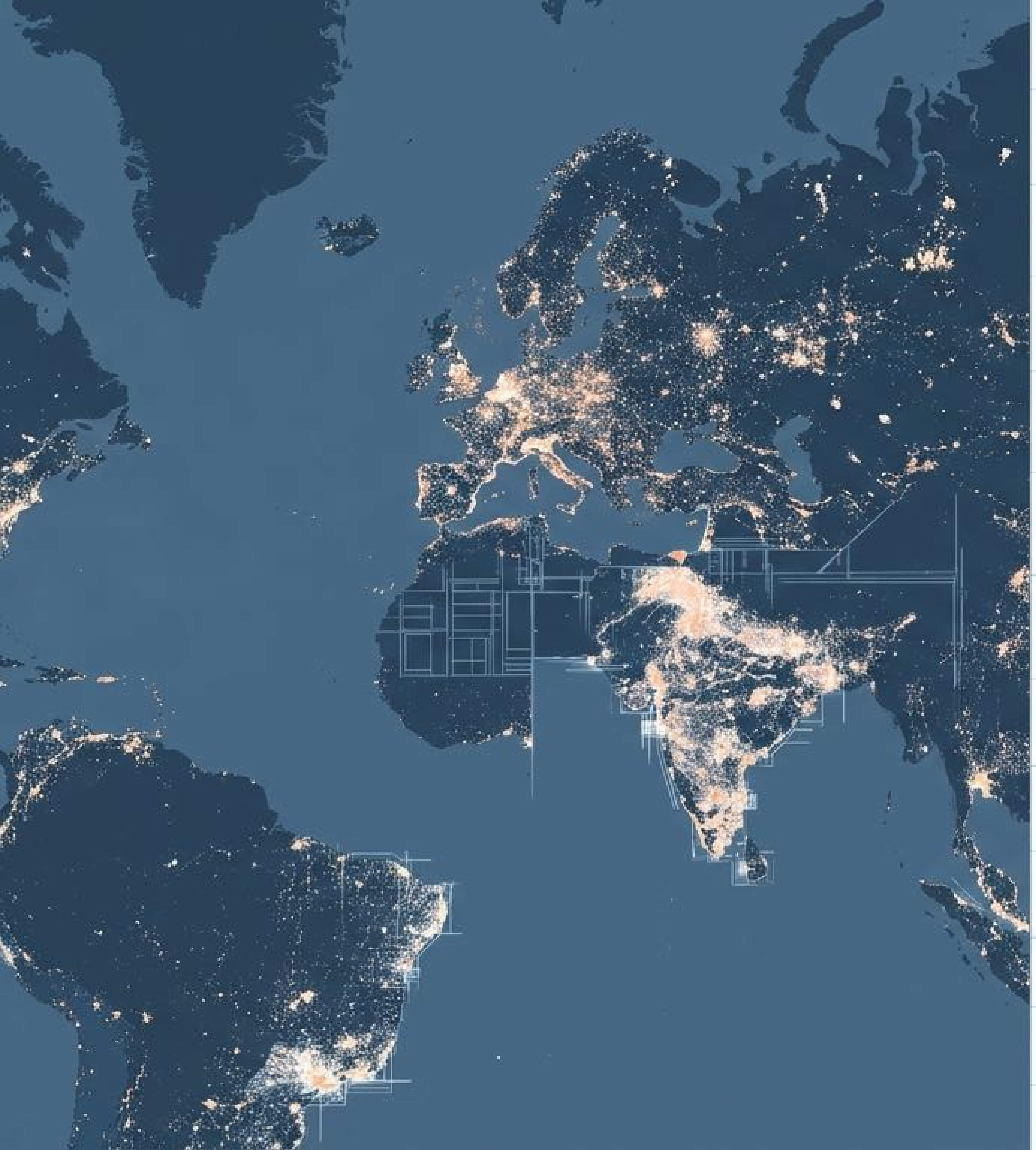
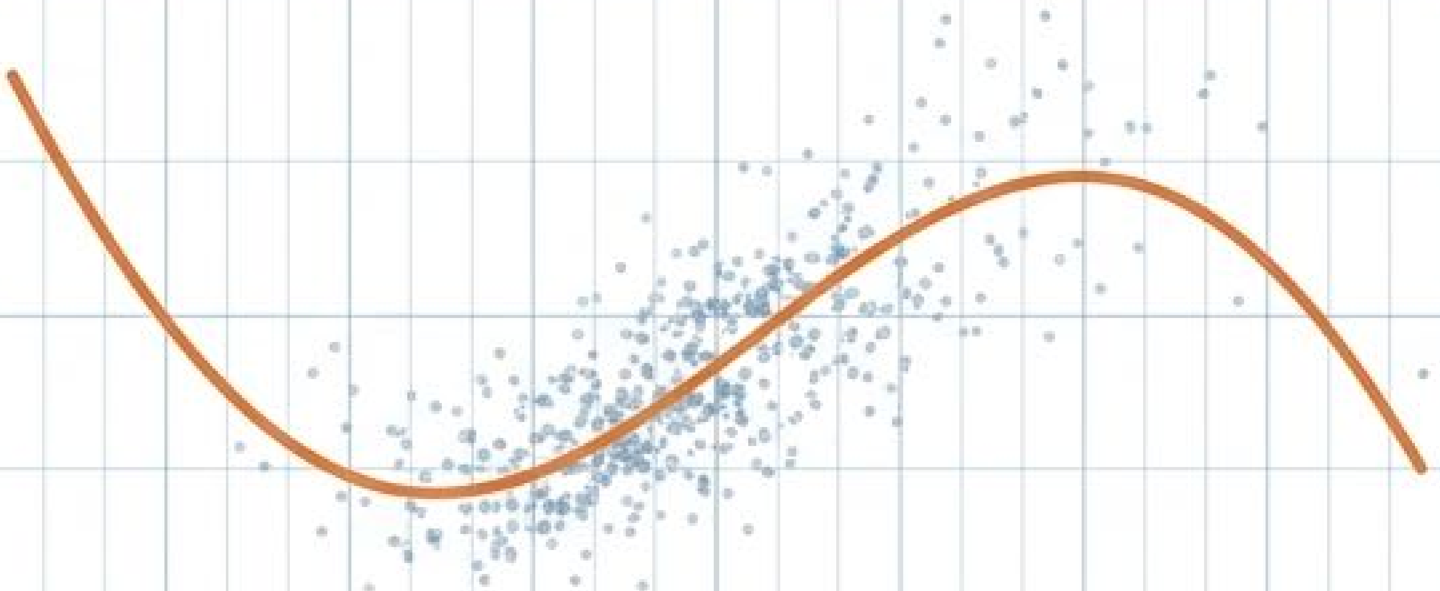


Regional Inequality from Outer Space: An Empirical Study Guide

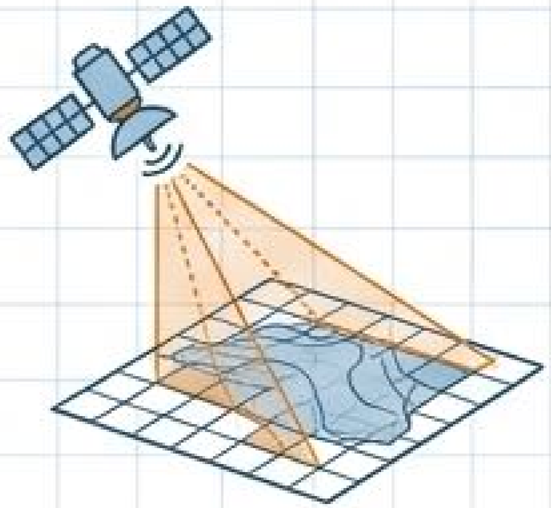
Predicting GDP from Nighttime Lights, Building Inequality Indices, and Estimating the Spatial Kuznets Curve in Python



A methodological breakdown based on the empirical pipeline of Lessmann & Seidel (2017).

STAGE 1: Predict

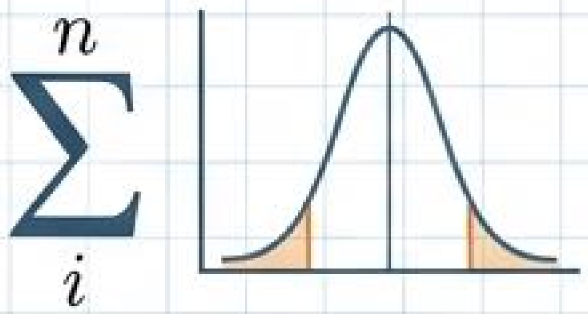
Light to Income



Calibrate OLS/RE models to translate physical lumens into regional GDP per capita.

STAGE 2: Construct

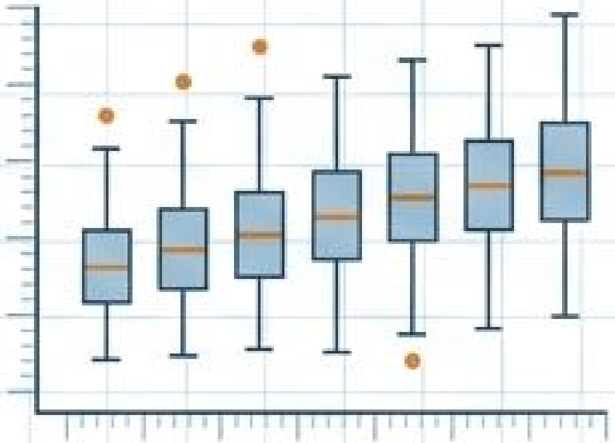
Income to Inequality



Compute population-weighted Gini and generalized-entropy indices from first principles.

STAGE 3: Explore

Inequality Dynamics



Map distributions across countries and time to identify global macroeconomic trends.

STAGE 4: Estimate

The Curve & Drivers

Identify the spatial Kuznets curve using panel fixed effects and test robustness via Conley standard errors.

Core Lexicon & Analogies

Nighttime Lights

Proxy for economic output. Brightness correlates with roads, infrastructure, and human activity.

Analogy: Guessing household wealth from an electricity bill—useful on average, though wrong for off-grid farmers.

Light-to-GDP Elasticity

The % change in predicted regional GDP per 1% change in light. Calibrated at a baseline of 0.102.

Analogy: The exchange rate between “lumens” and “dollars”.

Population Weights

Regions count proportionally by population rather than equally, tying the index to lived human experience.

Analogy: A class grade weighted by attendance, or voting by headcount vs. geographic district.

Spatial Kuznets Curve

Inequality follows an inverted-U or N-shape trajectory against logarithmic national income.

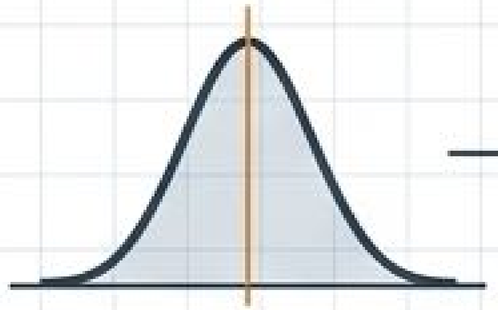
Analogy: A country's internal road trip where gaps widen leaving the village, narrow at the city, and fray in the suburbs.

The Data Landscape: Segregating the Panels

Panel A: Region-Year

Shape: 5,258 rows, 1,504 regions

Primary Purpose: Calibrate the lights model

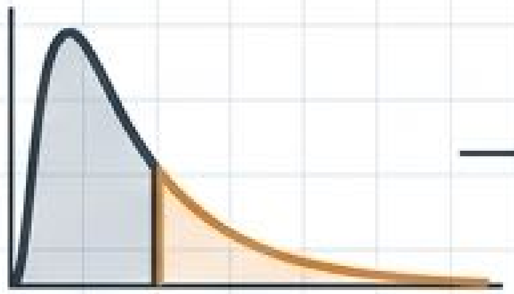


Distributions: Log light and log income are bell-shaped. Taking logs mathematically tames the heavy right skew.

Panel B: Country-Period

Shape: 3,675 rows, 180 countries, 1992–2012

Primary Purpose: Measure inequality and run Kuznets regressions



Distributions: Regional Gini is right-skewed and bounded at zero (mean 0.064, max 0.163).

Step 1: The Prediction Model

$$y_r = \beta_0 + \beta_1 l_r + \beta_2 g_c + \gamma' X_r + \mu_g + \tau_s + \varepsilon_r$$

Log regional income (The Target)

Log light per pixel. The key parameter calibrated at $\beta_1 = 0.102$

National log income (Absorbs overarching scale)

Fixed effects controlling for World Region and Satellite Generation

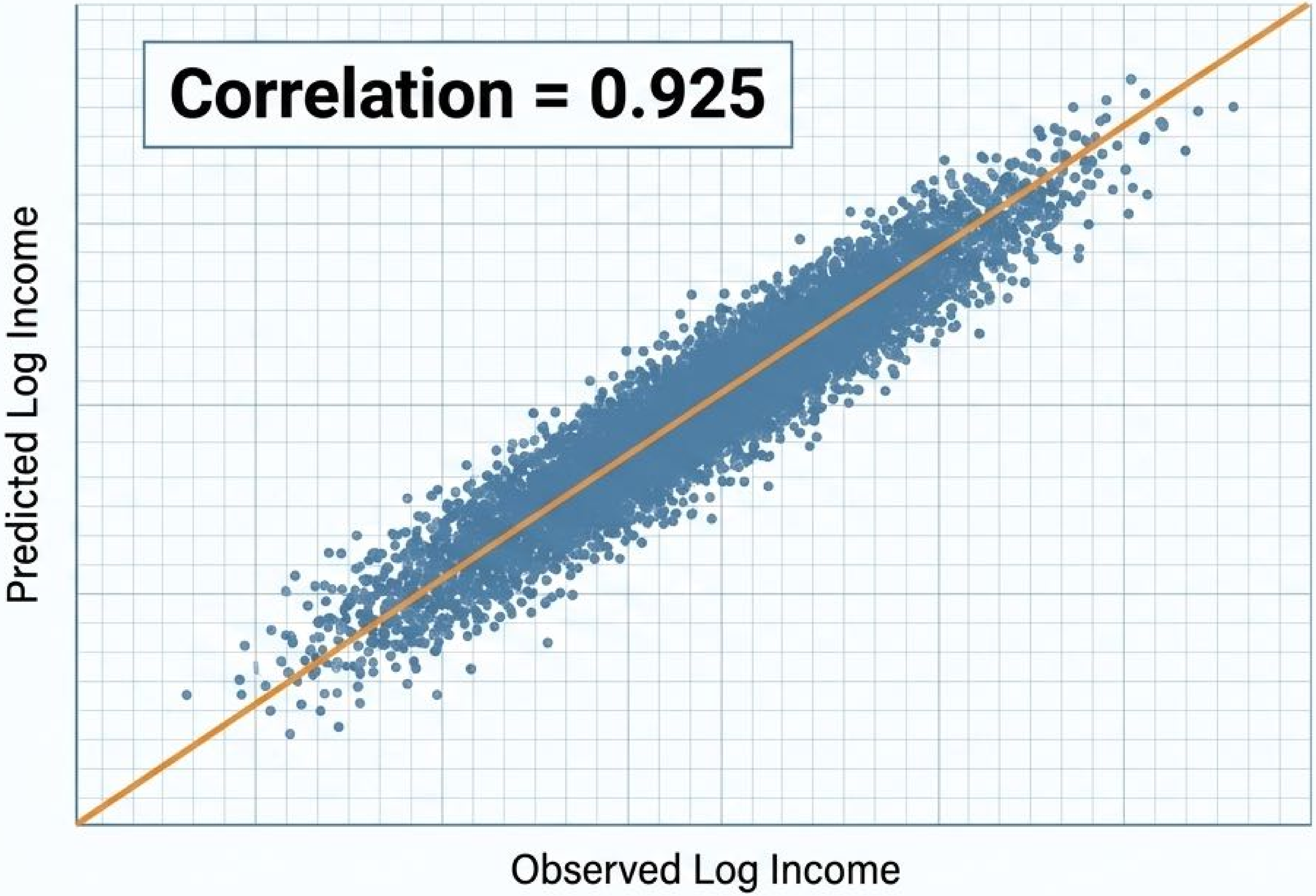
```
Implementation Note: Estimated using PyFixest (feols) with standard errors
clustered by country (CRV1).
```

Estimator Diagnostic: Fixed vs. Random Effects

	Fixed Effects (FE)	Random Effects (RE)
Variation Exploited	Only within-region changes	Between-region differences + within-region changes
Calculated Light Elasticity	$\beta_1 = 0.049$	$\boldsymbol{\beta_1 = 0.102}$
National GDP Elasticity	Absorbed	$\beta_2 = 0.889$ (Tracks almost one-for-one)

Core Insight: Random effects (linearmodels.RandomEffects in Python) are preferred for the calibration/prediction step. Once national income absorbs the overall scale, the between-region differences are exactly where nighttime lights provide crucial spatial detail.

Evaluating the Prediction Generalizability



Methodological Takeaway:

The model does not simply memorize one income band; it successfully generalizes across **four orders of magnitude.**

This high correlation licenses the critical next empirical move: applying these coefficients to **tens of thousands** of regions with zero official statistics to build a **complete global map.**

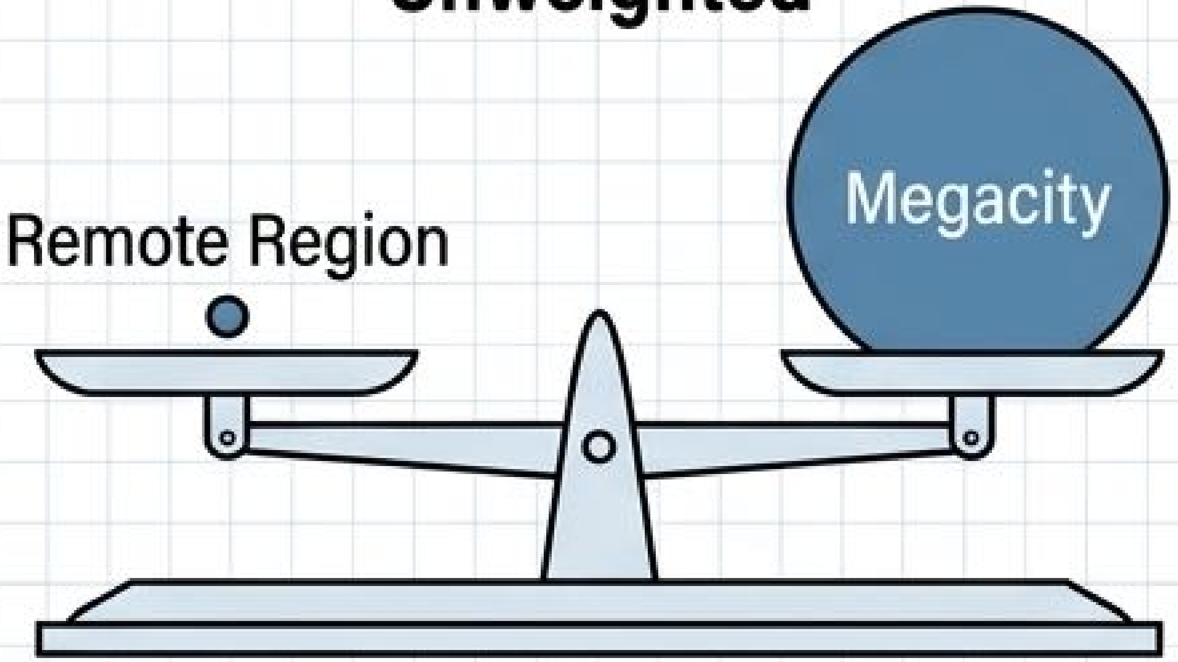
Step 2: Inequality Index Construction

Mathematical Theory		Programmatic Execution
Weighted Mean	$\bar{y} = \frac{\sum_i w_i y_i}{\sum_i w_i}$	`np.average(y, weights=w)`
Population Shares	$p_i = \frac{w_i}{\sum_j w_j}$	`w / np.sum(w)`
Regional Gini	$G = \frac{\sum_i \sum_j w_i w_j \lvert y_i - y_j \rvert}{2\left(\sum_i w_i\right)^2 \bar{y}}$	`absolute_diff = np.abs(np.subtract.outer(y, y))` **Warning Callout:** Note the absolute difference |y_i - y_j| is summed over all pairs, not a product. This absolute difference matrix logic is the classic trap when hand-coding a weighted Gini.

Parallel indices GE(0), GE(1), and CV also constructed to ensure results are not index-dependent.

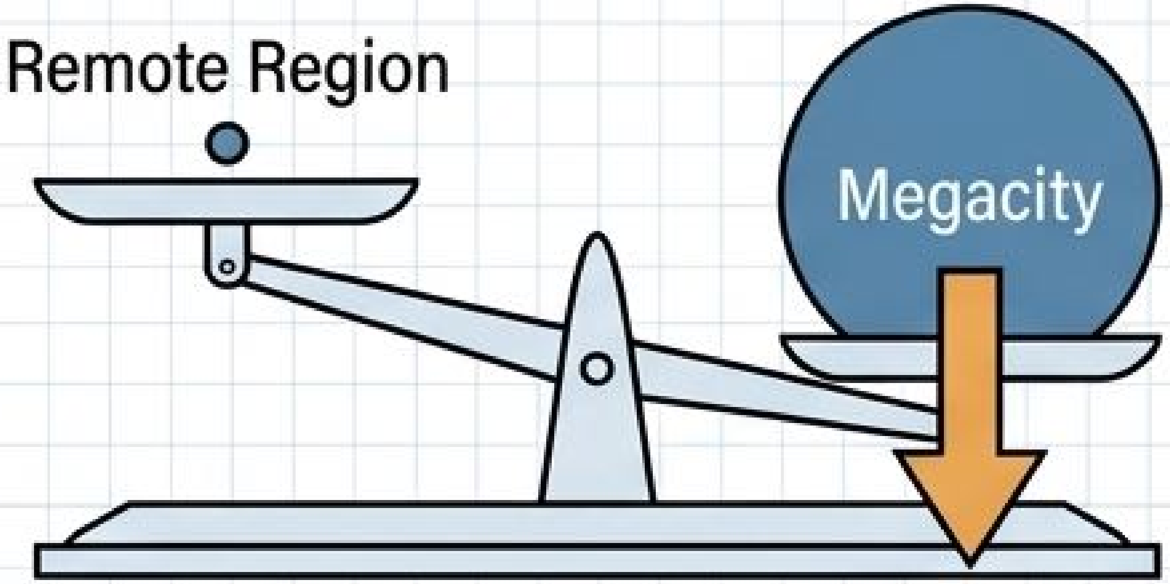
The Impact of Population Weighting

Unweighted



Tiny region and megacity counted equally.

Weighted



Megacity population dominance tips scale.

Methodological Effect

Weighted and unweighted Gini correlate at only 0.75. Weighting lowers average measured inequality by ~0.0034 because extreme, tiny regions lose statistical influence.

Worked Example: Germany 2010

Population-weighted Gini is just 0.028. German regions are close in income, and highly populous regions sit near the national mean, firmly anchoring the index.

Rule for Practice

Always report weighting.

Weighting “by people” rather than “by geography” is the policy-relevant choice.

The Payoff: Validating the Construction Pipeline

Naïve Approach
Raw light density Gini correlation with observed GDP

0.21

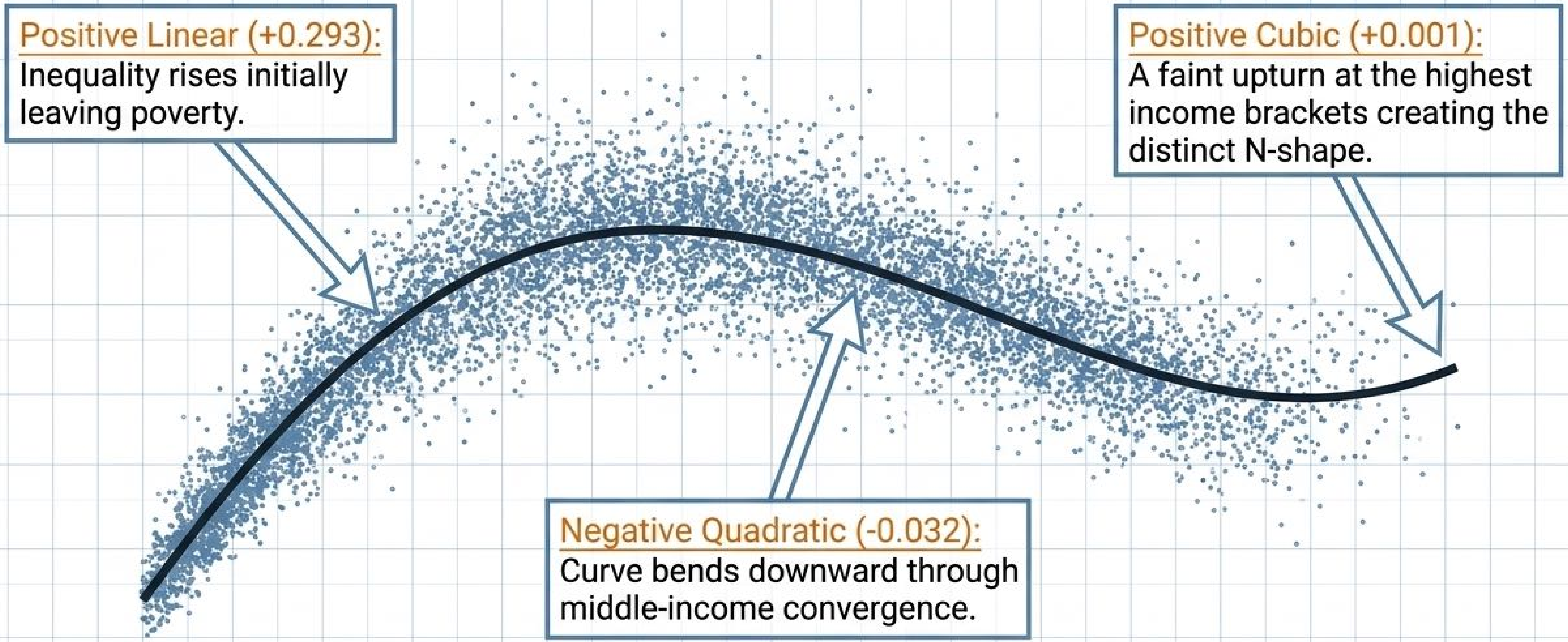
Empirical Approach
Predicted income Gini correlation with observed GDP

0.49

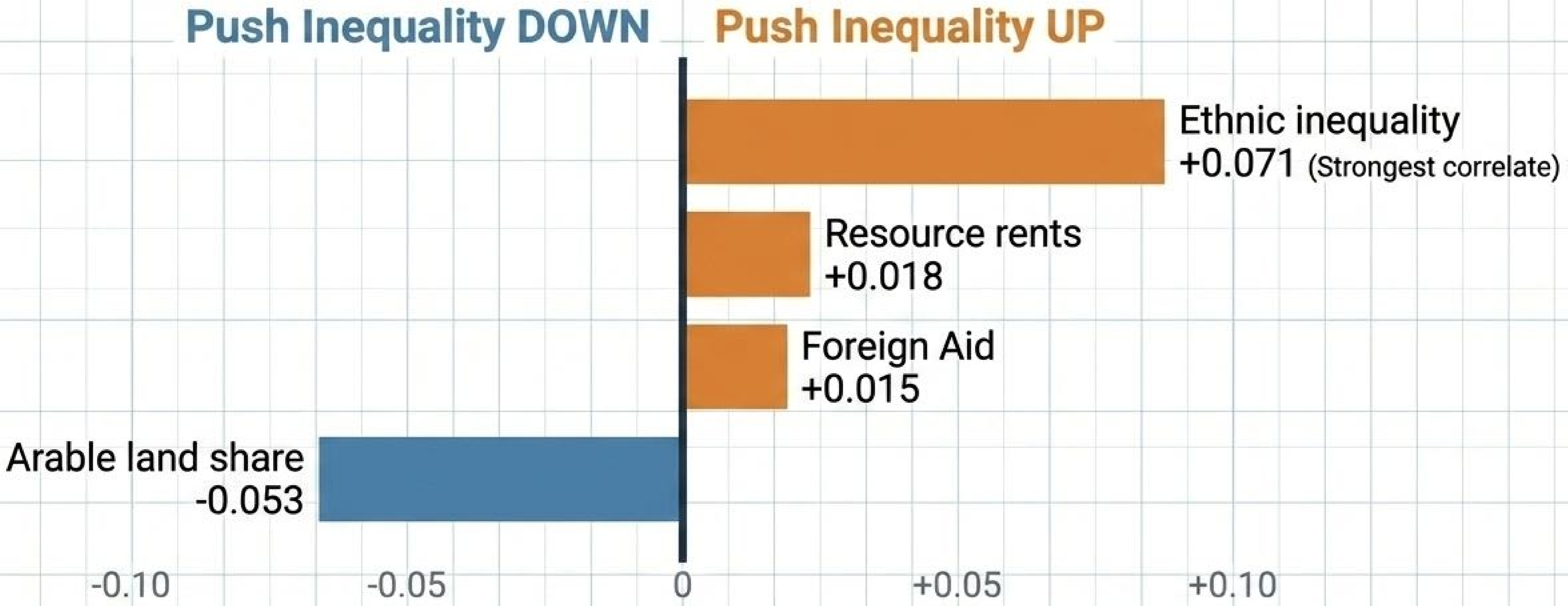
Synthesis Insight:
Predicting income first—rather than bluntly equating raw brightness with income—more than **doubles** the accuracy of the final inequality measure. The econometric heavy lifting of Step 1 was mandatory to validate Step 2.

Step 3: The Regional Kuznets Curve

$$\mathrm{GINIW}_{ct} = \beta_1 \ln Y + \beta_2 \ln Y^2 + \beta_3 \ln Y^3 + \alpha_c + \delta_t + \varepsilon$$



Structural Drivers of Inequality



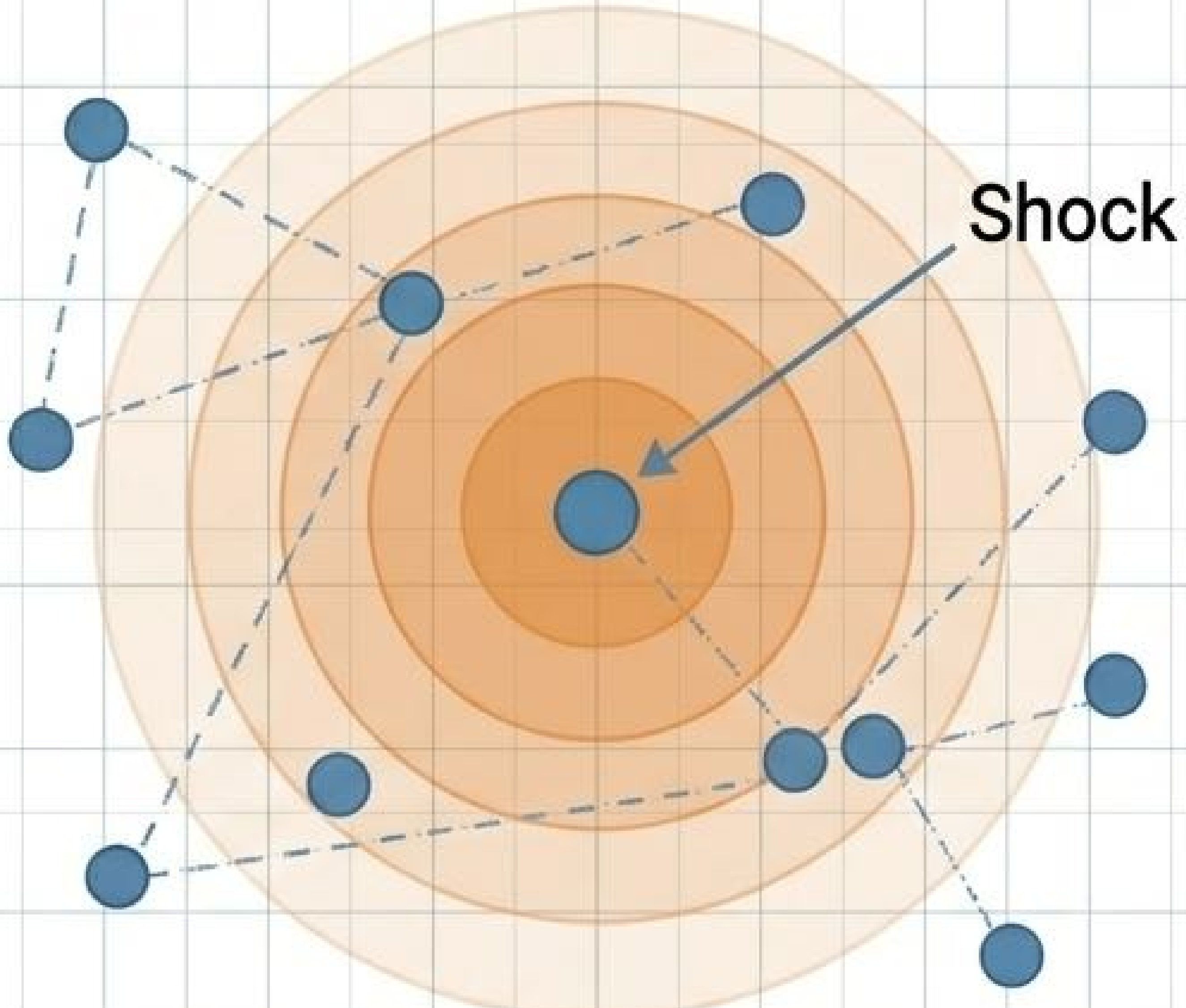
Methodological Caveat: These are descriptive associations estimated with country/period fixed effects, not causal identifications.

Step 4: Spatial Robustness & Conley Standard Errors

The Problem

Regions are **not independent.**

Ignoring spatial spillovers artificially shrinks standard errors, falsely inflating confidence.



The Analogy

Counting independent witnesses.

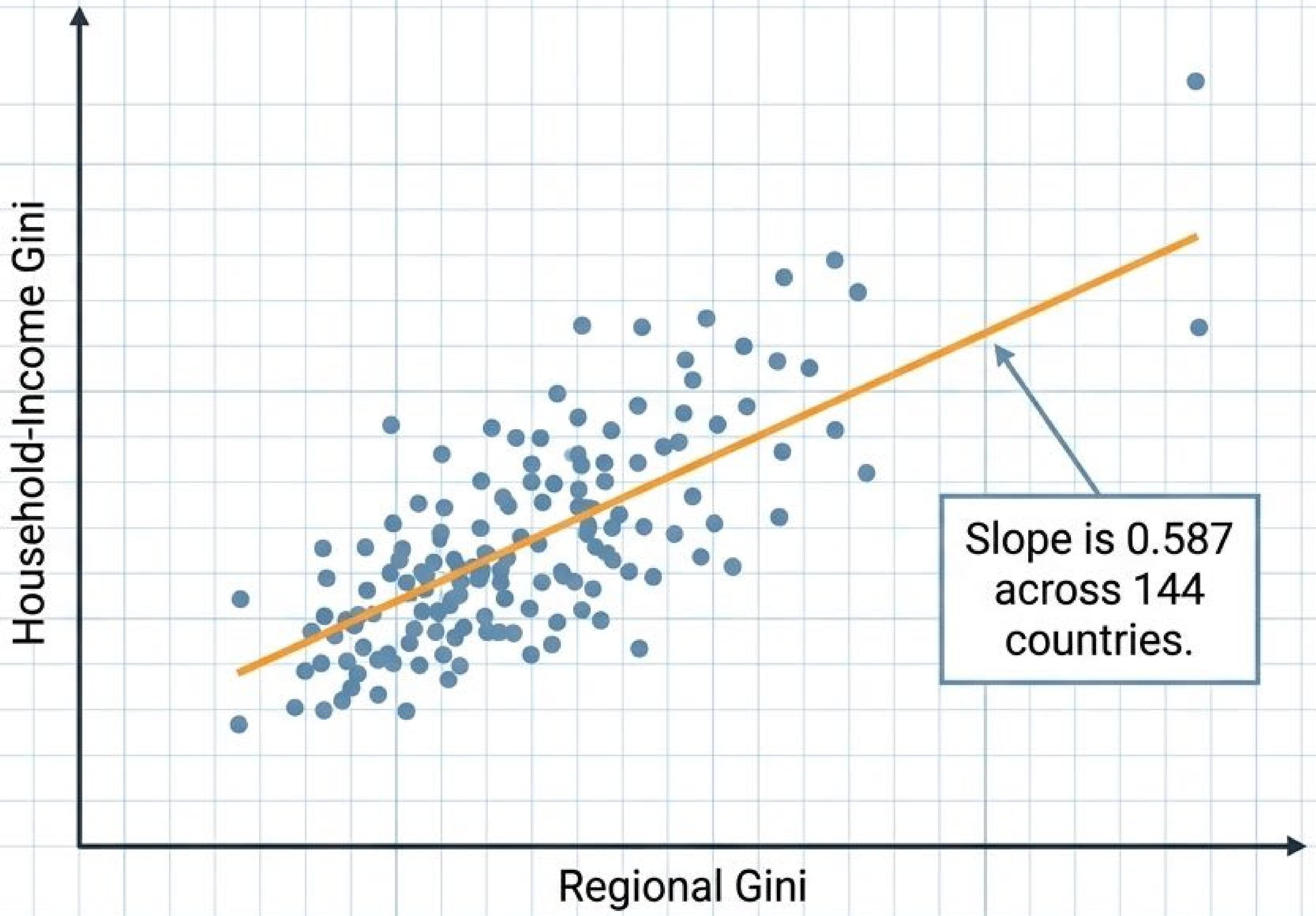
If ten people hear the same rumor, you have one fact, not ten.

Conley errors explicitly discount correlated neighbors.

The Result

Applying Conley SEs (1,000–5,000 km radii) doubles to triples the standard error (from 0.013 to 0.037). However, the core elasticity ($\beta_1 = 0.190$) easily survives with a t-statistic > 5.

Synthesis: Regional vs. Personal Inequality



Synthesis Insight

- Gaps between places heavily reflect gaps between people.
- The internal economic geography of a country is intimately bound up with its human geography.
- Policies that successfully narrow regional spatial gaps inherently act as macro-distributional policies for individual citizens.

Methodological Blueprint: Final Takeaways

Core Insights	Methodological Limitations	Next Steps for Researchers
▪ Lights reliably predict income (0.925 correlation via RE model). ▪ The regional Kuznets curve is robustly N-shaped. ▪ Conley SEs are mandatory for spatial regressions to account for spillover errors.	▪ Calculated income figures are predictions—accurate in aggregate but potentially flawed for unusual individual regions. ▪ Determinant models represent descriptive associations with fixed effects, lacking explicit causal identification.	▪ Swap legacy DMSP data for modern VIIRS satellite data for post-2012 series. ▪ Project the fitted coefficients globally to generate complete choropleth maps for data-desert regions.