirement. Appendix Q signs this channel for binary and categorical treatment measures and provides exact zero conditions.

Component B_2: within-unit sorting between shocks and vulnerability. The second component captures the interaction problem between vulnerability and shocks. Even if shocks and vulnerability are each aggregated in a mean-preserving way, so that each deviation is orthogonal to its own aggregate (satisfying the Berkson, 1950 conditions variable by variable), the product of the aggregates need not equal the aggregate of the product. Within each unit-year,

$$E[V_i S_{i,t} \mid g, t] = V_g S_{g,t} + c_{g,t},$$

where $c_{g,t}$ equals the within-unit covariance between shock intensity and vulnerability under mean-preserving aggregation. When the more vulnerable locations inside a unit are also the locations hit hardest, $c_{g,t} > 0$ and the coarsened regressor understates actual exposure in that unit-year.

The direction of the resulting slope bias is subtle. A sorting level common to all unit-years shifts measured exposure everywhere equally and is absorbed by the intercept or the fixed effects; only sorting that covaries with assigned exposure across unit-years, $\mathrm{Cov}(V_g S_{g,t}, c_{g,t}) \neq 0$, biases the coefficient. In the typical case, sorting concentrates precisely in the unit-years with high measured exposure, for example when a drought strikes the vulnerable part of a district: the proxy then understates exposure roughly in proportion to exposure itself, and the estimated coefficient is inflated away from zero. If sorting instead concentrates in unit-years with low measured exposure, the estimate attenuates. Either way, this is not standard noise in the treatment variable: the proxy removes the local interaction that determines exposure, and the sign and size of the distortion depend on the joint spatial distribution of shocks and vulnerability. Appendix R provides the covariance representation, exact zero conditions, and the sign discussion.

Component B_3: incorrect weighting of vulnerable populations. The third component arises when the aggregate vulnerability or shock measure is centered on the wrong population or spatial support. This channel has a parallel in the statistical literature (Gotway & Young, 2002) in the change of support problem: the spatial variable support is defined on the measure on which it is averaged. If vulnerable populations are not distributed uniformly across space, an area-weighted exposure measure can put the wrong weight on the places that matter for the outcome: in our application, rainfall and vulnerability are averaged over area, while outcomes are averaged over settlements which are non-uniformly distributed over space.

Component B_3 reflects a weighting channel. The problem is not only whether V_g or $S_{g,t}$ is high or low, but whether each preserves the relevant mean for the estimation sample. If the vulnerability or shock proxy is systematically too high or too low for the exposed population, the interaction coefficient loads that miscentering onto genuine variation in the other component. The bias vanishes when both proxies are aggregated using weights aligned with the estimation sample. Appendix S develops this point formally, and Appendix X formalizes how mismatched aggregation weights and supports generate such mean shifts.

Together, the three components show why aggregation can produce attenuation, amplification, or unstable inference. The overall bias can come from treatment misclassification, from removing the covariance between shocks and vulnerability, or from weighting exposed populations incorrectly. Which channel dominates, and hence the direction of the net distortion, depends on the spatial distribution of shocks, vulnerability, and outcomes. In our data the aggregated estimates attenuate toward zero, consistent with the misclassification and weighting channels dominating; the correction in Section 6 removes all three channels by construction, so it does not require taking a stand on their relative importance.

5.3 Why fixed effects do not solve the problem

Unit and time fixed effects do not generally remove this bias. Fixed effects absorb average differences across units and common changes over time. They do not reconstruct the local interaction between a time-varying shock and local vulnerability.

Let $\widetilde{Z}$ denote the two-way within transformation of a variable $Z_{i,t}$. Under the within-transformed analogue of the orthogonality condition, the two-way fixed effects estimator is

$$\hat{\beta}_{\mathrm{FE}} = \beta \frac{\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \widetilde{V_i S_{i,t}})}{\mathrm{Var}(\widetilde{V_g S_{g,t}})} = \beta \left[1 + \frac{\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \widetilde{\varepsilon}_x)}{\mathrm{Var}(\widetilde{V_g S_{g,t}})} \right], \tag{6}$$

where ε_x is the exposure wedge created by replacing $V_i S_{i,t}$ with $V_g S_{g,t}$. Appendix T shows that the fixed effects bias vanishes exactly when

$$\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \widetilde{\varepsilon}_x) = 0.$$

A sufficient condition is that the exposure wedge be additively separable into a unit component and a time component: then the within transformation absorbs it. But the aggregation wedge is generally a location-by-time object: it changes when shocks hit different parts of the same unit in different years. In that case, fixed effects remove the average level of vulnerability and common shocks, but they do not restore the missing within-unit covariance between shocks and vulnerability. In particular, fixed effects absorb a sorting level that is common across unit-years, but not sorting that varies with where shocks strike within units. The bias channels in B_1, B_2, and B_3 therefore persist.

6 Correcting Exposure Measures

The estimator's bias following aggregation points to a simple design principle elaborated in this section: construct exposure locally first and aggregate only afterward. We discuss this correction and then show how it improves performance in Ethiopian data as well as in a calibrated simulation in which we can control the parameters to be recovered. The section ends by turning the exercise into a diagnostic that can be used in other settings.

6.1 Implication for exposure construction

The bias decomposition implies that the bias disappears if shocks and vulnerability are constant within each unit, so that $S_{i,t} = S_{g,t}$ and $V_i = V_g$ for all $i \in g$. This removes any misclassification from coarsening as well as within-area covariances between vulnerability and shocks.

Given that researchers studying shocks often have access to high-resolution shock data, the plausible alternative to achieve this, is to form the exposure measure at the finest available support:

$$X_{i,t} \equiv V_i S_{i,t}.$$

Only then should exposure be aggregated to the unit used in estimation:

$$X_{g,t} \equiv \sum_{i \in g} \omega_{ig} X_{i,t}, \qquad \sum_{i \in g} \omega_{ig} = 1,$$

where the weights ω_{ig} are chosen to match the estimation sample. Relative to using $V_g S_{g,t}$, this measure preserves the within-unit interaction between shock intensity and vulnerability. For binary shocks, the resulting regressor is no longer a binary indicator. It is the vulnerability-weighted exposed share of the unit.

This logic also motivates a scale-sensitivity diagnostic. Researchers can construct exposure at different spatial scales while holding outcomes fixed, and trace the coefficient path as exposure is coarsened. If the estimated effect fades as exposure is aggregated, the design is sensitive to exposure construction. If a locally built exposure measure remains stable, it provides evidence that the original attenuation came from aggregation rather than from the absence of an effect.

6.2 Empirical performance of the vulnerability-weighted measure

Figure 4 compares the standard discrete drought indicator with the vulnerability-weighted continuous exposure measure. The standard indicator is shown in grey and the vulnerability-weighted measure in black. The difference is large. The standard indicator attenuates quickly as units grow. The vulnerability-weighted measure remains close to the fine-scale estimate over a much wider range of aggregation levels, and its median p-value stays below 0.05 for unit sizes up to roughly 10,600 km^2. This is about four times the area at which the standard binary indicator loses significance.

The result is that the corrected weighted exposure measure identifies the original coefficient up to much higher levels of data aggregation. In Appendix U, we show that the correction also improves robustness for the ERSS outcomes, though the gains are smaller in a sample where all surveyed villages lie in fully vulnerable regions.

The vulnerability-weighted measure's wider confidence band at coarse scales is itself informative. Because the continuous measure is an average over settlements, aggregation genuinely reduces its identifying variation, the widening band signals the increased imprecision. The binary indicator's band stays narrow only because re-thresholding pins its variance at every scale, while its correlation with true exposure erodes. This result is also consistent with the Berkson (1950) interpretation of our correction. Eliminating the covariances and incorrect averaging leads the exposure deviations to be closer to orthogonal to the aggregate average. In that case the coefficient estimate remains unbiased with aggregation, but loses efficiency, consistent with the widening confidence intervals reported in Figure 4.

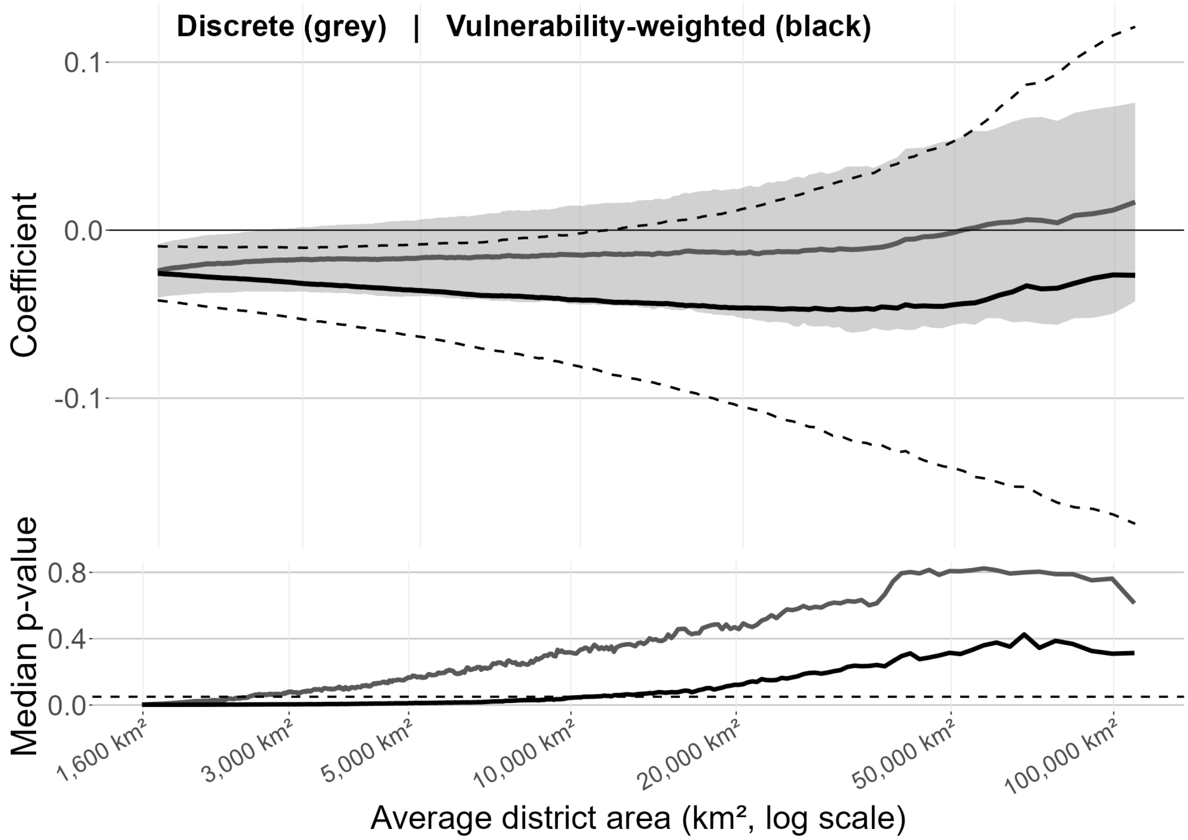


Figure 4: Aggregation sensitivity of the drought effect on log population: vulnerability-weighted versus discrete exposure. *Notes:* Rural rainfed settlements. Vulnerability-weighted continuous measure in black, based on 151 aggregation paths; standard discrete indicator in grey, based on 65 paths. Upper panel: mean coefficients; the shaded band is the mean coefficient $\pm$ 1.96 clustered standard errors (averaged across paths) for the discrete indicator, and the dashed lines mark the same interval for the vulnerability-weighted measure. Lower panel: median p-values, with the 0.05 threshold dashed.

6.3 A calibrated simulation of migration following shocks

The Ethiopian results are consistent with the mechanism in Section 5, but the empirical data still contain unobserved dynamics that explain part of the outcomes. We therefore use a calibrated simulation in which the true behavioral parameter is known. The simulation has no outcome measurement error and no confounding, so any attenuation that appears must come from how exposure is constructed.

The data-generating process is intentionally simple. A localized shock arrives with probability θ. When a location is shocked, a fraction γ of its population leaves. Movers then reallocate across unaffected destinations according to a gravity structure in population and bilateral distance.[11] The model is calibrated to the fine-scale Ethiopian estimates, so the magnitudes are empirically plausible, but the mechanism is general: it describes migration responses to any localized shock.

We model a multi-location economy in discrete time. For an origin-destination pair (j, i), bilateral inflows are

$$I_{ijt} = \frac{P_{it}}{NDP_t} \theta\gamma P_{jt} \left(1 - \frac{T_{ij}}{\sum_j T_{ij}}\right), \tag{7}$$

where P_{it} is destination population, P_{jt} is origin population, NDP_t is the total population of non-shocked destinations, and T_{ij} is the bilateral distance between i and j. The final term allocates a larger share of movers to destinations that are close relative to the full set of available destinations. Summing inflows across shocked origins and mapping the simulated population change to the regression estimand gives

$$\beta = \ln\left(\frac{(1-\gamma)P}{P+I}\right), \tag{8}$$

which links the estimated coefficient to the behavioral parameters (γ, θ). Appendix V provides the full derivation.

We take $(\hat{\beta}, \hat{\theta})$ from the fine-scale Ethiopian analysis. The drought probability is set to the realized incidence in the estimation sample, $\hat{\theta} = 0.1836$, and the departure rate to $\hat{\gamma} = 0.0188$, obtained by solving Equation (8). At these values, the model implies a fine-scale net population response of -0.023, well within one standard error of the estimates in Table 1. We initialize a distribution of settlements and populations, generate annual localized shocks, and apply the same randomized merger algorithm used in Section 4 to the simulated panel.[12]

Figure 5 reproduces the empirical pattern. As aggregation mixes shocked and non-

[11] See, among others, Gaubert (2018), McCann & Van Oort (2019), and Tombe & Zhu (2019).

[12] For computational tractability, bilateral flows are constructed among a random subset of locations, yielding over 40 million origin-destination pairs per year. Appendix V reports implementation details.

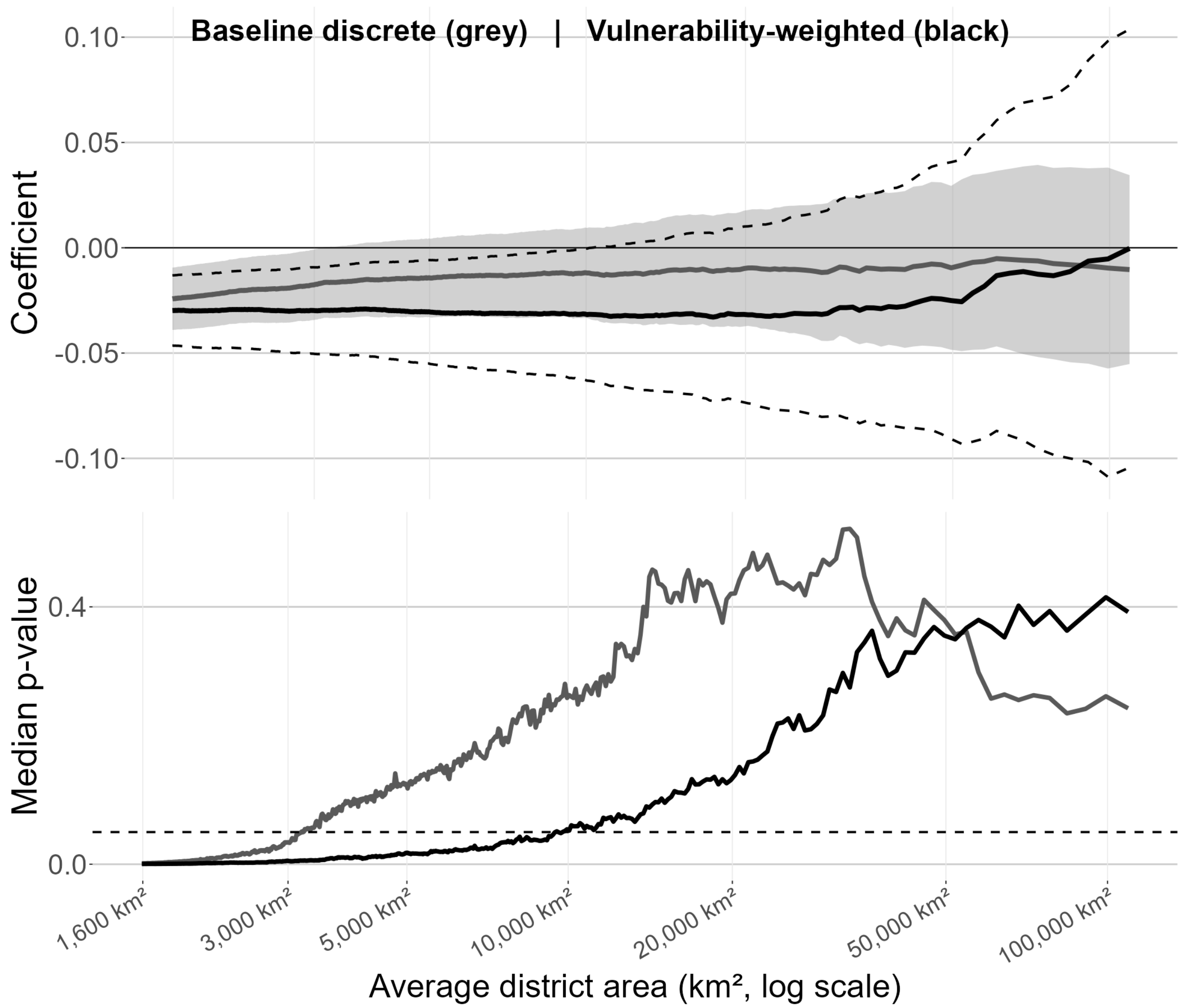


Figure 5: Aggregation sensitivity in the simulated gravity model. *Notes:* Results are based on 64 aggregation paths. Upper panel: estimated coefficients across increasing aggregation levels. The discrete indicator, shown in grey, attenuates toward zero; the vulnerability-weighted continuous measure, shown in black, remains close to the known fine-scale parameter. The shaded band is the mean coefficient ± 1.96 clustered standard errors (averaged across paths) for the discrete indicator, and the dashed lines mark the same interval for the vulnerability-weighted measure. Lower panel: median p-values, with the 0.05 threshold dashed.

shocked settlements, the discrete indicator loses identifying variation and the coefficient moves toward zero. The loss becomes pronounced once units exceed roughly 56×56 kilometers. The vulnerability-weighted continuous measure remains close to the known fine-scale parameter over a much wider range of aggregation levels, remaining stable up to roughly

28,000 km^2, with p-values below conventional thresholds up to about 9,500 km^2.

Because the true effect is known by construction, the simulation confirms the mechanism. The standard indicator fails even though the outcome is measured without error and there is no omitted-variable bias. The corrected measure performs better because it preserves the local exposure signal before aggregation.

6.4 A diagnostic for applied work

The results above translate into a test that applied researchers can run with data they already have. Exposure is constructed at the finest resolution of the shock data and then repeatedly coarsened, while outcomes, sample, and specification are held fixed. Re-estimating the main regression at each level traces a coefficient path, $\hat{\beta}(a)$, that shows how the estimate responds as exposure is measured over larger units. The exercise requires no new outcome data and no change to the estimator. It is a robustness check in the same sense as varying clustering or fixed effects, but it targets the step where, as we have shown, real effects can disappear.

The interpretation is direct. A stable path indicates that the exposure signal survives aggregation, so estimates at the available outcome scale can be trusted on this margin. Attenuation, sign changes, or collapsing precision indicate that the treatment measure, not the underlying response, is driving the result. The test remains feasible when outcomes exist only at a coarse level: coarsening exposure further above that level reveals whether the estimate is already on the unstable part of its path.

Comparing two exposure measures sharpens the test into a diagnosis. If the standard aggregated indicator fades while a locally constructed, vulnerability-weighted measure remains stable, the apparent null reflects measurement rather than the absence of an effect, and the corrected measure shows how much of the signal is recoverable. In the Ethiopian data, this comparison separates a real drought response from a false null that the standard construction produces at scales where, according to our audit, the large majority of the literature operates.

The value of the diagnostic lies in what it rules out. Without it, a null result estimated from aggregated exposure is ambiguous: it can mean that no response exists, or that the exposure measure was incapable of detecting one. With it, the two cases can be told apart at the cost of re-running one regression across aggregation levels. For research on localized shocks, and for the policies that rely on it, that distinction is the difference between concluding that people are unaffected and discovering that the measurement hid the people who are.

7 Conclusion

Weather shocks are local, but many of the outcomes used to study them are observed at coarser spatial scales. This paper shows that the way researchers bridge that mismatch can determine whether an effect is detected. In Ethiopia, drought reduces population in rural rainfed settlements by about 2.5 percent. The decline is concentrated among working-age adults, consistent with out-migration rather than mortality. The same droughts reduce food production, storage, income, and health. Yet when the same exposure is measured at coarser administrative scales, using the standard binary drought indicator common in the literature, these effects attenuate and eventually disappear.

The reason is not that the underlying response is fragile. It is that aggregation changes the treatment measure. When rainfall is averaged first and treatment is defined afterward, partly exposed areas can be classified as fully treated or untreated. When shocks and vulnerability are aggregated separately, the measure loses the local covariance between where shocks hit and where people are vulnerable to them. When vulnerability is measured on the wrong support or with the wrong weights, exposed populations can receive too little or too much weight. Fixed effects absorb average differences across units and common time shocks, but they do not recover the local interaction between a time-varying shock and local vulnerability.

The decomposition points to a simple empirical design principle: build exposure locally first and aggregate afterward. In the Ethiopian data, this means measuring local drought exposure, weighting it by dependence on rainfed agriculture, and only then aggregating to the unit used in estimation. This vulnerability-weighted measure recovers the fine-scale estimate and remains informative at spatial scales where the standard binary indicator fails. A calibrated gravity simulation, in which the true migration response is known and exposure construction is the only source of distortion, reproduces the same pattern: the standard measure attenuates, while the locally constructed measure remains stable.

These results have implications beyond Ethiopia and beyond drought. Empirical work on climate, conflict, pollution, disease, and other localized shocks increasingly relies on fine-resolution environmental data matched to coarser social and economic outcomes. That match is not innocuous. If exposure is averaged, thresholded, and assigned mechanically to large units, estimated null effects may reflect measurement rather than the absence of real responses. Scale sensitivity should therefore be treated as a core part of research design, not as a secondary robustness check.

The paper contributes in three ways. Substantively, it provides fine-scale evidence that drought induces population adjustment and welfare losses in Ethiopia. Conceptually, it explains why aggregated exposure can make these effects disappear. Practically, it provides a correction and a diagnostic that researchers can implement with existing data: construct exposure at the finest available scale, weight it by relevant vulnerability, aggregate only afterward, and examine whether estimates remain stable as exposure is coarsened. The central lesson is simple. For localized shocks, what matters is not only the shock itself, but how exposure to that shock is built.

Declarations

Declaration of competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. Computational resources were provided by SURF through the Dutch National Supercomputer Snellius. Beyond this computational support, this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability. All input data are publicly available: CHIRPS precipitation, WorldPop population and built-settlement rasters, and the Ethiopian Rural Socioeconomic Survey (World Bank LSMS). A replication package containing all code and processed data will be deposited in accordance with the journal's replication policy.

Declaration of generative AI use. During the preparation of this work the authors used a large language model (Anthropic Claude) to assist with editing, formatting, and literature search. The authors reviewed and edited all content and take full responsibility for the content of the published article.

References

Adhvaryu, A. R., Kala, N., & Nyshadham, A. (2018). The light and the heat: Productivity co-benefits of energy-saving technology. *Review of Economics and Statistics*, *102*, 779-792.

Aigner, D. J., et al. (1973). Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics*, *1*(1), 49–59.

Akyapı, B., Bellon, M., & Massetti, E. (2025). Estimating macrofiscal effects of climate shocks from billions of geospatial weather observations. *American Economic Journal: Macroeconomics*, *17*(3), 114–159.

Armand, A., Atwell, P., & Gomes, J. F. (2020). The reach of radio: Ending civil conflict through rebel demobilization. *American economic review*, *110*(5), 1395–1429.

Bai, Y., & Kung, J. K.-s. (2011). Climate shocks and sino-nomadic conflict. *Review of Economics and Statistics*, *93*(3), 970–981.

Bakkensen, L. A., & Barrage, L. (2026). Climate shocks, cyclones and economic growth: Bridging the micro-macro gap. *The Economic Journal*, *136*(673), 279–310.

Bazzi, S., & Blattman, C. (2014). Economic shocks and conflict: Evidence from commodity prices. *American Economic Journal: Macroeconomics*, *6*(4), 1–38.

Beegle, K., Dillon, A., Karlan, D., & Udry, C. (2024). Introduction to the journal of development economics special issue on methods and measurement. *Journal of Development Economics*, *170*, 103303.

Berkson, J. (1950). Are there two regressions? *Journal of the american statistical association*, *45*(250), 164–180.

Berlanda, A., Cervellati, M., Esposito, E., Rohner, D., & Sunde, U. (2024). Medication against conflict. *Journal of Development Economics*, *170*, 103306.

Boehm, C. E., Flaaen, A., & Pandalai-Nayar, N. (2019). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake. *Review of Economics and Statistics*, *101*(1), 60–75.

Bonomi, G., Gennaioli, N., & Tabellini, G. (2021). Identity, beliefs, and political conflict*. *The Quarterly Journal of Economics*, *136*(4), 2371-2411.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Burke, M., Gong, E., & Jones, K. (2015). Income shocks and hiv in africa. *The Economic Journal*, *125*(585), 1157–1189.

Cantelmo, A., Melina, G., & Papageorgiou, C. (2023). Macroeconomic outcomes in disaster-prone countries. *Journal of Development Economics*, *161*, 103037.

Caruso, G., & Miller, S. (2015). Long run effects and intergenerational transmission of natural disasters: A case study on the 1970 ancash earthquake. *Journal of development economics*, *117*, 134–150.

Carvalho, V. M., Nirei, M., Saito, Y. U., & Tahbaz-Salehi, A. (2016). Supply chain disruptions: Evidence from the great east japan earthquake. *ERN: Empirical Studies Using Input-Output Tables (Topic)*.

Casey, G. (2024). Energy efficiency and directed technical change: implications for climate change mitigation. *Review of Economic Studies*, *91*(1), 192–228.

Ciccone, A. (2011). Economic shocks and civil conflict: A comment. *American Economic Journal: Applied Economics*, *3*(4), 215–227.

Cole, S., Healy, A. J., & Werker, E. (2012). Do voters demand responsive governments? evidence from indian disaster relief . *Journal of Development Economics*, *97*, 167-181.

Conway, D., & Schipper, E. L. F. (2011). Adaptation to climate change in africa: Challenges and opportunities identified from ethiopia. *Global Environmental Change*, *21*(1), 227–237.

Corno, L., Hildebrandt, N., & Voena, A. (2020). Age of marriage, weather shocks, and the direction of marriage payments. *Econometrica*, *88*(3), 879–915.

Costinot, A., Donaldson, D., & Smith, C. (2016). Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, *124*(1), 205–248.

Czura, K., & Klonner, S. (2023). Financial market responses to a natural disaster: Evidence from credit networks and the indian ocean tsunami. *Journal of Development Economics*, *160*, 102996.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

Dell, M., Jones, B. F., & Olken, B. A. (2009). Temperature and income: Reconciling new cross-sectional and panel estimates. *American Economic Review*, *99*(2), 198–204.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, *4*, 66-95.

Del Valle, A., de Janvry, A., & Sadoulet, E. (2020). Rules for recovery: Impact of indexed disaster funds on shock coping in mexico. *American Economic Journal: Applied Economics*, *12*(4), 164–195.

Dercon, S. (2004). Growth and shocks: Evidence from rural Ethiopia. *Journal of Development Economics*, *74*(2), 309–329.

Dercon, S., Hill, R. V., Clarke, D., Outes-Leon, I., & Taffesse, A. S. (2014). Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in Ethiopia. *Journal of Development Economics*, *106*, 132–143.

Deryugina, T., Kawano, L., & Levitt, S. (2018). The economic impact of hurricane katrina on its victims: Evidence from individual tax returns. *American Economic Journal: Applied Economics*, *10*(2), 202–233.

Deryugina, T., & Molitor, D. (2018). Does when you die depend on where you live? evidence from hurricane katrina. *NBER Working Paper Series*.

Deschênes, O., & Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us. *American Economic Journal: Applied Economics*, *3*(4), 152–185.

Deschênes, O., Greenstone, M., & Guryan, J. (2009). Climate change and birth weight. *American Economic Review*, *99*(2), 211–217.

Dinkelman, T. (2017). Long-run health repercussions of drought shocks: evidence from south african homelands. *The Economic Journal*, *127*(604), 1906–1939.

Dube, O., & Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from colombia. *Review of Economic studies*, *80*(4), 1384–1421.

Fafchamps, M., Udry, C., & Czukas, K. (1998). Drought and saving in West Africa: Are livestock a buffer stock? *Journal of Development Economics*, *55*(2), 273–305.

Falcone, S., & Rosenberg, M. (2025). Agricultural modernization and redistributive conflict. *Journal of Development Economics*, 103529.

Fenske, J., & Kala, N. (2015). Climate and the slave trade. *Journal of Development Economics*, *112*, 19-32.

Fenske, J., & Kala, N. (2017). 1807: Economic shocks, conflict and the slave trade. *Journal of Development Economics*, *126*, 66–76.

Fried, S. (2018). Climate policy and innovation: A quantitative macroeconomic analysis. *American Economic Journal: Macroeconomics*, *10*(1), 90–118.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., … others (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, *2*(1), 1–21.

Gallagher, J., & Hartley, D. (2017). Household finance after a natural disaster: The case of hurricane katrina. *American Economic Journal: Economic Policy*, *9*(3), 199–228.

Gallea, Q. (2023). Weapons and war: The effect of arms transfers on internal conflict. *Journal of Development Economics*, *160*, 103001.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, *108*(11), 3117–3153.

Gehlke, C. E., & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, *29*(185), 169–170.

Gehring, K., Langlotz, S., & Kienberger, S. (2025). Stimulant or depressant? resource-related income shocks and conflict. *The Review of Economics and Statistics*, *107*(6), 1518-1536.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632–648.

Gray, C., & Mueller, V. (2012). Drought and population mobility in rural Ethiopia. *World Development*, *40*(1), 134–145.

Hanaoka, C., Shigeoka, H., & Watanabe, Y. (2018). Do risk preferences change? evidence from the great east japan earthquake. *American Economic Journal: Applied Economics*, *10*, 298-330.

Harari, M., & Ferrara, E. L. (2018). Conflict, climate, and cells: a disaggregated analysis. *Review of Economics and Statistics*, *100*(4), 594–608.

Henderson, J. V., Storeygard, A., & Deichmann, U. (2017). Has climate change driven urbanization in africa? *Journal of development economics*, *124*, 60–82.

Holland, S. P., Hughes, J. E., Knittel, C. R., & Parker, N. (2011). Some inconvenient truths about climate change policy: The distributional impacts of transportation policies. *Review of Economics and Statistics*, *97*, 1052-1069.

Hong, H., Wang, N., & Yang, J. (2023). Mitigating disaster risks in the age of climate change. *Econometrica*, *91*(5), 1763-1802.

Hull, P., & Imai, M. (2013). Economic shocks and civil conflict: Evidence from foreign interest rate movements . *Journal of Development Economics*, *103*, 77-89.

Janus, T. (2012). Natural resource extraction and civil conflict. *Journal of Development Economics*, *97*, 24-31.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of political Economy*, *114*(3), 538–575.

Jones, B. F., & Olken, B. A. (2010). Climate shocks and exports. *American Economic Review*, *100*(2), 454–459.

Josephson, A., Michler, J. D., Kilic, T., & Murray, S. (2026). The mismeasure of weather: Using earth observation data for estimation of socioeconomic outcomes. *Journal of Development Economics*, *178*, 103553.

Kassawmar, T., Zeleke, G., Bantider, A., Gessesse, G. D., & Abraha, L. (2018). A synoptic land change assessment of ethiopia's rainfed agricultural area for evidence-based agricultural ecosystem management. *Heliyon*, *4*(11), e00914.

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review*, *109*(10), 3585–3616.

Kazianga, H., & Udry, C. (2006). Consumption smoothing? livestock, insurance and drought in rural burkina faso. *Journal of Development economics*, *79*(2), 413–446.

König, M. D., Rohner, D., Thoenig, M., & Zilibotti, F. (2017). Networks in conflict: Theory and evidence from the great war of africa. *Econometrica*, *85*(4), 1093–1132.

Lagakos, D. (2020). Urban-rural gaps in the developing world: Does internal migration offer opportunities? *Journal of Economic Perspectives*, *34*(3), 174–192.

Lane, G. (2024). Adapting to climate risk with guaranteed credit: Evidence from bangladesh. *Econometrica*, *92*(2), 355–386.

Lloyd, C. T. (2016). Worldpop archive global gridded spatial datasets. version alpha 0.9. 100m nightlights v4 (tiled). *Harvard Dataverse*.

Maccini, S., & Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review*, *99*(3), 1006–1026.

Mahajan, P., & Yang, D. (2020). Taken by storm: Hurricanes, migrant networks, and us immigration. *American Economic Journal: Applied Economics*, *12*(2), 250–77.

McCann, P., & Van Oort, F. (2019). Theories of agglomeration and regional economic growth: a historical review. In *Handbook of regional growth and development theories* (pp. 6–23). Edward Elgar Publishing.

McGuirk, E. F., & Burke, M. (2017). The economic origins of conflict in africa. *Journal of Political Economy*, *128*, 3940 - 3997.

McGuirk, E. F., & Nunn, N. (2020). Transhumant pastoralism, climate change, and conflict in africa. *NBER Working Paper Series*.

McGuirk, E. F., & Nunn, N. (2024). Transhumant pastoralism, climate change, and conflict in africa. *The Review of Economic Studies*, *92*(1), 404-441.

Michler, J. D., Josephson, A., Kilic, T., & Murray, S. (2022). Privacy protection, measurement error, and the integration of remote sensing and socioeconomic survey data. *Journal of Development Economics*, *158*, 102927.

Miguel, E., & Satyanath, S. (2011). Re-examining economic shocks and civil conflict. *American Economic Journal: Applied Economics*, *3*, 228-232.

Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, *112*(4), 725–753.

Moscona, J., & Sastry, K. A. (2023). Does directed innovation mitigate climate damage? evidence from us agriculture. *The Quarterly Journal of Economics*, *138*(2), 637–701.

Mueller, V., Sheriff, G., Dou, X., & Gray, C. (2020). Temporary migration and climate variation in eastern africa. *World development*, *126*, 104704.

Nieves, J. J., Sorichetta, A., Linard, C., Bondarenko, M., Steele, J. E., Stevens, F. R., … others (2020). Annually modelling built-settlements between remotely-sensed observations using relative changes in subnational populations and lights at night. *Computers, environment and urban systems*, *80*, 101444.

Nordhaus, W. (2018). Projections and uncertainties about climate change in an era of minimal climate policies. *American economic journal: economic policy*, *10*(3), 333–360.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography*.

Pandya, S. S., & Venkatesan, R. (2016). French roast: consumer response to international conflict—evidence from supermarket scanner data. *Review of Economics and Statistics*, *98*(1), 42–56.

Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., … others (2013). A global human settlement layer from optical hr/vhr rs data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(5), 2102–2131.

Pouliot, G. A. (2023). Spatial econometrics for misaligned data. *Journal of Econometrics*, *232*(1), 168–190.

Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International journal of epidemiology*, *38*(2), 337–341.

Rohner, D., Thoenig, M., & Zilibotti, F. (2011). War signals: A theory of trade, trust and conflict. *Political Economy: Government Expenditures & Related Policies eJournal*.

Sarsons, H. (2015). Rainfall and conflict: A cautionary tale. *Journal of Development Economics*, *115*, 62–72.

Shah, M., & Steinberg, B. M. (2017). Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital. *Journal of Political Economy*, *125*(2), 527–561.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, *10*(2), e0107042.

Tombe, T., & Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, *109*(5), 1843–1872.

Traeger, C. P. (2021). Ace – analytic climate economy. *SSRN Electronic Journal*.

Voors, M., Nillesen, E. E. M., Verwimp, P., Bulte, E., Lensink, R., & van Soest, D. (2012). Violent conflict and behavior: A field experiment in burundi. *The American Economic Review*, *102*, 941-964.

Wong, D. W. (2004). The modifiable areal unit problem (maup). In *Worldminds: geographical perspectives on 100 problems: commemorating the 100th anniversary of the association of american geographers 1904–2004* (pp. 571–575). Springer.

Zappalà, G. (2024). Adapting to climate change accounting for individual beliefs. *Journal of Development Economics*.

Zhao, Y., Yang, Y., & Zhang, N. (2025). Flood shocks, heterogeneous risk exposure, and housing market dynamics in china. *Journal of Development Economics*.

Appendix

A Meta Study of Aggregation in Local Shock Literature

A.1 Motivation and purpose

This Appendix reports a structured meta study of influential empirical papers that use local shocks as identifying variation. The goal is descriptive: to document that spatial and temporal aggregation of realized shocks is common in closely related work, and therefore that the aggregation mechanism studied in this paper is plausibly relevant in standard empirical practice.

For each paper we record how exposure is constructed from underlying shock data, how vulnerability enters the design, and whether aggregation can plausibly alter the joint distribution of exposure and vulnerability within estimation units. The Appendix is not an evaluation of any individual study, and it makes no claim that aggregation necessarily induces bias in any given application.

A.2 Scope

The meta study is intentionally not an exhaustive review of all empirical work using local shocks. Instead, it targets a transparent subset of highly visible papers in which shocks are explicitly foregrounded. Many relevant papers will be missed because shocks are not referenced in the title, because they fall outside the journal set, or because they do not satisfy the influence threshold. These omissions would, if anything, strengthen the conclusion that aggregation is prevalent.

A.3 Journal universe and time window

The candidate pool is drawn from a prespecified set of journals that set widely used empirical standards in economics: American Economic Review (AER), Quarterly Journal of Economics (QJE), Journal of Political Economy (JPE), Econometrica, Review of Economic Studies (ReStud), American Economic Journal: Applied Economics (AEJ: Applied), American Economic Journal: Macroeconomics (AEJ: Macro), American Economic Journal: Economic Policy (AEJ: Policy), American Economic Journal: Microeconomics (AEJ: Micro), Review of Economics and Statistics (REStat), and Journal of Development Economics (JDE).

We restrict attention to research articles published between 2000 and 2025. The start date targets the modern period in which spatially explicit shock measures and quasi experimental designs become routine, making exposure aggregation choices central rather than incidental. The end date includes the most recent influential contributions while retaining a meaningful citation window for the influence screen. Paratext, comments, replies, and other non article content are excluded.

A.4 Search vocabulary and retrieval

We retrieve candidate papers using the OpenAlex API. For each journal, we issue a recall oriented search using a broad shock vocabulary as a retrieval hint: *rainfall, drought, flood, storm, hurricane, cyclone, precipitation, temperature, heat, earthquake, extreme weather, disaster, climate, climate shocks, conflict.* This vocabulary is not used as an inclusion filter. It is used only to construct an initial candidate pool.

A.5 Title based inclusion rule and deduplication

Inclusion is determined by a hard mechanical rule: at least one vocabulary term must appear in the paper title. This restriction ensures that shocks are a first order object rather than an auxiliary control and yields a transparent selection rule. Titles are matched using a case insensitive regular expression. Records are then deduplicated using DOIs when available and OpenAlex identifiers otherwise.

A.6 Influence screen and final sample construction

To focus on influential contributions and keep full text coding tractable, we restrict the deduplicated set to papers with at least 30 citations at the time of collection and order the remaining papers by citation count and publication year. This procedure yields a final sample of 59 papers. PDFs are downloaded programmatically when available via open access links, publisher landing pages, or DOI resolution.

A.7 Coding protocol and scope restrictions

Each paper is coded using a structured pipeline that combines automated text extraction with rule based classification via the OpenAI API. Coding records whether exposure is built by aggregating fine scale shock realizations to coarser units, how vulnerability enters the design, and whether exposure–vulnerability interactions are explicitly modeled or plausibly altered by aggregation. All classifications are descriptive. When text evidence is not found, the field is coded as unclear rather than inferred.

Uniform national or global shocks such as policies, prices, or interest rates are excluded from the substantive interpretation in this appendix because they do not involve aggregation of realized local shock intensities, and their heterogeneity operates primarily through vulnerability rather than exposure aggregation.

A.8 Outputs reported in this appendix

To preserve readability, the appendix reports a slimmed down set of tables that isolates the two pillars emphasized in the paper: shock construction and aggregation, and vulnerability construction and interaction with shocks. The summary tables below print 57 specification-level codings covering 55 of the 59 sampled papers; Dell et al. (2012) and Moscona & Sastry (2023) each appear twice because they employ two distinct analysis scales, and the remaining four papers are contained in the complete coding. All counts cited in the main text refer to these printed codings, except the sample size of 59, which refers to the full sample. The online appendix provides the full coding, including paper specific notes and extended detail.

Table 3: Shock construction and aggregation.

Citation	Unit	Local shock	Aggregation	Agg. to country	Shock level	Exposure scale
Gallea (2023)	country	yes	yes	yes	global	discrete
Berlanda et al. (2024)	country	yes	yes	yes	sub-national regions	unclear
Fenske & Kala (2017)	region	yes	yes	no	regional	discrete
Deschênes et al. (2009)	district	yes	yes	no	county	discrete
Dell et al. (2012)	district	yes	yes	no	country	continuous
Jones & Olken (2010)	country	yes	yes	yes	unclear	discrete
Ciccone (2011)	other	yes	yes	no	unclear	unclear
Lane (2024)	district	yes	yes	no	local	discrete
Falcone & Rosenberg (2025)	district	yes	yes	no	municipality	discrete
Akyapı et al. (2025)	country	yes	yes	yes	global	discrete
Armand et al. (2020)	grid cell	yes	yes	no	local	discrete
Bazzi & Blattman (2014)	country	yes	yes	yes	unclear	unclear
Burke et al. (2015)	district	yes	yes	no	local	continuous
Fenske & Kala (2015)	other	yes	yes	no	local	continuous
Deschênes & Greenstone (2011)	other	yes	yes	no	local	discrete
Bai & Kung (2011)	region	yes	yes	no	local	discrete
Dube & Vargas (2013)	district	yes	yes	no	local	continuous
Kazianga & Udry (2006)	village	yes	yes	no	local	continuous
Harari & Ferrara (2018)	grid cell	yes	yes	no	subnational	continuous
Del Valle et al. (2020)	district	yes	yes	no	local	discrete
Deryugina et al. (2018)	district	yes	yes	no	local	continuous
Moscona & Sastry (2023)	district	yes	yes	no	local	continuous
König et al. (2017)	other	yes	yes	no	unclear	unclear

Continued on next page

Citation	Unit	Local shock	Aggregation	Agg. to country	Shock level	Exposure scale
Miguel et al. (2004)	country	yes	yes	yes	country	discrete
Casey (2024)	country	yes	yes	yes	global	unclear
Costinot et al. (2016)	other	yes	yes	no	global	continuous
Czura & Klonner (2023)	other	yes	yes	no	unclear	unclear
Zhao et al. (2025)	district	yes	yes	no	local	discrete
Pandya & Venkatesan (2016)	other	yes	yes	no	unclear	discrete
Fried (2018)	other	yes	yes	no	unclear	continuous
Henderson et al. (2017)	district	yes	yes	no	Sub-Saharan Africa	discrete
Gallagher & Hartley (2017)	other	yes	yes	no	local	discrete
Boehm et al. (2019)	other	yes	no	unclear	local	continuous
Caruso & Miller (2015)	district	yes	yes	no	large area of Peru	discrete
Cantelmo et al. (2023)	country	yes	yes	yes	unclear	continuous
Nordhaus (2018)	country	yes	yes	yes	unclear	continuous
Sarsons (2015)	district	yes	yes	no	district	continuous
Gehring et al. (2025)	district	yes	yes	no	district	discrete
Holland et al. (2011)	district	yes	yes	no	district	discrete
Carvalho et al. (2016)	other	yes	yes	no	regional	continuous
Dell et al. (2012)	country	yes	yes	yes	country	continuous
McGuirk & Nunn (2024)	grid cell	yes	yes	no	subnational	continuous
Adhvaryu et al. (2018)	other	yes	yes	no	local	continuous
McGuirk & Burke (2017)	grid cell	yes	yes	no	local	continuous
Rohner et al. (2011)	other	yes	no	unclear	unclear	discrete
Miguel & Satyanath (2011)	country	no	no	unclear	unclear	unclear
Voors et al. (2012)	other	no	no	unclear	unclear	unclear
Zappalà (2024)	other	yes	no	unclear	unclear	unclear

Continued on next page

Citation	Unit	Local shock	Aggregation	Agg. to country	Shock level	Exposure scale
Shah & Steinberg (2017)	district	yes	no	unclear	local	discrete
Moscona & Sastry (2023)	other	no	yes	no	unclear	unclear
Hull & Imai (2013)	country	yes	no	unclear	unclear	continuous
Hanaoka et al. (2018)	other	yes	no	unclear	local	continuous
Bonomi et al. (2021)	other	no	no	unclear	unclear	unclear
Mahajan & Yang (2020)	country	no	no	unclear	unclear	unclear
Janus (2012)	other	no	no	unclear	unclear	unclear
Hong et al. (2023)	other	no	no	unclear	unclear	unclear
Traeger (2021)	region	no	no	unclear	unclear	unclear

Table 4: Vulnerability construction and exposure–vulnerability interaction.

Citation	Vulnerability observed	Vulnerability scale	Shock × Vulnerability	Relevance	In scope
Gallea (2023)	no	not applicable	yes	direct	yes
Berlanda et al. (2024)	yes	unclear	yes	indirect	yes
Fenske & Kala (2017)	yes	continuous	yes	direct	yes
Deschênes et al. (2009)	yes	continuous	no	direct	yes
Dell et al. (2012)	yes	continuous	no	direct	yes
Jones & Olken (2010)	yes	unclear	yes	indirect	yes
Ciccone (2011)	no	not applicable	no	none	yes
Lane (2024)	yes	continuous	no	direct	yes
Falcone & Rosenberg (2025)	yes	continuous	no	direct	yes
Akyapı et al. (2025)	yes	continuous	no	direct	yes
Armand et al. (2020)	no	not applicable	yes	direct	yes
Bazzi & Blattman (2014)	yes	unclear	no	indirect	yes
Burke et al. (2015)	yes	continuous	no	direct	yes
Fenske & Kala (2015)	yes	continuous	yes	direct	yes
Deschênes & Greenstone (2011)	yes	discrete	yes	direct	yes
Bai & Kung (2011)	yes	continuous	yes	direct	yes
Dube & Vargas (2013)	yes	unclear	yes	direct	yes
Kazianga & Udry (2006)	yes	continuous	yes	direct	yes
Harari & Ferrara (2018)	yes	continuous	no	direct	yes
Del Valle et al. (2020)	yes	continuous	yes	direct	yes
Deryugina et al. (2018)	yes	discrete	yes	direct	yes
Moscona & Sastry (2023)	yes	continuous	yes	direct	yes
König et al. (2017)	no	not applicable	no	direct	yes

Continued on next page

Citation	Vulnerability observed	Vulnerability scale	Shock × Vulnerability	Relevance	In scope
Miguel et al. (2004)	yes	discrete	no	direct	yes
Casey (2024)	no	not applicable	yes	indirect	yes
Costinot et al. (2016)	yes	continuous	yes	direct	yes
Czura & Klonner (2023)	yes	unclear	yes	direct	yes
Zhao et al. (2025)	yes	continuous	yes	direct	yes
Pandya & Venkatesan (2016)	yes	continuous	yes	indirect	yes
Fried (2018)	no	not applicable	yes	indirect	yes
Henderson et al. (2017)	yes	continuous	yes	direct	yes
Gallagher & Hartley (2017)	yes	continuous	yes	direct	yes
Boehm et al. (2019)	yes	continuous	no	none	yes
Caruso & Miller (2015)	yes	continuous	yes	direct	yes
Cantelmo et al. (2023)	yes	unclear	yes	direct	yes
Nordhaus (2018)	no	not applicable	no	indirect	yes
Sarsons (2015)	yes	continuous	yes	direct	yes
Gehring et al. (2025)	yes	continuous	yes	direct	yes
Holland et al. (2011)	no	not applicable	yes	direct	yes
Carvalho et al. (2016)	yes	continuous	yes	direct	yes
Dell et al. (2012)	yes	continuous	no	direct	yes
McGuirk & Nunn (2024)	yes	continuous	yes	direct	yes
Adhvaryu et al. (2018)	yes	continuous	yes	direct	yes
McGuirk & Burke (2017)	yes	continuous	yes	direct	yes
Rohner et al. (2011)	no	not applicable	no	none	yes
Miguel & Satyanath (2011)	no	not applicable	no	none	no
Voors et al. (2012)	no	not applicable	no	none	no
Zappalà (2024)	yes	unclear	yes	direct	no

Continued on next page

Citation	Vulnerability observed	Vulnerability scale	Shock × Vulnerability	Relevance	In scope
Shah & Steinberg (2017)	no	not applicable	no	none	no
Moscona & Sastry (2023)	yes	unclear	yes	indirect	no
Hull & Imai (2013)	yes	unclear	yes	direct	no
Hanaoka et al. (2018)	no	not applicable	yes	indirect	no
Bonomi et al. (2021)	no	not applicable	no	none	no
Mahajan & Yang (2020)	no	not applicable	yes	indirect	no
Janus (2012)	yes	unclear	yes	indirect	no
Hong et al. (2023)	no	not applicable	no	none	no
Traeger (2021)	no	not applicable	no	none	no

Meta Study Repository and Coding Protocol

The papers listed above form the fixed repository for our meta study. The repository is generated by a reproducible search procedure that restricts attention to a prespecified set of journals and publication years and requires explicit local shock terminology in the paper title. This construction is intended to minimize discretionary selection while capturing a visible and influential subset of empirical work that uses shocks as identifying variation.

We analyzed and coded each paper in full. For each study we record the primary shock, the spatial unit at which exposure is defined, and the exposure construction pipeline, with particular attention to whether fine scale primitives are aggregated to coarser units and whether additional transformations are applied such as thresholds, binning, or temporal pooling. We also code the role of vulnerability, distinguishing measured vulnerability proxies from implicit vulnerability captured by fixed effects or controls, recording the scale at which vulnerability is defined, and whether vulnerability is discrete or continuous. Finally, we code whether exposure–vulnerability interactions are explicitly modeled, and whether interactions are formed before or after aggregation. The coding is descriptive and does not assess whether any individual estimate is biased or invalid.

To preserve space, the printed appendix reports a slimmed down set of tables that isolates the two pillars emphasized in the paper: exposure construction and aggregation, and vulnerability construction and exposure–vulnerability interaction. The online appendix provides the full coding, including paper specific descriptions of the shock and aggregation steps, the operationalization of vulnerability, and a qualitative relevance rationale.

Relevance Classification and Table Links

Using the coding, we classify papers into four relevance categories: *Very High*, *High*, *Medium*, and *Low*. The classification is structural rather than evaluative. It summarizes how closely a paper matches the mechanism studied here through exposure aggregation, within unit vulnerability variation, and the implied scope for exposure–vulnerability covariance effects.

A paper is classified as *Very High Relevance* when exposure is constructed by aggregating fine scale local shock realizations to relatively coarse units, vulnerability plausibly varies within those units but is represented coarsely or implicitly, and exposure–vulnerability interactions are not modeled prior to aggregation. The remaining categories relax one or more of these features, for example through less aggregation, finer vulnerability measurement, or design elements that partially mitigate aggregation concerns.

Tables 5–8 report the classification. Direct links are Very High Relevance Table, High Relevance Table, Medium Relevance Table, and Low Relevance Table.

Table 5: Very high relevance papers.

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Dell et al. (2012)	Temperature and precipitation	Grid cell to country year	Population weighted country-year averages of gridded climate variables	High	Poor country indicator plus fixed effects and macro controls	High	No	Local nonlinear climate realizations are aggregated to national means while vulnerability is coarse, making both pillars and their covariance plausibly salient
Dell et al. (2009)	Temperature and precipitation	Grid cell to municipality to country	Fine climate inputs aggregated using population weights to municipal and national panels	High	Development status and income group heterogeneity plus fixed effects	High	No	Multi-layer aggregation of exposure smooths local extremes, and vulnerability is represented coarsely relative to within-unit exposure variation

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Deschênes et al. (2009)	Gestational temperature extremes	Station to county trimester	Inverse-distance-weighted county temperature, discretized into bin counts by trimester	High	Demographic subgroup vulnerability and timing, largely via fixed effects and subgroup analyses	High	No	Exposure is averaged and discretized before temporal aggregation, while vulnerability differs across subgroups and likely covaries with exposure within counties
Jones & Olken (2010)	Temperature and precipitation	Grid cell to country year	Population weighted country-year climate averages used in export regressions	High	Poor country status and fixed effects	High	No	Country averaging compresses heterogeneous local exposure, while vulnerability is coarse and plausibly aligned with exposure through economic geography
Miguel et al. (2004)	Rainfall shocks	Grid cell to country year	Country-year rainfall averages from coarse gridded precipitation	High	Vulnerability implicit via fixed effects and macro structure	High	No	National averaging removes within-country alignment between exposure and vulnerability that plausibly drives conflict responses

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Burke et al. (2015)	Extreme heat and precipitation	Grid cell to county	County aggregated degree-day measures and seasonal summaries from fine weather data	High	Vulnerability implicit via fixed effects and long differences	Medium–High	No	Nonlinear exposure is summarized at county level, and unmeasured within-county vulnerability can align with where extreme heat occurs
Shah & Steinberg (2017)	Rainfall shocks	Grid cell to district year	District rainfall coded into discrete percentile shocks from aggregated series	High	Vulnerability implicit via age timing and cohort exposure	High	No	Exposure is discretized after aggregation and vulnerability is not directly measured, making both pillars and their interaction plausibly important
Sarsons (2015)	Rainfall	Grid cell to district year	District rainfall deviations and percentile categories built from gridded precipitation	High	Irrigation exposure as a coarse vulnerability proxy plus fixed effects	High	No	District aggregation smooths localized rainfall while irrigation vulnerability is coarse and likely spatially correlated with rainfall regimes

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
McGuirk & Nunn (2020)	Rainfall scarcity	Grid cell to ethnic territory year	Annual rainfall averaged over large ethnic territories	High	Pastoralism-based vulnerability at group level plus fixed effects	High	No	Large-territory averaging compresses local drought variation while group vulnerability is coarse, with strong plausibility of within-territory sorting
Bai & Kung (2011)	Drought and flood disasters	Region decade	Decadal regional frequency measures of disasters	High	Vulnerability implicit via economic structure and geography	High	No	Extreme spatial and temporal aggregation collapses localized shocks and vulnerability heterogeneity likely covaries with exposure across vast areas
Mahajan & Yang (2020)	Hurricane exposure	Grid cell to country year	Country-year hurricane exposure index averaging wind intensity across populated locations	High	Migrant network stocks at country level	High	No	Local hurricane intensity is averaged into national exposure while vulnerability is proxied at coarse scale, with strong scope for covariance

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Zappalà (2024)	Drought SPEI	Grid cell to union season	SPEI averaged to union-season indices	High	Individual beliefs measured at household level	High	No	Exposure is aggregated while vulnerability is fine, creating a clear case where aggregation can mask the joint distribution of exposure and vulnerability
Henderson et al. (2017)	Moisture index	Grid cell to district	Spatially averaged district moisture plus temporal smoothing	High	Industrial capacity and baseline aridity proxies	High	No	District aggregation smooths local shocks central to migration, while vulnerability proxies operate at coarse scale and likely align with exposure
Harari & Ferrara (2018)	Drought SPEI	Cell year	Cell-level SPEI aggregated over growing season with discretization	Medium–High	Vulnerability absorbed via fixed effects and geography	Medium–High	No	More disaggregated than country panels but still aggregates in space and time, leaving scope for exposure–vulnerability covariance loss

Table 6: High relevance papers.

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Akyapı et al. (2025)	Extreme temperature and drought	Grid cell to country year	Country-year aggregates of extreme moments built from dense local weather observations	High	Country characteristics and fixed effects	Medium–High	No	Upstream aggregation remains central despite rich primitives, but vulnerability is mostly coarse at country level
Cole et al. (2012)	Monsoon rainfall and disaster relief	Grid cell to district; relief at state year	District seasonal rainfall aggregates from gridded data combined with state-year relief spending	High	Agricultural share and literacy proxies	High	No	Both exposure and vulnerability are coarse relative to village conditions, with plausible within-district sorting
Del Valle et al. (2020)	Extreme rainfall triggering transfers	Station to municipality	Municipal rainfall index and threshold-based assignment from station data	High	Municipal resilience and infrastructure proxies	Medium–High	No	Aggregation exists but the design partially mitigates concerns via a sharp policy threshold, leaving remaining within-municipality heterogeneity

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Deryugina & Molitor (2018)	Hurricane disaster	City-wide treated cohort	Binary city-wide exposure with intensity analyzed post hoc	High	Household and demographic vulnerability measures	High	No	Coarse treatment collapses local intensity, while vulnerability heterogeneity is rich and plausibly aligned with intensity within the city
Gallagher & Hartley (2017)	Flood depth	Flood maps to census block	Block-mean flood depth discretized into bins	High	Household financial variables and insurance status	High	No	Exposure is aggregated and discretized while vulnerability is fine, creating strong scope for within-block heterogeneity
Czura & Klonner (2023)	Tsunami intensity	Survey points to branch area	Branch-level tsunami intensity assigned from sparse geophysical points	High	Occupational vulnerability categories	High	No	Strong upstream aggregation of intensity with coarse vulnerability, likely correlated within coastal areas

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Zhao et al. (2025)	Flood events	District event indicator	Binary district flood event measure	High	Compound elevation and topography within districts	High	Yes	Aggregated exposure paired with fine vulnerability and explicit interaction, partially mitigating but not eliminating heterogeneity concerns
Caruso & Miller (2015)	Earthquake intensity	Department	Department-level Mercalli intensity scores, sometimes binned	High	Cohort timing and demographic characteristics	High	No	Regional exposure averages collapse localized destruction while vulnerability varies across cohorts and locations
Cantelmo et al. (2023)	Natural disasters	Country year	Country-year disaster counts and damage measures from global databases	High	Vulnerability embedded structurally, not measured locally	Medium–High	No	National aggregates mechanically compress local disasters and obscure joint exposure–vulnerability distributions

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Carvalho et al. (2016)	Disaster origin exposure	Municipality disaster area	Coarse disaster-area indicator at municipality of origin	Medium–High	Network position and firm characteristics	Medium	No	Initial spatial discretization collapses intensity heterogeneity, while vulnerability is network based rather than spatial
Moscona & Sastry (2023)	Extreme heat	Grid cell to county and crop national	County crop decade degree days combined with national crop exposure weights	High	Crop tolerance and planting patterns	High	Yes	Explicit interaction and layered aggregation of both exposure and vulnerability channels
Costinot et al. (2016)	Climate-driven yield shifts	Grid cell to country crop	Micro yield predictions aggregated to country-crop productivity	High	Vulnerability implicit through productivity structure	Medium	No	Strong aggregation pipeline for shocks, with vulnerability less explicit but aligned with the mechanism

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Dube & Vargas (2013)	Commodity prices	Uniform by design, interacted locally	National or global price series interacted with local production	Medium–High	Municipality production intensity and crop presence	High	Yes	No spatial intensity variation, but heterogeneous exposure and vulnerability are interacted, with plausible covariance across locations
Bazzi & Blattman (2014)	Commodity export prices	Country year	National export price index weighted by export shares	Medium–High	Institutions and regime type proxies	Medium–High	No	Aggregation is inherent to the national shock object, with likely within-country exposure heterogeneity
Gehring et al. (2025)	International opium prices	Uniform by design interacted with suitability	Global price series scaled by local suitability and market access	Medium	Suitability and market access proxies	Medium–High	Yes	Uniform shock with heterogeneous exposure mapping, creating aggregation concerns in both exposure and vulnerability

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
McGuirk & Nunn (2024)	Global food prices	Uniform by design mapped to cells	Global prices scaled by local crop structure	Medium	Producer–consumer structure at cell level	High	Yes	Uniform shock but explicit local exposure mapping, strong relevance to vulnerability and covariance channels
Deschênes & Greenstone (2011)	Temperature bins and precipitation	Station to county and state	County weather aggregation and binning, with state aggregation for some outcomes	High	Age group heterogeneity and fixed effects	Medium–High	No	Strong exposure aggregation and discretization, with vulnerability largely implicit
König et al. (2017)	Rainfall instruments	Grid cell to group homeland	Rainfall averaged over large group territory hulls	High	Vulnerability absorbed via fixed effects	Medium–High	No	Clear upstream spatial averaging of a local climate field over heterogeneous areas
Miguel & Satyanath (2011)	Rainfall shocks	Grid cell to country year	National rainfall averages and growth rates	High	Vulnerability implicit via macro structure and fixed effects	High	No	Canonical IV design with aggregation central and vulnerability pillar largely unmeasured

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Falcone & Rosenberg (2025)	Agricultural modernization shocks	Fine biophysical inputs to municipality	Municipality averages of suitability gains interacted with post period	High	Land inequality and indigenous presence proxies	High	Yes	Suitability is aggregated and then interacted with vulnerability proxies, with strong plausibility of within-municipality sorting

Table 7: Medium relevance papers.

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Berlanda et al. (2024)	ART availability scaled by HIV prevalence	Country–year and region–year	National and regional treatment coverage aggregated from individual data	Medium	Pre-treatment HIV prevalence aggregated to region or country	Medium–High	Yes	Both exposure and vulnerability are aggregated but explicitly interacted, reducing but not eliminating scope for loss of joint variation
Kazianga & Udry (2006)	Rainfall deviations	Village–year	Village-level rainfall aggregates from local measurements	Medium	Household assets and land characteristics at micro level	Medium	Yes	Exposure is aggregated at village level rather than plot level, but vulnerability is rich and interaction is explicit
Voors et al. (2012)	Conflict exposure	Community and household	Community-level conflict intensity aggregated over events and time	Medium	Household-level victimization and baseline characteristics	Medium	No	Aggregation occurs over time and space, but the shock is not a spatial physical field and vulnerability is relatively well measured

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Armand et al. (2020)	Radio exposure and income shocks	Grid cell with auxiliary aggregation	Mixed: radio intensity at cell level, auxiliary shocks aggregated	Medium	Crop exposure and income channels proxied at cell level	Medium	Yes	Aggregation exists mainly for auxiliary shocks, while the primary treatment preserves fine spatial variation
Fenske & Kala (2015)	Temperature anomalies	Coarse grid to port and region	Paleoclimate reconstructions at very coarse spatial resolution	Medium–High	Agro-ecological zones and crop environment proxies	Medium	No	Exposure is inherently coarse due to historical data limits; aggregation is present but not a modeling choice
Pandya & Venkatesan (2016)	Diplomatic conflict and media salience	Uniform by design	Uniform national shock with local variation only in salience	Medium	Local consumer identity composition proxies	Medium	Yes	Aggregation operates through exposure and vulnerability proxies rather than spatial averaging of a local shock field

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Hull & Imai (2013)	Foreign interest rate movements	Uniform by design	National exposure to global interest rates	Medium	Ethnolinguistic fractionalization and macro controls	Medium	Yes	Uniform shocks generate heterogeneous exposure through national structure, but spatial aggregation of local shocks is not central
Holland et al. (2011)	Fuel price and subsidy policies	Uniform by design	National policy shocks mapped to local outcomes	Medium	County-level production and consumption structure	Medium	No	Aggregation is inherent to the policy simulation and not driven by spatial smoothing of a physical shock
Gallea (2023)	Arms imports	Country–year	Country-level arms import totals constructed via network weights	Low–Medium	No explicit vulnerability; heterogeneity absorbed by controls	Low–Medium	No	Aggregation reflects network construction rather than spatial smoothing of local shocks
Harari & Ferrara (2018)	Drought SPEI	Cell–year	Cell-level SPEI aggregated over growing season months	Medium	Vulnerability absorbed via fixed effects and geography	Medium	No	More disaggregated than country studies, but still aggregates temporally and within cells

Continued on next page

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Ciccone (2011)	Rainfall measurement	Country–year	National rainfall averages from gridded data	Medium	Vulnerability not explicitly constructed	Medium	No	Included as a methodological reference highlighting sensitivity to aggregated rainfall measures rather than as a primary empirical case

Table 8: Low relevance papers.

Citation	Primary shock	Shock spatial unit	Exposure construction	Likely exposure heterogeneity	Vulnerability construction	Likely exposure–vulnerability covariance	Interaction explicitly modeled	Relevance rationale
Boehm et al. (2019)	Supply-chain disruptions	Disaster origin aggregated to firm-level shocks	Firm-level import contraction measures	Medium	Firm dependence on Japanese inputs	Medium	No	Aggregation is upstream and indirect, not a spatial averaging of local exposure relevant for migration or welfare
Costinot et al. (2016)	Climate-driven yield changes	Grid cell to country–crop	Micro yield predictions aggregated to country–crop productivity	Medium	Heterogeneity treated as productivity rather than vulnerability	Low–Medium	No	Aggregation is explicit and disciplined, but migration-relevant vulnerability and local tail exposure are abstracted from

B Rainfall Data and Drought Identification

We construct drought identifiers using the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset (Funk et al., 2015). CHIRPS combines satellite-based precipitation estimates from TRMM and GPM with gauge observations from global station networks. Geostatistical interpolation merges these sources to produce quasi-global precipitation fields at approximately 5.5×5.5 km resolution.

We extract monthly rainfall for Ethiopia over 1981–2016 and overlay the rasters with district boundaries from openAFRICA. When raster tiles span multiple administrative units, rainfall is assigned proportionally to the area overlap so that total precipitation is preserved.

There is no consensus on the optimal meteorological measure of drought. Some indices incorporate evapotranspiration, soil moisture, or vegetation stress (e.g. CDI, SPEI), but most economic applications use nominal rainfall. To match this literature, we aggregate rainfall to the annual level, fit a gamma distribution to each district's historical rainfall series, and classify a year as a drought if annual rainfall falls below the twentieth percentile of the fitted distribution. As shown in a catalog of leading studies (Corno et al., 2020; Jayachandran, 2006; Burke et al., 2015; Maccini & Yang, 2009; Miguel et al., 2004), roughly 80 percent of papers aggregate annually, and more than half use a gamma distribution and a twentieth percentile cutoff.

Appendix D shows that these district-level drought shocks display strong positive spatial autocorrelation (average Moran's $I = 0.6494$), implying that localized clusters of drought are empirically common.

To ensure that rainfall variation maps to meaningful agricultural risk, we restrict our main analysis to rainfed regions, where approximately 95 percent of agricultural water originates directly from rainfall (Kassawmar et al., 2018). Appendix C provides the classification procedure.

C Classification of Rainfed Regions

Following Kassawmar et al. (2018), we classify locations as rainfed if more than 95 percent of agricultural water input derives from direct precipitation rather than irrigation or groundwater. This classification combines maps of irrigation infrastructure, topography, river basin hydrology, and agricultural system characteristics. The majority of rural Ethiopia falls under this category. We restrict our baseline sample accordingly.

D Spatial Autocorrelation in Shocks

This appendix documents the spatial dependence of the district level drought indicator used in the main analysis. We compute Moran's I for each year between 2000 and 2016 using a standard contiguity based spatial weights matrix at the district level. The results show strong and highly significant positive spatial autocorrelation. Droughts tend to cluster geographically rather than occurring as isolated point shocks, which is important for understanding how aggregation mixes treated and untreated locations and for interpreting the simulations in Section 6. The years 2007, 2008, 2012, and 2014 do not appear in the table because no rainfed district was in (lagged) drought in those years, so the statistic is not defined for the drought indicator.

E Settlement Boundary Construction

Settlement boundaries are constructed from WorldPop's Built-Settlement Growth Model (BSGM, version v0a), which provides annual 100×100 meter binary settlement grids, interpolated between remotely sensed observations for 2001–2014 (BSGMi) and extrapolated for later years (BSGMe) by integrating land-cover classifications with subnational population and nighttime-lights data (Nieves et al., 2020). Built-up structures are defined following Pe-

Table 9: Spatial autocorrelation of district level drought shocks in Ethiopia between 2000 and 2016, identified by Moran's I.

	Year	Moran I statistic standard deviate	p-value	Moran I statistic	Expectation	Variance	Alternative hypothesis
1	2001	27.7387	1.192567e-169	0.6705	-0.0016	0.0006	greater
2	2002	25.8831	5.166891e-148	0.6173	-0.0016	0.0006	greater
3	2003	26.2493	3.642580e-152	0.6322	-0.0016	0.0006	greater
4	2004	26.5196	2.882103e-155	0.6376	-0.0016	0.0006	greater
5	2005	25.1126	1.812894e-139	0.6047	-0.0016	0.0006	greater
6	2006	22.6302	1.094106e-113	0.5402	-0.0016	0.0006	greater
7	2009	23.1857	3.177519e-119	0.5511	-0.0016	0.0006	greater
8	2010	25.9995	2.511006e-149	0.6233	-0.0016	0.0006	greater
9	2011	29.6730	8.574227e-194	0.6969	-0.0016	0.0006	greater
10	2013	22.0282	7.738195e-108	0.5225	-0.0016	0.0006	greater
11	2015	34.9465	7.310424e-268	0.8329	-0.0016	0.0006	greater
12	2016	36.1216	5.202922e-286	0.8633	-0.0016	0.0006	greater
13	**Means**	**27.1740**	**6.448505e-109**	**0.6494**	**-0.0016**	**0.0006**	**greater**

Notes: The years 2007, 2008, 2012, and 2014 are omitted because no rainfed district was in (lagged) drought in those years, so Moran's I is undefined for the drought indicator.

saresi et al. (2013) as enclosed constructions above ground intended for shelter or economic production.

The construction proceeds in three steps:

1. **Buffering:** Each built-up grid cell is surrounded by a 500 meter buffer to connect proximate structures and capture low-density settlement sprawl.

2. **Clustering:** Overlapping buffer zones are merged into continuous polygons, which constitute settlement boundaries.

3. **Longitudinal ID assignment:** When previously distinct settlements merge over time, a unified identifier is assigned to maintain longitudinal consistency, generating an unbalanced panel of settlement boundaries from 2000 to 2016.

This procedure yields 13,376 unique settlements. Visual checks and manual overlays verify that settlement clusters align with known urban and rural areas.

F Population Data Construction

Population estimates come from the WorldPop project (Lloyd, 2016), which uses machine-learning (random forest) methods to disaggregate census counts into fine-resolution grid cells (Stevens et al., 2015). For Ethiopia, the 2007 Population and Housing Census provides the benchmark. Predictive covariates include:

- built-up area and settlement extent,
- land cover classifications,
- road networks,
- building footprints,
- satellite night lights,
- topography and vegetation indices,
- service locations (schools, health centers).

Validation exercises for nearby countries (notably Uganda) report precision rates above 90 percent. Annual rasters for 2000–2016 are overlaid on the settlement polygons to obtain total population counts for each settlement-year observation.

Some of these predictive covariates, in particular vegetation indices and satellite night lights, can respond to weather conditions directly, raising the possibility that drought lowers predicted population mechanically rather than through actual population change. Section 2 discusses why this channel is unlikely to explain the baseline estimates: meteorological drought, and any vegetation response to it, occurs in rainfed and non-rainfed districts alike while the population response appears only in rainfed districts; the ERSS-based welfare and demographic results do not rely on WorldPop; and the implied departure rates match survey-based migration estimates for rural Ethiopia.

G Proving migration as the primary channel of demographic change in the baseline regression

A natural concern with using population change as evidence of migration is that drought may also affect fertility or mortality. If these margins respond meaningfully to shocks, the coefficient in our baseline specification could reflect demographic adjustment rather than population movement. We use the Ethiopian Rural Socioeconomic Survey (ERSS) to unpack this mechanism and measure how the demographic composition of settlements changes after drought exposure.

We use georeferenced ERSS households matched to settlement boundaries to construct age-specific population ratios for children (0–14), working-age adults (15–59), and elderly individuals (60+). Each variable is defined as the share of the corresponding group in the total settlement population, so all demographic outcomes are ratios rather than absolute counts. Appendix AB reports the full set of TWFE estimates. The results are straightforward. Following a drought, the share of children and elderly increases, while the share of working-age adults declines. The effect on the working-age group is large, negative, and statistically significant, whereas the other two groups move in the opposite direction. We also observe that the probability of being in a committed relationship rises after drought exposure, consistent with younger singles being more likely to leave. None of these patterns is consistent with fertility or mortality being the primary channel. Instead, the demographic shifts align with selective out-migration of mobile working-age individuals.

Having established that migration is the main driver of population change, we examine whether cross district spillovers contaminate the baseline estimates. If migrants relocate into neighboring, unaffected districts, those districts would experience population inflows that are mechanically correlated with drought elsewhere, which could bias the coefficient toward zero. The literature finds that most rural to rural migration in Sub Saharan Africa occurs within a radius of roughly 50 kilometers. Using this distance rule, we classify all rural settlements in non drought districts as *adjacent* or *non adjacent* to drought areas and compare their post shock population growth. The TWFE results, summarized in Appendix L, show no difference between the two groups (p value ≈ 0.35). This indicates that spillovers of rural migrants into neighboring districts are not driving our baseline migration effect.

The urban case is less clear. Because the ERSS does not allow us to distinguish between within district and cross district rural to urban migration, or between different types of urban to urban flows, any migration driven spillovers into urban areas are absorbed at the district level. As a result, the estimated drought coefficients for urban settlements (Appendix L) should be interpreted conservatively. In particular, upward bias cannot be ruled out for urban settlements that receive migrants from drought affected neighboring districts.[13]

Taken together, the ERSS evidence confirms that the baseline population response to drought reflects migration dynamics rather than demographic shifts. Spillovers do not materially affect the rural estimates, and the demographic composition of settlements evolves in a pattern that is consistent with selective out migration by working age individuals.

H ERSS Survey Design and Variables

The Ethiopian Rural Socioeconomic Survey (ERSS) is a panel survey implemented jointly by the Central Statistics Agency and the World Bank LSMS-ISA program. It covers three waves (2011/12, 2013/14, 2014/15) and is representative of rural town areas, defined as towns with fewer than 10,000 inhabitants.

Sampling follows a two-stage probability design: enumeration areas (EAs) are selected with probability proportional to size, and households are sampled within each EA. The first wave includes 3,996 households; later waves expand to more than 5,000 households across

[13]The lack of high frequency, geocoded urban movement data prevents us from decomposing the various migration channels in urban areas.

433 settlements.

The ERSS comprises household, agricultural, post-planting, post-harvest, and community surveys covering hundreds of micro-variables. We focus on ten georeferenced outcomes that capture production shocks, consumption, nutrition, income, assets, and health:

- self-reported productivity shock,
- crop damage,
- adverse shock to food production,
- adverse shock to food stocks,
- harvest in storage (kg),
- cereal intake (past seven days),
- meat intake (past seven days),
- illness in past two months,
- adverse shock to income,
- adverse shock to assets.

Geographic coordinates allow these outcomes to be linked precisely to rainfall and settlement data.

I Two Way Fixed Effects Weights

This appendix reports the implied weights in the two way fixed effects estimator following De Chaisemartin & d'Haultfoeuille (2020). The TWFE coefficient can be expressed as a weighted average of group time specific treatment effects. Negative weights can generate unintuitive estimates if treatment effects are heterogeneous.

In all relevant specifications the weights are almost entirely positive. This supports the interpretation of the reported TWFE coefficients as meaningful average treatment effects rather than arbitrary weighted averages with large sign reversals.

I.1 Baseline model, all rural settlements

Under the common trends assumption, the baseline TWFE coefficient $\beta = -0.0263$ is a weighted sum of 1794 ATT terms. Almost all weights are positive.

Treat. var: DROUGHT_l1	**ATT's**	**$\sum$ weights**
Positive weights	1791	1.0001
Negative weights	3	-0.0001
Total	1794	1

I.2 Baseline model, rural settlements in rainfed districts

For the restricted sample of rural rainfed settlements, the TWFE coefficient $\beta = -0.0161$ averages 1772 ATT terms with almost all mass on positive weights.

Treat. var: DROUGHT_l1	**ATT's**	**$\sum$ weights**
Positive weights	1768	1.0002
Negative weights	4	-0.0002
Total	1772	1

I.3 Baseline model, rural settlements in non rainfed districts

For rural non rainfed settlements, the TWFE coefficient $\beta = -0.0193$ is a weighted sum of 603 ATT terms, all with positive weights.

Treat. var: DROUGHT_l1	**ATT's**	**$\sum$ weights**
Positive weights	603	1
Negative weights	0	0
Total	603	1

I.4 Baseline model, urban settlements in rainfed districts

For urban rainfed settlements, the TWFE coefficient $\beta = -0.0248$ is a weighted sum of 132 ATT terms, all with positive weights.

Treat. var: DROUGHT_l1	**ATT's**	**$\sum$ weights**
Positive weights	132	1
Negative weights	0	0
Total	132	1

I.5 Baseline model, urban settlements in non rainfed districts

For urban non rainfed settlements, the TWFE coefficient $\beta = -0.0031$ averages 9 ATT terms, all positively weighted.

Treat. var: DROUGHT_l1	**ATT's**	**$\sum$ weights**
Positive weights	9	1
Negative weights	0	0
Total	9	1

J Spatial Control Variables

This appendix describes additional spatial controls that are used in robustness specifications of the baseline two way fixed effects model. These variables absorb variation related to borders, neighborhood drought density, repeated shocks, and district size. The goal is to verify that the main migration response is not an artifact of simple omitted spatial structure.

J.1 Constructing Neighborhood and Repeat Shock Controls

We construct three additional spatial controls that capture neighborhood drought density, repeated shocks, and district area.

Neighborhood drought density. The first covariate, *Drought Neighborhood*, measures the fraction of nearby area that is classified as drought affected. It captures the local density of shock exposure in the migration relevant hinterland and helps to account for correlated shocks in destination choice.

We base the buffer radius on a 50 kilometer rural migration distance threshold, consistent with stylized facts from Sub Saharan Africa. For each district polygon P we compute the average distance from internal points to the district boundary $B(P)$,

$$\text{Average Distance to Border} = \frac{1}{|P|} \sum_{p \in P} \min_{b \in B(P)} d(p, b),$$

and define the buffer thickness as

$$\text{Buffer Distance} = 50 - \text{Average Distance to Border}.$$

The resulting ring around each district defines the migration neighborhood. For each year we then compute the share of that neighborhood suffering a drought, which yields the *Drought Neighborhood* variable.

For notational consistency we also report the following expression, which is used in the creation of the buffer zone:

$$\text{Buffer Zone} = \frac{1}{|P|} \sum_{p \in P} \min_{b \in B(P)} d(p, b).$$

Repeat drought. The second control, *Repeat Drought*, is an indicator for districts that experience shocks in two consecutive years, that is in both $t-1$ and $t-2$. This variable captures within sample persistence in shock exposure that is not absorbed by fixed effects. Consecutive shocks may influence investment, adaptation, and selection in ways that differ from isolated events.

Area. The third control, *Area*, records district size in square meters. Larger districts mechanically encompass more internal movement and may show different population responses for purely geometric reasons. Including area ensures that the main effect is not driven by size related heterogeneity.

J.2 Adjusted Baseline TWFE Model

Table 10 reports the baseline specification for rainfed rural settlements with and without the additional spatial controls. The estimated drought effect is stable across specifications, which indicates that the main result is not sensitive to these refinements.

Table 10: Effect of drought shocks on rainfed rural log population relative to unaffected settlements, adjusted for spatial control variables.

Dependent Variable:	Log Population (Settlement Level)				
Data Selection:	Rural & Rainfed	Rural & Rainfed	Rural & Rainfed	Rural & Rainfed	Rural & Rainfed
Drought Shock	-0.0247***	-0.0246***	-0.0220***	-0.0247***	-0.0219***
(First Lag)	(0.0081)	(0.0082)	(0.0071)	(0.0081)	(0.0072)
Drought Neighborhood		-0.0008			-0.0010
		(0.0066)			(0.0066)
Repeat Drought			-0.0164		-0.0164
			(0.0144)		(0.0144)
Area				-0.0062	-0.0062
				(27.19)	(27.19)
Settlement fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	84,705	84,705	84,705	84,705	84,705
R^2	0.27148	0.27148	0.27148	0.27148	0.27148
Within R^2	7.87×10^{-5}	7.88×10^{-5}	8.79×10^{-5}	7.87×10^{-5}	8.81×10^{-5}

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

J.3 Border Districts

If drought-induced population change operates primarily through migration, then the magnitude of the response should depend on how easily residents can move to nearby destinations. Border districts have fewer neighboring districts within short travel distance and weaker market linkages, which increases the cost of out-migration. The migration mechanism therefore predicts attenuated population responses in these locations.

To assess this implication, we split the rural rainfed sample into border and non-border districts and re-estimate the baseline specification.

Table 11: Effect of drought shocks on rainfed rural log population, split by border and non-border districts.

Dependent Variable:		Log Population (Settlement Level)	
Data Selection:	Rural & Rainfed All	Rural & Rainfed Excluding Border Districts	Rural & Rainfed Only Border Districts
Drought Shock	-0.0247***	-0.0370***	-0.0128
(First Lag)	(0.0081)	(0.0113)	(0.0319)
Settlement fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	84,705	83,330	4,650
R^2	0.27148	0.22384	0.17308
Within R^2	7.87×10^{-5}	0.00012	8.25×10^{-6}

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results accord tightly with the migration-based interpretation. The drought effect is larger and more precisely estimated in non-border districts, where migration options are plentiful, and much weaker in border districts, where out-migration is relatively costly. This pattern is inconsistent with alternative channels such as fertility or mortality change and provides supportive evidence that the baseline semi-elasticity reflects mobility responses.

K Primary Assumptions of the TWFE Model

This appendix states the main identifying assumptions underlying the two way fixed effects specification used in the empirical analysis and relates them to features of the rainfall and population data.

Assumption 1: Sharp design. $\forall(i,g,t) \in \{1,...,N_{g,t}\} \times \{1,...,G\} \times \{1,...,T\}, \quad D_{i,g,t} = D_{g,t}$

Treatment is defined at the district year level. All settlements in a given district share the same drought status. For the baseline specification with 690 districts this assignment is precise because districts are relatively small compared to the spatial scale of rainfall fields. As aggregation becomes coarser, the sharp design assumption is progressively violated and the mapping from local exposure to measured treatment weakens, which is exactly the mechanism analyzed in the paper.

Assumption 2: Strong exogeneity. $E[u_{it} \mid X_{it}, \alpha_i, \lambda_t] = 0$

Conditional on settlement fixed effects α_i and year fixed effects λ_t, the error term is uncorrelated with the treatment and covariates. In this setting the key justification is that rainfall anomalies are driven by atmospheric dynamics rather than local economic outcomes. The spatial structure is absorbed by fixed effects and by the drought construction based on long run local distributions. Precipitation is not a choice variable for households or local authorities.

Assumption 2b: No anticipation. $\rho = \mathrm{corr}(E_{t-s}[D_{g,t}], D_{g,t}) = 0 \quad \forall\ s \geq 1$

Agents do not systematically forecast district level drought shocks beyond information that is already embedded in long run climate. Behavior in year $t-s$ therefore does not respond to future anomalies in year t. Empirically, rainfall expectations are usually formed from recent history, so one can treat deviations from the district specific gamma distribution as unanticipated.

Assumption 3: Parallel trends. Untreated potential outcomes in treated and untreated districts follow the same trend once fixed effects are controlled for. In the empirical setting most rural rainfed districts experience drought at some point over the sample, and the socioeconomic structure of these districts is relatively homogeneous with a large agricultural labor share. This reduces the scope for systematic divergence in absence of shocks. The near zero correlation between treatment status and pre treatment trends in population is consistent with this assumption.

L Spillover Effects for Local Migration Patterns

Population losses in treated districts may be offset by inflows into neighboring districts. Such spillovers can bias estimated treatment effects if migration is primarily local. To assess this, we compare log population in non drought districts that are adjacent to drought districts with non drought districts that are not adjacent.

Table 12: Effect of drought shocks on log population in non drought districts adjacent to drought districts relative to non adjacent non drought districts. Settlement level.

Dependent Variable:	Log Population	
	Rural	**Urban**
Drought Shock	0.0030	0.0302**
(First Lag)	(0.0076)	(0.0124)
Settlement fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	70,896	2,614
R^2	0.34912	0.84286
Within R^2	1.4×10^{-6}	0.00065

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For rural settlements, there is no significant difference in population growth between adjacent and non adjacent non drought districts, which suggests that rural spillovers do not materially bias the baseline estimate. For urban settlements, positive and significant spillovers are more plausible, and the urban coefficients in the main text should be interpreted with more caution.

M Additional Socioeconomic Outcomes

We complement the migration based analysis with a set of micro level productivity and welfare outcomes from the ERSS survey. These variables provide a micro foundation for the channels through which drought affects households and allow us to test whether aggregation bias extends beyond population outcomes.

Tables 13 and 14 report baseline TWFE estimates that use the same drought definition and fixed effects structure as in the population regressions.

M.1 Productivity related outcomes

Table 13: Effect of drought shocks on productivity-related socioeconomic outcomes from ERSS.

Dependent Variables:	Self-reported Productivity Shock	Self-reported Crop Damage	Adverse Shock to Food Production	Adverse Shock to Food Stocks	Harvest in Storage (kg)
Drought Shock	0.1570***	0.1652***	0.1478***	0.1432***	-69.37***
(First Lag)	(0.0338)	(0.0459)	(0.0333)	(0.0329)	(24.10)
District fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	10,849	2,718	10,849	10,849	8,162
R^2	0.34862	0.29270	0.32691	0.32120	0.24882
Within R^2	0.02416	0.01569	0.02250	0.02137	0.00344

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

M.2 Welfare related outcomes

Across outcomes, drought shocks reduce stored harvests and meat consumption and increase reported production and income shocks, assets shocks, and health problems, which is consistent with the idea that migration responds to genuine welfare losses.

Table 14: Effect of drought shocks on welfare-related socioeconomic outcomes from ERSS.

Dependent Variables:	Cereal Intake Past 7 Days (g)	Meat Intake Past 7 Days (g)	Health Problems Past 2 Months	Adverse Shock to Income	Adverse Shock to Assets
Drought Shock	-151.3	-297.9***	0.0273**	0.1456***	0.1319***
(First Lag)	(653.2)	(87.12)	(0.0121)	(0.0323)	(0.0273)
District fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	10,833	10,833	15,426	10,849	10,849
R^2	0.27472	0.07018	0.03103	0.33202	0.28607
Within R^2	2.16×10^{-5}	0.00226	0.00031	0.02169	0.02195

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

N Aggregation Bias in the Dependent Variable

The main analysis keeps the outcome at the settlement level and only aggregates treatment. As a separate exercise, we aggregate population itself to the district level and re estimate the baseline model. This isolates the effect of coarsening the dependent variable rather than the shock regressor.

Table 15: Effect of drought shocks on district-level log population when the dependent variable is aggregated from settlement to district level.

Dependent Variable:	Log Population (District Level)	
Data Selection:	Rural & Rainfed	Urban & Rainfed
Drought Shock	-0.0039	-0.0176
(First Lag)	(0.0114)	(0.0156)
District fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	8,551	2,870
R^2	0.94104	0.96228
Within R^2	1.9×10^{-5}	0.00079

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Once population is aggregated to the district level, the estimated effect on rural rainfed population becomes very small and statistically insignificant. This confirms that aggregation of the dependent variable itself can mask economically meaningful local responses, and it reinforces the argument that fine scale outcome data are crucial for credible identification of localized shock effects.

O Bias Decomposition Algebra and the Remainder Term

This appendix derives the exact OLS estimand when the true exposure is $V_i S_{i,t}$ but the regression uses the coarsened regressor $V_g S_{g,t}$. All moments are taken under the estimation sample measure (Appendix Z). The non-i.i.d. component of the regression error is

$$\varepsilon_x \equiv V_i S_{i,t} - V_g S_{g,t} = \varepsilon_{V_i} S_{g,t} + \varepsilon_{S_{i,t}} V_g + \varepsilon_{S_{i,t}} \varepsilon_{V_i}, \tag{9}$$

and the OLS wedge is

$$\hat{\beta} = \beta + \beta \, \frac{\mathrm{Cov}(V_g S_{g,t}, \varepsilon_x)}{\mathrm{Var}(V_g S_{g,t})}, \tag{10}$$

where $\mathrm{Cov}(V_g S_{g,t}, \varepsilon_{i,t}) = 0$ by the maintained orthogonality condition. Substituting (9) splits the wedge into one covariance per component of ε_x:

$$\mathrm{Cov}(V_g S_{g,t}, \varepsilon_x) = \mathrm{Cov}(V_g S_{g,t}, V_g \varepsilon_{S_{i,t}}) + \mathrm{Cov}(V_g S_{g,t}, S_{g,t} \varepsilon_{V_i}) + \mathrm{Cov}(V_g S_{g,t}, \varepsilon_{V_i} \varepsilon_{S_{i,t}}). \tag{11}$$

The third term of (11) is the sorting component, and it can be evaluated exactly without any expansion. Because V_g and $S_{g,t}$ are measurable with respect to the unit-year cell (g, t), the law of iterated expectations gives

$$\mathrm{Cov}(V_g S_{g,t}, \varepsilon_{V_i} \varepsilon_{S_{i,t}}) = \mathrm{Cov}\left(V_g S_{g,t},\ c_{g,t}\right), \qquad c_{g,t} \equiv E[\varepsilon_{V_i} \varepsilon_{S_{i,t}} \mid g, t]. \tag{12}$$

Appendix R shows that under mean-zero within-unit deviations $c_{g,t}$ equals the within-unit covariance $\mathrm{Cov}(V_i, S_{i,t} \mid g, t)$, the sorting between vulnerability and shock intensity inside the unit. Equation (12) makes precise the sense in which sorting biases the slope: what matters is not the average level of $c_{g,t}$ but its covariance with assigned exposure across unit-years. A sorting level common to all cells shifts $E[\varepsilon_x]$, and hence the intercept or the fixed effects, but not the slope coefficient.

For the first two terms of (11) we use the exact identity

$$\operatorname{Cov}(X, YZ) = E[Y]\operatorname{Cov}(X,Z) + E[Z]\operatorname{Cov}(X,Y) + \kappa(X,Y,Z),$$
$$\text{with } \kappa(X,Y,Z) \equiv E\big[(X - E[X])(Y - E[Y])(Z - E[Z])\big], \tag{13}$$

where κ denotes the third central moment. Applying (13) twice to each term, once to peel off the discrepancy and once to expand the product $V_g S_{g,t}$, yields

$$\begin{aligned}\operatorname{Cov}(V_g S_{g,t}, V_g \varepsilon_{S_{i,t}}) &= E[V_g]^2 \operatorname{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) + E[S_{g,t}]E[\varepsilon_{S_{i,t}}] \operatorname{Var}(V_g) \\ &\quad + E[V_g]E[S_{g,t}]\operatorname{Cov}(V_g, \varepsilon_{S_{i,t}}) + E[V_g]E[\varepsilon_{S_{i,t}}]\operatorname{Cov}(V_g, S_{g,t}) + \Gamma_1,\end{aligned} \tag{14}$$

with $\Gamma_1 \equiv E[V_g]\,\kappa(V_g, S_{g,t}, \varepsilon_{S_{i,t}}) + E[\varepsilon_{S_{i,t}}]\,\kappa(V_g, S_{g,t}, V_g) + \kappa(V_g S_{g,t}, V_g, \varepsilon_{S_{i,t}})$, and, symmetrically,

$$\begin{aligned}\operatorname{Cov}(V_g S_{g,t}, S_{g,t} \varepsilon_{V_i}) &= E[S_{g,t}]^2 \operatorname{Cov}(V_g, \varepsilon_{V_i}) + E[V_g]E[\varepsilon_{V_i}] \operatorname{Var}(S_{g,t}) \\ &\quad + E[S_{g,t}]E[V_g]\operatorname{Cov}(S_{g,t}, \varepsilon_{V_i}) + E[S_{g,t}]E[\varepsilon_{V_i}]\operatorname{Cov}(V_g, S_{g,t}) + \Gamma_2,\end{aligned} \tag{15}$$

with Γ_2 defined analogously.

Substituting (12), (14), and (15) into (10) and grouping terms yields the exact OLS estimand

$$\begin{aligned}\hat{\beta} = \beta\Bigg[1 &+ \underbrace{\frac{E[V_g]^2 \operatorname{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) + E[S_{g,t}]^2 \operatorname{Cov}(V_g, \varepsilon_{V_i})}{\operatorname{Var}(V_g S_{g,t})}}_{B_1} + \underbrace{\frac{\operatorname{Cov}\left(V_g S_{g,t},\ c_{g,t}\right)}{\operatorname{Var}(V_g S_{g,t})}}_{B_2} \\ &+ \underbrace{\frac{E[V_g]E[\varepsilon_{V_i}] \operatorname{Var}(S_{g,t}) + E[S_{g,t}]E[\varepsilon_{S_{i,t}}] \operatorname{Var}(V_g)}{\operatorname{Var}(V_g S_{g,t})}}_{B_3} + \frac{\Gamma}{\operatorname{Var}(V_g S_{g,t})}\Bigg],\end{aligned} \tag{16}$$

with remainder

$$
\begin{aligned}
\Gamma \equiv\ & E[V_g]E[S_{g,t}]\big(\operatorname{Cov}(V_g, \varepsilon_{S_{i,t}}) + \operatorname{Cov}(S_{g,t}, \varepsilon_{V_i})\big) \\
&+ \big(E[V_g]E[\varepsilon_{S_{i,t}}] + E[S_{g,t}]E[\varepsilon_{V_i}]\big)\operatorname{Cov}(V_g, S_{g,t}) + \Gamma_1^\kappa + \Gamma_2^\kappa,
\end{aligned} \tag{17}
$$

where Γ_1^κ and Γ_2^κ denote the third central moment collections in Γ_1 and Γ_2. This is the expression reported in the main text.

Every term in B_1, B_3, and Γ involves a lone discrepancy $\varepsilon_{S_{i,t}}$ or ε_{V_i}, entering through its mean, through a covariance with an aggregated proxy, or through a third central moment. Under mean-preserving aggregation with weights and support aligned to the estimation sample, that is,

$$
E[\varepsilon_{S_{i,t}} \mid g, t] = 0 \qquad \text{and} \qquad E[\varepsilon_{V_i} \mid g, t] = 0,
$$

every such term is zero by iterated expectations, because V_g and $S_{g,t}$ are constant within cells. In that case $B_1 = B_3 = \Gamma = 0$ and the entire aggregation bias reduces to the sorting component,

$$
\hat{\beta} = \beta \left[1 + \frac{\operatorname{Cov}\left(V_g S_{g,t},\ c_{g,t}\right)}{\operatorname{Var}(V_g S_{g,t})}\right]. \tag{18}
$$

This is the sense in which separately aggregated proxies can satisfy the Berkson (1950) conditions variable by variable and still yield a biased interaction coefficient: linear mean-preserving aggregation makes each discrepancy orthogonal to its own aggregate, but it cannot deliver $E[V_i S_{i,t} \mid g, t] = V_g S_{g,t}$ when shocks and vulnerability covary within units. Conversely, nonlinear constructions such as thresholding, binning, or weight and support mismatch violate mean preservation and activate B_1, B_3, and Γ; Appendices Q and S analyze these channels in detail. The optional centering restriction $E[S_{g,t} \mid g] = 0$ discussed in Appendix AA.3 eliminates the terms scaled by $E[S_{g,t}]$ together with the covariances $\operatorname{Cov}(S_{g,t}, V_g)$ and $\operatorname{Cov}(S_{g,t}, \varepsilon_{V_i})$.

Equation (16) is an exact decomposition. The main text emphasizes B_1, B_2, and B_3

because they map into interpretable failures of exposure construction: proxy–discrepancy covariances created by nonlinear coarsening (B_1), within-unit sorting that covaries with assigned exposure (B_2), and mean shifts from mismatched weights or supports (B_3). The remainder Γ collects cross-support covariances and third central moments; it vanishes exactly under mean-preserving aggregation and is small whenever the discrepancies behave like approximately centered noise within cells. We therefore treat Γ as a completeness term and build the theoretical and empirical discussion around the three named components.

P Outcome Aggregation and the Exposure Wedge

This appendix shows that aggregating outcomes does not eliminate the bias created by constructing exposure as the product of separately aggregated proxies.

Suppose the observed unit level outcome is an aggregation of location level log outcomes,

$$\log Y_{g,t} = \sum_{i \in g} \omega_i \log Y_{i,t}, \qquad \sum_{i \in g} \omega_i = 1,$$

for some aggregation weights $\{\omega_i\}$.

Substituting the location level model (2) yields

$$\log Y_{g,t} = \beta \sum_{i \in g} \omega_i V_i S_{i,t} + \sum_{i \in g} \omega_i \varepsilon_{i,t}.$$

The relevant exposure object at the unit level is therefore the aggregated interaction

$$X^*_{g,t} \equiv \sum_{i \in g} \omega_i V_i S_{i,t}.$$

Using the product of separately aggregated proxies, $V_g S_{g,t}$, replaces $X^*_{g,t}$ with a generally different object. The resulting exposure wedge is

$$X_{g,t}^{*} - V_g S_{g,t} = \sum_{i \in g} \omega_i V_i S_{i,t} - V_g S_{g,t},$$

which does not vanish in general when V and S vary within g and are aggregated separately.

Outcome aggregation can introduce additional attenuation or sensitivity to the outcome aggregation weights, but it does not resolve the noncommutativity between aggregation and interaction construction that generates the exposure wedge above.

Q Component (i) details: proxy endogeneity from coarsening

This appendix develops component (i) in (19). It provides (i) exact zero conditions, (ii) a sharp identity for binary or binned shocks, and (iii) a catalog of common aggregation choices that make the key covariance terms nonzero. All moments are defined under the estimation sample measure (Appendix Z).

Q.1 Bias term and two primitive covariances

Recall

$$B_1 = \frac{E[V_g]^2 \ \mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) + E[S_{g,t}]^2 \ \mathrm{Cov}(V_g, \varepsilon_{V_i})}{\mathrm{Var}(V_g S_{g,t})}, \qquad \varepsilon_{S_{i,t}} \equiv S_{i,t} - S_{g,t}, \quad \varepsilon_{V_i} \equiv V_i - V_g. \tag{19}$$

Thus B_1 is governed by two objects:

$$\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) \quad \text{and} \quad \mathrm{Cov}(V_g, \varepsilon_{V_i}),$$

which capture whether the aggregated proxy covaries with its own within-unit discrepancy.

Q.2 Exact zero conditions

Assume $\mathrm{Var}(V_g S_{g,t}) > 0$. Then $B_1 = 0$ if and only if

$$E[V_g]^2\ \mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) + E[S_{g,t}]^2\ \mathrm{Cov}(V_g, \varepsilon_{V_i}) = 0.$$

There are two empirically relevant routes.

Route 1: conditional mean unbiasedness. Sufficient conditions are

$$E[\varepsilon_{S_{i,t}} \mid S_{g,t}] = 0 \quad \Longrightarrow \quad \mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) = 0, \qquad E[\varepsilon_{V_i} \mid V_g] = 0 \quad \Longrightarrow \quad \mathrm{Cov}(V_g, \varepsilon_{V_i}) = 0. \tag{20}$$

These restrictions rule out systematic within-unit misclassification by assigned shock status and by assigned vulnerability. They hold under sharp within-unit exposure ($S_{i,t} = S_{g,t}$ for all $i \in g$), sharp within-unit vulnerability ($V_i = V_g$ for all $i \in g$), or correct linear aggregation under weights and support aligned with the estimation sample. They typically fail under nonlinear constructions (thresholding, bins, maxima, winsorization, index truncation) or weight and support mismatch.

Route 2: centering multipliers. Since $E[V_g]^2 = 0$ if and only if $E[V_g] = 0$, explicitly centering V_g eliminates the term involving $\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}})$. Likewise, centering $S_{g,t}$ so that $E[S_{g,t}] = 0$ eliminates the term involving $\mathrm{Cov}(V_g, \varepsilon_{V_i})$. Fixed effects do not mechanically imply either centering, because the within transformation demeans the product $V_g S_{g,t}$ rather than V_g or $S_{g,t}$ separately.

Q.3 Binary and binned shock constructions

When the shock proxy is an indicator (or a bin), $\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}})$ has a transparent interpretation.

Lemma Q.1 (Binary regressor covariance identity) *Let $X \in \{0,1\}$ and Y have finite second moments. Let $p \equiv P(X=1)$. Then*

$$\mathrm{Cov}(X,Y) = p(1-p)\Big(E[Y \mid X=1] - E[Y \mid X=0]\Big).$$

Proof. $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$. Since $X \in \{0,1\}$, $E[XY] = pE[Y \mid X=1]$ and $E[Y] = pE[Y \mid X=1] + (1-p)E[Y \mid X=0]$. Substitute and simplify. $\square$

Applying Lemma Q.1 with $X = S_{g,t}$ and $Y = \varepsilon_{S_{i,t}}$ yields

$$\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) = P(S_{g,t}=1)P(S_{g,t}=0)\Big(E[\varepsilon_{S_{i,t}} \mid S_{g,t}=1] - E[\varepsilon_{S_{i,t}} \mid S_{g,t}=0]\Big). \tag{21}$$

Mechanical interpretation. Equation (21) is nonzero if and only if the average within-unit discrepancy differs between units coded as high exposure and units coded as low exposure. This is exactly the misclassification induced by coarsening. For example, if $S_{g,t}=1$ is assigned when a unit statistic crosses a cutoff, then units coded as treated typically include many locations with $S_{i,t}$ below the cutoff, so $\varepsilon_{S_{i,t}} = S_{i,t} - S_{g,t}$ tends to be negative when $S_{g,t}=1$. Units coded as untreated can still include some high exposure locations, making $\varepsilon_{S_{i,t}}$ tend to be positive when $S_{g,t}=0$. The difference in conditional means drives $\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}})$.

The same logic extends to multi-bin $S_{g,t}$ by conditioning on each bin and comparing mean discrepancies across bins.

Q.4 Common sources of nonzero covariances

The conditional mean restrictions in (20) fail in several recurring applied situations.

Case A. Thresholding, discretization, or binning. If $S_{g,t}$ is built by applying cutoffs to an aggregated statistic, then $E[\varepsilon_{S_{i,t}} \mid S_{g,t}]$ typically varies with the assigned bin, making $\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}) \neq 0$.

Case B. Nonlinear aggregation of continuous shocks. If $S_{g,t}$ is a max, median, quantile, winsorized mean, capped index, or any construction that aggregates and then applies a nonlinear map, the sign and magnitude of $S_{i,t} - S_{g,t}$ typically depend on $S_{g,t}$, violating $E[\varepsilon_{S_{i,t}} \mid S_{g,t}] = 0$.

Case C. Weight or support mismatch. If $S_{g,t}$ is built using weights that differ from the regression weights (for example area weights versus population weights), or using a construction sample that differs from the estimation sample, then $\varepsilon_{S_{i,t}}$ under the estimation measure need not be mean zero conditional on $S_{g,t}$ even under linear aggregation.

Case D. Endogenous within-unit sample composition. If the observed micro locations are not representative of the population implicitly targeted by $S_{g,t}$, then $E[\varepsilon_{S_{i,t}} \mid S_{g,t}]$ can be nonzero. This is a sample selection channel distinct from purely mechanical coarsening.

Case E. Measurement quality correlated with intensity. If measurement error in $S_{i,t}$ varies with realized conditions, then the induced discrepancy $\varepsilon_{S_{i,t}}$ can be systematically related to $S_{g,t}$.

Q.5 Symmetric results for vulnerability

All results above apply verbatim with V replacing S. In particular,

$$\operatorname{Cov}(V_g, \varepsilon_{V_i}) \neq 0$$

is plausible when V_g is discretized, top-coded, winsorized, constructed via a nonlinear index, or built under mismatched weights or support. The key sufficient condition for $\operatorname{Cov}(V_g, \varepsilon_{V_i}) = 0$ is $E[\varepsilon_{V_i} \mid V_g] = 0$, which is most defensible under correct linear aggregation with weights aligned to the estimation sample or under sharp within-unit vulnerability.

Q.6 Denominator scaling under coarsening

The denominator $\operatorname{Var}(V_g S_{g,t})$ scales all bias components. Coarsening often reduces $\operatorname{Var}(V_g S_{g,t})$ by smoothing $S_{g,t}$ across space and compressing dispersion in V_g (for example through bins or truncation). This does not create bias by itself, but it magnifies any nonzero numerator covariances in (19).

R Component (ii) details: within unit sorting between vulnerability and exposure

This appendix develops the sorting component B_2 derived in (12). It provides (i) a characterization of the within-cell moment $c_{g,t}$ that drives the component, (ii) exact zero conditions in the same spirit as Appendix Q, together with the level-versus-slope distinction and the sign of the bias, and (iii) a catalog of common mechanisms that make within unit sorting nonzero. All moments are defined under the estimation sample measure (Appendix Z).

R.1 Bias term and its primitive object

Recall

$$B_2 = \frac{\mathrm{Cov}\left(V_g S_{g,t},\, c_{g,t}\right)}{\mathrm{Var}(V_g S_{g,t})}, \qquad c_{g,t} \equiv E[\varepsilon_{V_i}\varepsilon_{S_{i,t}} \mid g,t], \qquad \varepsilon_{S_{i,t}} \equiv S_{i,t} - S_{g,t}, \quad \varepsilon_{V_i} \equiv V_i - V_g. \tag{22}$$

Thus B_2 is governed by a single primitive object, the within-cell moment $c_{g,t}$, and by its joint distribution with assigned exposure $V_g S_{g,t}$ across unit-years. Decomposing the conditional moment,

$$c_{g,t} = \mathrm{Cov}(\varepsilon_{V_i}, \varepsilon_{S_{i,t}} \mid g,t) + E[\varepsilon_{V_i} \mid g,t]\, E[\varepsilon_{S_{i,t}} \mid g,t], \tag{23}$$

shows that $c_{g,t}$ combines within unit sorting between the two discrepancies with the product of any within-cell mean shifts. Under mean-preserving aggregation the second term is zero and $c_{g,t}$ is pure sorting.

R.2 Exact zero conditions

Assume $\mathrm{Var}(V_g S_{g,t}) > 0$. Then $B_2 = 0$ if and only if

$$\mathrm{Cov}\left(V_g S_{g,t},\, c_{g,t}\right) = 0. \tag{24}$$

Three routes are empirically relevant.

Route 1: no sorting. If $c_{g,t} = 0$ in every cell, for example because V_i or $S_{i,t}$ is constant within units, or because within unit deviations in vulnerability are unrelated to within unit deviations in exposure in the sense formalized below, then $B_2 = 0$. This is strong in spatial settings, where settlement and geography often align vulnerability and exposure.

Route 2: homogeneous sorting. If $c_{g,t} = c$ is constant across unit-years, the covariance in (24) is zero even though sorting is present everywhere. The proxy then understates true exposure by the same amount in every cell, which shifts the intercept, or is absorbed by the fixed effects, but leaves the slope unbiased.

Route 3: sorting orthogonal to assigned exposure. More generally, $B_2 = 0$ whenever the variation in $c_{g,t}$ across unit-years is uncorrelated with assigned exposure $V_g S_{g,t}$. This holds only by accident in applications, because the same spatial structure that creates sorting typically also determines where measured exposure is high.

R.3 Level versus slope, and the sign of the bias

The representation $B_2 = \mathrm{Cov}(V_g S_{g,t}, c_{g,t}) / \mathrm{Var}(V_g S_{g,t})$ separates two effects of sorting that are easily conflated. The level effect is that positive sorting makes true exposure exceed the proxy on average, since $E[V_i S_{i,t} \mid g, t] = V_g S_{g,t} + c_{g,t}$. By itself this level effect does not bias the slope: a constant understatement of the regressor moves the intercept, not the coefficient. The slope effect operates through the covariance: sorting biases the coefficient only insofar as it is systematically stronger or weaker in the unit-years where assigned exposure is high.

The sign follows directly. In the typical spatial configuration, sorting is concentrated precisely in high-exposure cells. With a binary local shock, $c_{g,t}$ is nonzero only in unit-years in which the shock hits part of the unit, and it is largest when the shock hits the vulnerable part; these are also the cells in which $V_g S_{g,t}$ is high. Then $\mathrm{Cov}(V_g S_{g,t}, c_{g,t}) > 0$, the proxy understates exposure roughly in proportion to exposure itself, and $\hat{\beta}$ is inflated away from zero. A simple example: suppose half of a unit is hit, the hit half has $V_i = 1$, and the other half has $V_i = 0$. The proxy is $V_g S_{g,t} = 0.5 \times 0.5 = 0.25$, while true average exposure is 0.5; if this configuration recurs whenever the unit is coded as exposed, the regression on the proxy recovers roughly twice the true coefficient. Conversely, if sorting concentrates in low-exposure cells, for example when shocks systematically strike the least vulnerable parts of units with high average vulnerability, the covariance is negative and the estimate attenuates. The direction of the sorting bias is therefore an empirical property of the joint spatial distribution of shocks and vulnerability, not a general theorem.

R.4 Sorting part: the within-cell moment $c_{g,t}$

Recall

$$\varepsilon_{V_i} = V_i - V_g, \qquad \varepsilon_{S_{i,t}} = S_{i,t} - S_{g,t}.$$

Because conditioning on the cell (g, t) makes V_g and $S_{g,t}$ constants,

$$\mathrm{Cov}(\varepsilon_{V_i}, \varepsilon_{S_{i,t}} \mid g, t) = \mathrm{Cov}(V_i, S_{i,t} \mid g, t). \tag{25}$$

Key special case: mean zero deviations within cells. If the aggregation rules and estimation weights imply

$$E[\varepsilon_{V_i} \mid g, t] = 0 \qquad \text{and} \qquad E[\varepsilon_{S_{i,t}} \mid g, t] = 0, \tag{26}$$

then the mean product term in (23) vanishes and

$$c_{g,t} = \mathrm{Cov}(V_i, S_{i,t} \mid g, t) \qquad \text{under (26)}. \tag{27}$$

Proof. [Proof of (27)] By definition,

$$c_{g,t} = E[\varepsilon_{V_i}\varepsilon_{S_{i,t}} \mid g, t] = \mathrm{Cov}(\varepsilon_{V_i}, \varepsilon_{S_{i,t}} \mid g, t) + E[\varepsilon_{V_i} \mid g, t]\, E[\varepsilon_{S_{i,t}} \mid g, t].$$

Under (26) the second term is zero, and (25) gives the result. □

Interpretation. Equation (27) shows that $c_{g,t}$ is the within unit covariance between vulnerability and shock intensity in cell (g,t). It captures whether, inside the aggregate, the more vulnerable micro locations are also systematically the more exposed micro locations in that year. Averaging over cells, the law of total covariance links $c_{g,t}$ to the unconditional object,

$$\mathrm{Cov}(\varepsilon_{V_i}, \varepsilon_{S_{i,t}}) = E[c_{g,t}] \qquad \text{under (26)}, \tag{28}$$

so the unconditional covariance between the two discrepancies measures the average level of sorting. As emphasized above, this level enters the intercept; the slope bias B_2 depends on how $c_{g,t}$ varies with assigned exposure across unit-years.

Sufficient conditions for $c_{g,t} = 0$ in every cell. Any of the following is sufficient.

- Either discrepancy is identically zero, meaning V_i is constant within g or $S_{i,t}$ is constant within g.
- Within each cell, V_i and $S_{i,t}$ are orthogonal, meaning $\mathrm{Cov}(V_i, S_{i,t} \mid g,t) = 0$ for all (g,t), and mean zero within unit deviations hold as in (26).
- A conditional mean independence restriction, for example $E[\varepsilon_{V_i} \mid \varepsilon_{S_{i,t}}, g, t] = 0$ or $E[\varepsilon_{S_{i,t}} \mid \varepsilon_{V_i}, g, t] = 0$.

Why continuity does not help. Even if both V and S are continuous at both scales, within unit sorting can still generate $c_{g,t} \neq 0$. This is a joint spatial structure object, not a discretization artifact.

R.5 Five mechanisms that make within unit sorting likely

Mechanism S1. Settlement patterns and geographic gradients. If vulnerable households locate in marginal micro areas inside each unit that are also more exposed, the within unit covariance is positive.

Mechanism S2. Targeted protection and infrastructure. If irrigation, drainage, or flood protection is placed in particular micro areas that also differ in exposure, the covariance can be positive or negative depending on targeting.

Mechanism S3. Nonuniform population correlated with micro climate. If population mass within polygons is concentrated in micro regions that also differ systematically in shock intensity, then V_i and $S_{i,t}$ will covary within g.

Mechanism S4. Aggregation that mixes distinct micro climates and socioeconomic zones. Larger units combine subregions that differ both in exposure and vulnerability. If high vulnerability subregions tend to be high exposure subregions, the covariance becomes more positive in magnitude as coarsening grows.

Mechanism S5. Shared inputs in measurement construction. If vulnerability and shock measures share spatial predictors or data sources, their within unit deviations can be correlated even if the underlying causal objects are unrelated.

R.6 Putting the pieces together for B_2

Given (22) and (23), component (ii) is absent when sorting is absent in every cell, when sorting is homogeneous across cells, or when its variation is orthogonal to assigned exposure. The practically relevant assessment is therefore whether the vulnerable parts of units are hit differentially in exactly the unit-years that provide the identifying variation in $V_g S_{g,t}$. In spatial settings this is the norm rather than the exception, and the resulting bias can amplify or attenuate the estimate depending on where sorting concentrates.

R.7 Fixed effects and omitted vulnerability

Unit fixed effects can absorb additive time invariant vulnerability levels when V_i enters in levels, and they absorb the unit-average level of sorting, that is, the time-invariant component of $c_{g,t}$. They do not eliminate sorting whose intensity varies with where shocks strike within the unit over time. That time-varying component is a within unit covariance in fine scale heterogeneity, and it survives within transformations unless the aggregation moments above are satisfied.

S Additional results for bias component (iii)

This appendix develops component (iii) in full detail. It provides the signal identity (29), conditions for $E[\varepsilon_{V_i}] = 0$, a binary signal decomposition for $\mathrm{Cov}(S_{g,t}, S_{i,t})$, and common cases where the mean shift is nonzero. All moments are defined under the estimation sample measure (Appendix Z).

S.1 Algebraic form and the signal interpretation

Start from the form of B_3 in the main-text decomposition (5),

$$B_3 = \frac{E[V_g]E[\varepsilon_{V_i}]\ \mathrm{Var}(S_{g,t}) + E[S_{g,t}]E[\varepsilon_{S_{i,t}}]\ \mathrm{Var}(V_g)}{\mathrm{Var}(V_g S_{g,t})}.$$

Each mean shift is scaled by the variance of the other aggregated proxy. The remainder of this appendix focuses on the vulnerability mean shift, $E[\varepsilon_{V_i}]$; the shock mean shift term is symmetric, with the roles of V and S exchanged. Using $\varepsilon_{S_{i,t}} = S_{i,t} - S_{g,t}$,

$$\mathrm{Cov}(S_{g,t}, S_{i,t}) = \mathrm{Cov}(S_{g,t}, S_{g,t} + \varepsilon_{S_{i,t}}) = \mathrm{Var}(S_{g,t}) + \mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}}), \tag{29}$$

so whenever the misclassification covariance $\mathrm{Cov}(S_{g,t}, \varepsilon_{S_{i,t}})$ from component (i) is zero, as under mean-preserving shock aggregation, the variance factor equals the shock signal,

$$\mathrm{Var}(S_{g,t}) = \mathrm{Cov}(S_{g,t}, S_{i,t}).$$

Interpretation. The factor $\mathrm{Cov}(S_{g,t}, S_{i,t})$ is the shock signal term. It measures how much the aggregated shock tracks micro exposure. It is zero when the aggregated shock is uninformative about the micro shock, which is typically an identification failure rather than a benign special case. The vulnerability mean shift $E[\varepsilon_{V_i}]$ therefore loads onto exactly the variation that identifies the effect.

S.2 Exact zero and nonzero conditions

Given $\mathrm{Var}(V_g S_{g,t}) > 0$,

$$B_3 = 0 \quad \Longleftrightarrow \quad E[V_g]E[\varepsilon_{V_i}] \ \mathrm{Var}(S_{g,t}) + E[S_{g,t}]E[\varepsilon_{S_{i,t}}] \ \mathrm{Var}(V_g) = 0.$$

Outside knife-edge cancellations, this requires shutting down each product separately. The conditions below target the vulnerability piece; symmetric conditions, centering $S_{g,t}$ or mean preserving shock aggregation with $E[\varepsilon_{S_{i,t}}] = 0$, shut down the shock piece.

Condition Z1. Centered vulnerability level. $E[V_g] = 0$ eliminates the vulnerability piece. This requires explicit centering or deviation coding of V_g prior to constructing the interaction.

Condition Z2. Mean preserving vulnerability aggregation. $E[\varepsilon_{V_i}] = 0$ eliminates the vulnerability piece. Since $E[\varepsilon_{V_i}] = E[V_i] - E[V_g]$, this requires $E[V_i] = E[V_g]$ in the estimation sample. A sufficient condition is correct linear aggregation of V under weights and support aligned with the estimation sample.

Condition Z3. No shock variation. $\mathrm{Var}(S_{g,t}) = 0$ eliminates the vulnerability piece, but it also means the aggregated shock has no identifying variation, which is an identification failure rather than a benign special case.

S.3 Binary decomposition of the shock signal

If $S_{g,t} \in \{0, 1\}$, then the signal term admits the standard binary identity,

$$\begin{aligned} \mathrm{Cov}(S_{g,t}, S_{i,t}) = P(S_{g,t} = 1)P(S_{g,t} = 0) \\ \times \Big(E[S_{i,t} \mid S_{g,t} = 1] - E[S_{i,t} \mid S_{g,t} = 0]\Big). \end{aligned} \tag{30}$$

Coarsening that increases within unit mixing typically shrinks the conditional mean gap in (30), reducing signal toward zero.

S.4 Five common sources of mean vulnerability shifts, $E[\varepsilon_{V_i}] \neq 0$

Case B3.1. Discretizing or binning V_g. If V_g is constructed as categories rather than a mean under the regression weights, it need not preserve the micro mean.

Case B3.2. Nonlinear index construction. If V_g is a median, max, capped index, or any nonlinear function of micro inputs, then $E[V_g]$ can systematically differ from $E[V_i]$.

Case B3.3. Weight mismatch between construction and estimation. If V_g is area weighted but the regression weights are effectively population weights, then $E[V_i] - E[V_g] \neq 0$ is expected.

Case B3.4. Support mismatch and sample composition. If the construction frame for V_g differs from the estimation sample support, the distribution of V_i among observed micro locations can differ from what is embedded in V_g.

Case B3.5. Nonclassical measurement error and missingness. If measurement error or missingness in V_i is correlated with vulnerability, then $E[V_i]$ in the estimation sample can differ directionally from $E[V_g]$.

S.5 Fixed effects and this component

Fixed effects remove additive time invariant confounding in outcomes, but they do not guarantee $E[\varepsilon_{V_i}] = 0$ because that is a mean preservation property of the aggregation rule and the weights. Fixed effects also do not guarantee $E[V_g] = 0$ unless V_g is explicitly centered before constructing the interaction.

T Fixed Effects and Interaction Identification

This Appendix derives the estimator under aggregation with annihilation of unity and time fixed effects. It shows under what conditions fixed effects can alleviate the bias associated with aggregation of the independent variable.

The matrix $\mathbf{D} = [\mathbf{D}_{\text{unit}} \ \mathbf{D}_{\text{time}}]$ stacks the unit and time dummies the associated annihilator is

$$M_{\mathbf{D}} = I - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-}\mathbf{D}',$$

where $(\cdot)^{-}$ signifies the generalized inverse. $M_{\mathbf{D}}$ is symmetric and idempotent, and $M_{\mathbf{D}}\mathbf{D} = 0$. We defined a residualized (on time and unit fixed effects) variable as $\tilde{Z} \equiv M_{\mathbf{D}} Z$.

The fixed effects estimator, i.e. the regression of the residualized outcome on the residualized regressor, is

$$\hat{\beta}_{\mathrm{FE}} = \frac{\mathrm{Cov}\left(\widetilde{V_g S_{g,t}}, \widetilde{\log Y}\right)}{\mathrm{Var}(\widetilde{V_g S_{g,t}})}. \tag{31}$$

Using the true model $\log \tilde{Y} = \beta \widetilde{V_i S_{i,t}} + \tilde{\varepsilon}$ in the fixed effects estimator gives:

$$\hat{\beta}_{\mathrm{FE}} = \beta \, \frac{\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \widetilde{V_i S_{i,t}})}{\mathrm{Var}(\widetilde{V_g S_{g,t}})} + \frac{\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \tilde{\varepsilon})}{\mathrm{Var}(\widetilde{V_g S_{g,t}})}.$$

Linearity of $M_{\mathbf{D}}$ gives $\widetilde{V_i S_{i,t}} = \widetilde{V_g S_{g,t}} + \widetilde{\varepsilon}_x$, so

$$\mathrm{Cov}(\widetilde{V_g S_{g,t}}, V_i \tilde{S}_{i,t}) = \mathrm{Var}(\widetilde{V_g S_{g,t}}) + \mathrm{Cov}(\widetilde{V_g S_{g,t}}, \widetilde{\varepsilon}_x),$$

and dividing by $\mathrm{Var}(\widetilde{V_g S_{g,t}})$ yields

$$\hat{\beta}_{\mathrm{FE}} = \beta \left[1 + \frac{\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \widetilde{\varepsilon}_x)}{\mathrm{Var}(\widetilde{V_g S_{g,t}})} \right] \tag{32}$$

This is the partialled out version of the OLS estimator with the bias expression.

Using this, the bias term in the fixed effects estimator (32) is zero if and only if $\mathrm{Cov}(\widetilde{V_g S_{g,t}}, \tilde{\varepsilon}_x) = 0$. Given the covariance structure, the sufficient condition is $\tilde{\varepsilon}_x = 0$. This holds if and only if the wedge ε_x is additively separable into a unit and a time component.

By linearity of $M_{\mathbf{D}}$, and following the steps for the OLS bias composition, the within-transformed wedge decomposes into the three structural parts analogous to the OLS case:

$$\tilde{\varepsilon}_x = \underbrace{M_{\mathbf{D}}\left(\varepsilon_i^V S_{g,t}\right)}_{\text{vulnerability discrepancy}} + \underbrace{M_{\mathbf{D}}\left(\varepsilon_{i,t}^S V_g\right)}_{\text{shock discrepancy}} + \underbrace{M_{\mathbf{D}}\left(\varepsilon_{i,t}^S \varepsilon_i^V\right)}_{\text{within-unit sorting}} . \tag{33}$$

The bias discussion is similar to the OLS case, provided that the within transformation does not absorb the interacted variation listed in the three sources of bias.

The formal question is whether the annihilator equates the interactions of discrepances and the spatially aggregated vulnerability and shock to zero. Generally, it does not. To see this, first note that given the exhaustive fixed effects structure in $\mathbf{D}$, $M_{\mathbf{D}}$ also demeans, so for any Z_1, Z_2 the within second moments are the inner products of their annihilator residuals,

$$\mathrm{Cov}(\tilde{Z}_1, \tilde{Z}_2) = \tfrac{1}{N}\, \tilde{Z}_1' \tilde{Z}_2 = \tfrac{1}{N}\, Z_1' M_{\mathbf{D}} Z_2, \qquad \mathrm{Var}(\tilde{Z}) = \tfrac{1}{N}\, \tilde{Z}' \tilde{Z}, \tag{34}$$

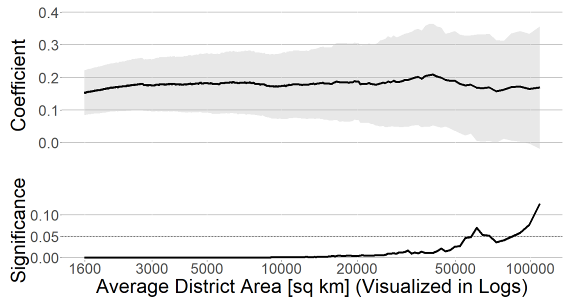
where the variance uses the idempotence of $M_{\mathbf{D}}$.

Using this, none of the three sources of bias is zero in general. The "vulnerability discrepancy" is the product of a time-invariant cross-sectional discrepancy ε_i^V and a time-varying aggregate $S_{g,t}$; such a product is additively separable only if $\varepsilon_i^V \equiv 0$, or $S_{g,t}$ does not vary over time within units. The "shock discrepancy" is the product of a time-varying discrepancy $\varepsilon_{i,t}^S$ and the aggregate V_g, separable only if $\varepsilon_{i,t}^S \equiv 0$. The "within-unit sorting" is a product of two discrepancies and is separable only if one of them is identically zero. These exclusions are the knife-edge no-bias cases described in Appendix Y.

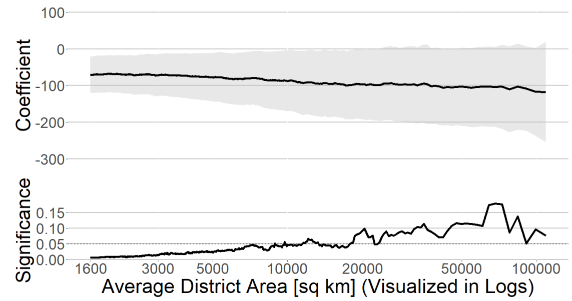
U Aggregation Sensitivity for Socioeconomic Microvariables

We apply the same random aggregation algorithm used for population to four representative ERSS variables: adverse shock to food production, harvest in storage, adverse shock to income, and recent health problems. For each variable we re estimate the baseline model across aggregation levels and report mean coefficients and median p values.

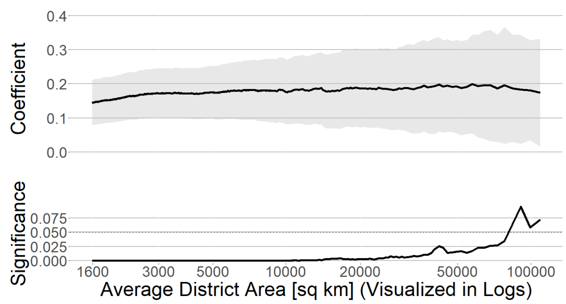
For most variables, coefficients move toward zero and significance deteriorates as spatial units become larger, although the exact threshold differs by outcome. Even in a relatively



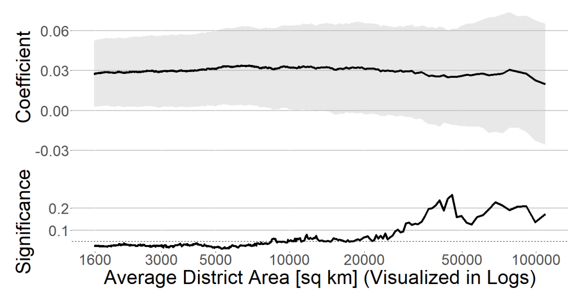
(a) Shock to Food Production



(b) Harvest in Storage



(c) Shock to Income



(d) Health Problems

Figure 6: Effect of drought shocks on four representative socioeconomic microvariables related to productivity and welfare. *Notes:* Adjusted specification that interacts continuous rainfall with vulnerability.

homogeneous survey sample, aggregation weakens measured impacts. The adjusted specification with continuous rainfall and vulnerability interaction improves robustness somewhat, although the gains are smaller than for the migration outcome because survey villages are all located in highly vulnerable regions.

V Derivations for the Gravity Model

This appendix provides the derivations connecting the generic migration data-generating process in Section 6 to the reduced-form coefficient β defined in Equation (8). We first show the algebraic link between outmigration, inflows, and the log population ratio, and then derive the expected inflow term implied by the gravity migration structure in Equation (7).

A. Log population ratio (Equation (8))

Let each location have initial population P. When a location is struck by a shock, a fraction γ of its residents leave, while non-affected locations receive inflows I:

$$\text{Population (shocked)} = (1-\gamma)P, \qquad \text{Population (non-shocked)} = P + I.$$

The reduced-form parameter in Equation (8) is

$$\beta = \ln\left(\frac{(1-\gamma)P}{P+I}\right) = \ln(1-\gamma) - \ln\left(1 + \frac{I}{P}\right).$$

This expression establishes the mapping from the underlying migration parameters

$$(\gamma, I/P)$$

to the empirical estimand β.

B. Expected inflows under the gravity specification (linking to Equation (7))

We now derive the expected inflow term I implied by the gravity-based bilateral migration rule in Equation (7). The bilateral inflow from a shocked origin j into a non-shocked destination i is

$$I_{ijt} = \frac{P_{it}}{NDP_t}\,\theta\gamma P_{jt}\left(1 - \frac{T_{ij}}{\sum_j T_{ij}}\right),$$

where T_{ij} is an inverse-distance weight and NDP_t denotes total population across non-shocked destinations.

Summing over all shocked origins $j \in A_t$ yields total inflows:

$$I_{it} = \sum_{j \in A_t} I_{ijt} = \frac{P_{it}}{NDP_t} \theta\gamma \sum_{j \in A_t} P_{jt} \left(1 - \frac{T_{ij}}{\sum_j T_{ij}}\right).$$

Under the simplifying assumptions used in the simulation calibration (i.i.d. pre-shock populations $P_{jt} = P$, symmetry across bilateral frictions T_{ij}, and $|A_t|$ shocked origins), this expression reduces to

$$I_{it} = \frac{P_{it}}{NDP_t} \theta\gamma P \, |A_t|.$$

Since the total population of shocked locations is θP_t and $NDP_t = (1-\theta)P_t$, taking expectations yields

$$\mathbb{E}_t \left[\frac{I_{it}}{P_{it}}\right] = \frac{\gamma\theta}{1-\theta},$$

which is the expression used in Section 6 to calibrate the simulated migration environment.

C. Mapping the structural inflow expression into the reduced-form parameter

Substituting the inflow ratio into the reduced-form coefficient in Equation (8), we obtain

$$\beta = \ln(1-\gamma) - \ln\left(1 + \frac{\gamma\theta}{1-\theta}\right),$$

which provides the exact mapping from structural behavioral parameters (γ, θ) to the empirical semi-elasticity β used in the simulation. This expression is the basis for calibrating (γ, θ) to match the empirical estimates in Section 3.

W The Modifiable Areal Unit Problem

The Modifiable Areal Unit Problem presents a core challenge when using remotely sensed data. Spatial units are typically defined by arbitrary grid choices rather than by the underlying structure of the variable of interest, and this mismatch can distort empirical relationships. Resampling intensifies the problem because averaging or otherwise combining neighboring raster cells introduces artificial smoothing and measurement error. Since natural systems vary sharply across space, analyses must verify that results are stable across alternative aggregation scales. Our approach explicitly tests this stability to ensure that spatial patterns remain consistent when units are redefined. Accounting for these MAUP effects is essential when working with large scale spatial data, and especially critical for extreme weather events where fine scale spatial precision is necessary for credible inference.

X Aggregation Operators, Weights, and Mean Shifts

This appendix formalizes mean shifts induced by aggregation.

Let $V_g = \sum_{i \in g} a_i V_i$ be constructed using weights $\{a_i\}$, while the regression estimand is defined under weights $\{w_{i,t}\}$. When $a_i \neq w_{i,t}$ or when the construction sample differs from the estimation sample,

$$E[V_g] \neq E[V_i] \quad \Rightarrow \quad E[\varepsilon_{V_i}] \neq 0.$$

Discretization, thresholding, binning, or nonlinear index construction can further induce mean shifts even when linear aggregation uses correct weights. These mean shifts play a central role in bias component (iii) and interact with the signal content of the aggregated shock.

X.1 Optional weights centering of the aggregated shock

The maintained restriction $\text{Cov}(V_g S_{g,t}, \varepsilon_{i,t}) = 0$ is the exogeneity condition required for identification of β in (3). It requires that the idiosyncratic outcome disturbance $\varepsilon_{i,t}$ is orthogonal to the aggregated exposure measure $V_g S_{g,t}$ under the estimation sample measure.

This condition holds, for example, when the structural shock process $S_{i,t}$ is orthogonal to $\varepsilon_{i,t}$ at the location level and aggregation affects only the measurement of exposure rather than the assignment of shocks. More generally, it is satisfied whenever any dependence between outcomes and exposure operates through the interaction term $V_i S_{i,t}$ in (2), and the remaining disturbance $\varepsilon_{i,t}$ is mean independent of $V_g S_{g,t}$.

Y Special Cases with No Aggregation Bias

Aggregation does not distort the OLS estimand in several knife edge cases.

First, if exposure is sharp within units, so that $S_{i,t} = S_{g,t}$ for all $i \in g$, then $\varepsilon_{S_{i,t}} \equiv 0$.

Second, if vulnerability is constant within units, so that $V_i = V_g$ for all $i \in g$, then $\varepsilon_{V_i} \equiv 0$.

Third, if the interaction is constructed before aggregation, $X_{i,t} \equiv V_i S_{i,t}$ and $X_{g,t} \equiv \sum_{i \in g} \omega_i X_{i,t}$ under the estimation weights, then the regressor equals the aggregated true exposure and the aggregation wedge vanishes. Equivalently, if V_g and $S_{g,t}$ are correctly weighted within-unit means and $\text{Cov}(V_i, S_{i,t} \mid g) = 0$ for all g, then $E[V_i S_{i,t} \mid g] = V_g S_{g,t}$.

Outside these special cases, aggregation bias is generically present when shocks and vulnerability vary within units.

Z Expectation Operators, Weights, and the OLS Estimand

This appendix defines the expectation operators used throughout the bias decomposition.

All expectations, variances, and covariances are taken with respect to the estimation sample measure. If the regression is unweighted, this corresponds to the empirical distribution over observed (i, t). If the regression uses weights $w_{i,t}$, define normalized weights $\tilde{w}_{i,t} \equiv w_{i,t} / \sum_{i,t} w_{i,t}$ and interpret

$$E[Z] \equiv \sum_{i,t} \tilde{w}_{i,t} Z_{i,t}, \qquad \mathrm{Cov}(Z_1, Z_2) \equiv E[(Z_1 - E[Z_1])(Z_2 - E[Z_2])], \qquad \mathrm{Var}(Z) \equiv \mathrm{Cov}(Z, Z).$$

This convention ensures that all moment objects in the bias expressions are defined under the same weights that determine the OLS estimand. In particular, mean shifts such as $E[\varepsilon_{V_i}]$ and $E[\varepsilon_{S_{i,t}}]$ are defined with respect to the estimation sample measure, making explicit the role of weight and support mismatch.

AA Error Decomposition and the OLS Estimand

This appendix derives the error decomposition in (4) and the corresponding OLS estimand. All expectations and covariances are taken with respect to the estimation sample measure, including any regression weights, as defined in Appendix Z.

AA.1 Error decomposition

Start from the fine scale model (2) and the empirical specification (3). The regression error can be written as

$$u_{i,t} = \beta\big(V_i S_{i,t} - V_g S_{g,t}\big) + \varepsilon_{i,t}.$$

Define discrepancies relative to the aggregated proxies,

$$\varepsilon_{S_{i,t}} \equiv S_{i,t} - S_{g,t}, \qquad \varepsilon_{V_i} \equiv V_i - V_g,$$

so that $S_{i,t} = S_{g,t} + \varepsilon_{S_{i,t}}$ and $V_i = V_g + \varepsilon_{V_i}$. Then

$$\begin{aligned} V_i S_{i,t} - V_g S_{g,t} &= (V_g + \varepsilon_{V_i})(S_{g,t} + \varepsilon_{S_{i,t}}) - V_g S_{g,t} \\ &= V_g \varepsilon_{S_{i,t}} + \varepsilon_{V_i} S_{g,t} + \varepsilon_{V_i} \varepsilon_{S_{i,t}}, \end{aligned}$$

which delivers (4).

AA.2 OLS estimand

The OLS estimand from regressing $\log Y_{i,t}$ on $V_g S_{g,t}$ is

$$\hat{\beta} = \beta + \frac{\mathrm{Cov}(V_g S_{g,t}, u_{i,t})}{\mathrm{Var}(V_g S_{g,t})}.$$

Under the maintained orthogonality condition $\mathrm{Cov}(V_g S_{g,t}, \varepsilon_{i,t}) = 0$, the bias is driven entirely by the aggregation induced component of $u_{i,t}$.

AA.3 Optional centering of the aggregated shock in OLS

Some applications construct $S_{g,t}$ as a within unit anomaly. A generic way to represent this is

$$S_{g,t} \equiv \tilde{S}_{g,t} - E[\tilde{S}_{g,t} \mid g],$$

where $\tilde{S}_{g,t}$ denotes the raw aggregated shock. Empirically, the same restriction can be imposed by residualizing $\tilde{S}_{g,t}$ on unit fixed effects, and any additional time invariant unit level controls, and defining $S_{g,t}$ as the residual. If $E[S_{g,t} \mid g] = 0$ holds, then $E[S_{g,t} V_g] = 0$

and $E[S_{g,t}\varepsilon_{V_i}] = 0$ because V_g and ε_{V_i} are time invariant.

AB Shock Impact by Age Groups and Demographics

To clarify whether population changes reflect migration rather than fertility or mortality responses, we estimate the effect of drought on age structure and household composition using ERSS microdata. Table 16 reports the results.

Table 16: Effect of drought shocks on demographic composition in rainfed rural areas. Dependent variables are subgroup population shares.

Dependent Variables:	Children (0–14 years)	Working Age (15–59 years)	Elderly (60+ years)	In a Committed Relationship	Women
Drought Shock (First Lag)	0.0145** (0.0055)	-0.0186*** (0.0057)	0.0041* (0.0020)	0.0292*** (0.0054)	0.0164** (0.0063)
District fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	824	824	824	824	824
R^2	0.588042	0.59100	0.812098	0.56533	0.428248
Within R^2	0.00914	0.01551	0.00885	0.03758	0.010421

Notes: Standard errors, clustered at the district level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Drought exposure reduces the working age share and increases the shares of children and elderly, and it raises the fraction of individuals in committed relationships. These patterns are consistent with selective out migration of working age and single individuals rather than with fertility or mortality shifts, which supports the migration based interpretation of the baseline population response.