Spatial Correlation

Applied Economics Research Course

Bas Machielsen

Motivation: The Persistence Puzzle

Historical Persistence in Economics

- A large literature asks: do historical events shape modern economic outcomes?
- Examples:
  - Colonial institutions $\rightarrow$ modern income (Acemoglu, Johnson & Robinson, 2001)
  - Slave trade $\rightarrow$ modern distrust (Nunn & Wantchekon, 2011)
  - Plough use $\rightarrow$ gender norms today (Alesina, Giuliano & Nunn, 2013)
  - Roman roads $\rightarrow$ modern economic activity
- These studies typically find **strong, statistically significant** results
- t-statistics of 4, 5, even 6 are common

The Paradox

- Conley & Kelly (2025) notice something striking:
  "The historical persistence literature reports an unusual prevalence of large t-statistics."
- If history truly determines modern outcomes, some persistence is plausible...
- ...but t-statistics this large imply implausibly precise estimates
- Across **30 seminal papers**, rejection rates are far above what we'd expect
- **The claim:** standard errors may be systematically too small

Why Might Standard Errors Be Too Small?

- Regression with geographic units: countries, regions, municipalities
- These units are **not independent**: nearby places share
  - Climate, geography, culture
  - Historical shocks
  - Institutional spillovers
- When errors are **spatially correlated**, observations carry less independent information than their count implies
- Classical OLS standard errors treat all observations as independent $\rightarrow$ understate uncertainty
- This is the **core problem** we will study today

What is Spatial Correlation?

Tobler's First Law of Geography

> *"Everything is related to everything else, but near things are more related than distant things."*

- Waldo Tobler (1970) — arguably the most cited sentence in geography
- It is an empirical regularity, not a theorem
- But it holds remarkably widely:
  - Income, institutions, culture cluster geographically
  - Soils, climate, vegetation cluster geographically
  - Historical shocks (wars, plagues, colonisation) have spatial footprints

Implication for econometrics: residuals from spatial regressions will typically be spatially correlated.

Formal Definition

Consider a regression with n geographic units:

$$y_i = \mathbf{x}_i'\beta + u_i, \quad i = 1, \ldots, n$$

Spatial correlation means:

$$\text{Cov}(u_i, u_j) \neq 0 \quad \text{for units } i, j \text{ that are spatially close}$$

This is analogous to **serial correlation** in time series — but across space rather than time.

Key difference from time series: space is two-dimensional and has no natural ordering. There is no single "lag".

The Spatial Weight Matrix

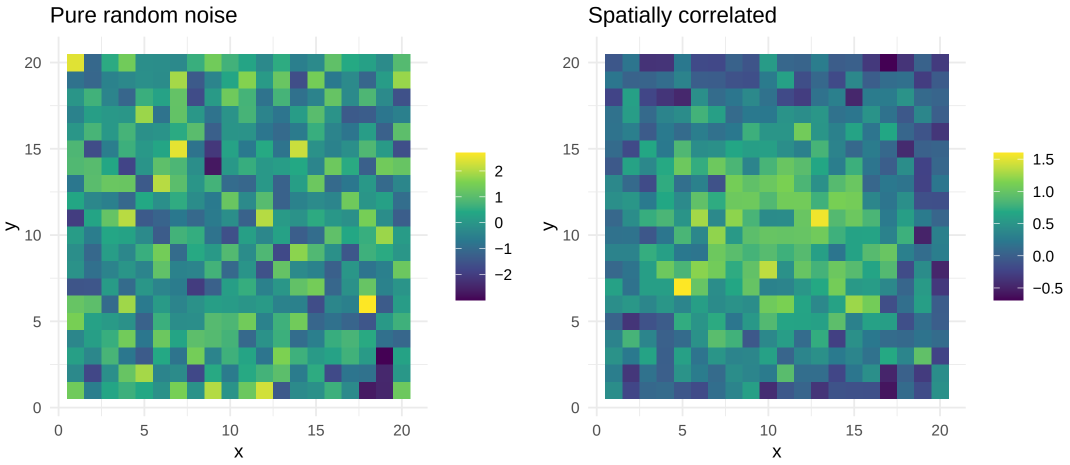
- To formalise "who is close to whom", we use a **spatial weight matrix** W
- W is an $n \times n$ matrix where w_{ij} encodes the spatial relationship between units i and j
- Common choices:
  - **Contiguity:** $w_{ij} = 1$ if i and j share a border, 0 otherwise
  - **Distance-based:** $w_{ij} = d_{ij}^{-1}$ (inverse distance)
  - k**-nearest neighbours:** $w_{ij} = 1$ if j is among i's k nearest neighbours
- By convention: $w_{ii} = 0$ (no self-neighbours)
- Usually **row-standardised**: each row sums to 1, so $\sum_j w_{ij} = 1$

Spatial Lag

- The **spatial lag** of a variable y is Wy
- Row-standardised: $(Wy)_i = \sum_j w_{ij} y_j$ — a weighted average of neighbours' values
- Spatial correlation means y_i and $(Wy)_i$ are correlated
- This is directly analogous to y_t and y_{t-1} in time series

Intuition: if your neighbours have high income, you probably do too. If the regression residual is positive for Paris, it is probably positive for Lyon too.

Visualising Spatial Correlation



Why Does Geography Cluster?

Spatial correlation in regression residuals can arise from:

1. **Omitted variables** that are themselves spatially distributed (climate, soil quality, disease environment)
2. **Spatial spillovers**: outcomes in one location affect neighbouring locations (trade, migration, institutional imitation)
3. **Common shocks**: historical events affect entire regions (wars, epidemics, colonial borders)
4. **Measurement error** that is geographically structured

Important: even after controlling for observed spatial covariates, residuals are often spatially correlated. This is the relevant concern for inference.

Consequences for OLS

OLS Assumptions Revisited

The standard OLS variance formula assumes **spherical errors**:

$$E[\mathbf{uu}'] = \sigma^2\mathbf{I}_n$$

Under this assumption:

$$\widehat{\mathsf{Var}}(\hat{\beta}) = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

When this holds, OLS is **BLUE** (Gauss-Markov theorem) and standard errors are valid.

What Breaks with Spatial Correlation?

When $E[\mathbf{u}\mathbf{u}'] = \Sigma \neq \sigma^2 \mathbf{I}$, the **true** variance of $\hat{\beta}$ is:

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

But we compute $s^2(\mathbf{X}'\mathbf{X})^{-1}$ instead. The **ratio** depends on the off-diagonal elements of Σ.

- If $\text{Cov}(u_i, u_j) > 0$ for nearby units (the typical case), classical SEs **understate** true uncertainty
- t-statistics are **inflated** — we reject H_0 too often
- OLS is still **unbiased** for β, but **inference is invalid**

Effective Sample Size

Intuition: imagine $n = 100$ municipalities, but due to strong spatial clustering they behave like 10 independent "blocs". Then:

- We effectively have $n_{\text{eff}} = 10$ independent observations, not 100
- Standard errors should scale as $1/\sqrt{10}$, not $1/\sqrt{100}$
- Classical SEs are too small by a factor of $\sqrt{10}$, inflating t-statistics by $\sqrt{10} \approx 3.2$

General formula for effective sample size under AR(1)-like spatial dependence with correlation ρ:

$$n_{\text{eff}} \approx n \cdot \frac{1-\rho}{1+\rho}$$

At $\rho = 0.5$: $n_{\text{eff}} \approx n/3$. At $\rho = 0.8$: $n_{\text{eff}} \approx n/9$.

Simulating Spatial Correlation

The formula borrows the AR(1) effective sample size from time series, where ρ is the correlation between adjacent errors and decays geometrically: Corr$(u_t, u_{t+k}) = \rho^k$.

The spatial analogue is the **Spatial Error Model (SEM)**:

$$\mathbf{u} = \rho W \mathbf{u} + \varepsilon, \quad \varepsilon \sim N(\mathbf{0}, \sigma^2 I)$$

Solving: $\mathbf{u} = (I - \rho W)^{-1}\varepsilon$

- $\rho = 0$: errors are i.i.d. — standard OLS inference is valid
- $\rho > 0$: neighbouring units have positively correlated errors
- For row-standardised W, ρ is approximately the pairwise correlation between a unit and its neighbours

Caveat: $n_{\text{eff}} \approx n\frac{1-\rho}{1+\rho}$ is exact for a 1D chain, approximate for 2D space (variable neighbour counts, no natural ordering).

Simulating a DGP with a Given ρ

```
library(spdep); library(Matrix)

W_mat <- listw2mat(W)           # convert listw → dense matrix
n     <- nrow(W_mat)
I_n   <- diag(n)

rho <- 0.5
eps <- rnorm(n, sd = 1)

# Draw spatially correlated errors from the SEM
u <- solve(I_n - rho * W_mat) %*% eps

# Generate outcome under H0: beta = 0
y_sim <- u
```

When $\rho = 0$: $u = \varepsilon$ (independent noise). As $\rho \to 1$: $(I - \rho W)$ approaches singularity and errors cluster strongly across neighbours.

The Conley & Kelly (2025) Simulation

They demonstrate the problem starkly:

1. Draw a **"historical variable"** x: spatial noise with a north-south trend on a unit square ($n = 250$)
2. Draw an **independent "modern outcome"** y: independent spatial noise
3. Regress y on x — by construction, $\beta = 0$
4. Repeat many times

Results without spatial correction:

- Nominal 5% test rejects the true null **27% of the time**
- In one draw: $t = -3.8$, nominal $p = 0.0001$
- Placebo p-value (the true one): $p = 0.10$

Lesson: spatially structured data can generate convincingly significant results from pure noise.

Rejection Rates Under Spatial Correlation

Method	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$
HC (robust, no spatial correction)	5%	25%	45%
Conley SEs (distance cutoff)	5%	10%	15%
Large clusters (BCH, $k = 6$)	5%	7%	10%
Spatial basis regression	5%	6%	7%

Source: Conley & Kelly (2025), Figure 2. Nominal size: 5%.

All methods overreject, but **spatial basis regression** and **large cluster inference** perform best.

Testing for Spatial Correlation: Moran's I

Moran's I — Definition

Moran's I is the most widely used test for **global** spatial autocorrelation:

$$I = \frac{n}{S_0} \cdot \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

where $S_0 = \sum_i \sum_j w_{ij}$ is the sum of all weights.

- Proposed by Patrick Moran (1950)
- Defined for any variable y and weight matrix W
- In practice, we apply it to **OLS residuals** to test for model misspecification

Moran's I — Application to Residuals

Instead of the raw outcome variable, you plug in $\hat{u}_i$ — the OLS residual for unit i. Since OLS residuals sum to zero (when there's an intercept), $\bar{y} = \bar{\hat{u}} = 0$, so the formula simplifies to:

$$I = \frac{n}{S_0} \cdot \frac{\sum_i \sum_j w_{ij}, \hat{u}_i, \hat{u}_j}{\sum_i \hat{u}_i^2}$$

The weights w_{ij} represent the same spatial weight matrix defined earlier in the slides (contiguity, distance-based, k-NN).

Moran's I — Intuition

Moran's I is the spatial analogue of Pearson's correlation coefficient.

Rewrite it as:

$$I \propto \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

This is a weighted sum of cross-products $(y_i - \bar{y})(y_j - \bar{y})$ — exactly what we compute in a correlation, but between unit i and its spatial neighbours.

- $I > 0$: **positive spatial autocorrelation** — similar values cluster together
- $I < 0$: **negative spatial autocorrelation** — dissimilar values are neighbours (rare)
- $I \approx 0$: **spatial randomness**

Moran's I — Inference

Expected value under the null (spatial randomness):

$$E[I] = \frac{-1}{n-1} \approx 0 \text{ for large } n$$

Variance under normality assumption:

$$\mathrm{Var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)} - E[I]^2$$

where $S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$, $S_2 = \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2$.

In practice: use a **permutation test** — randomly permute labels across locations, recompute I many times, compare observed I to the permutation distribution.

The Moran Scatterplot

- Plot $y_i - \bar{y}$ (horizontal axis) against the spatial lag $(Wy)_i - \bar{y}$ (vertical axis)
- The **slope** of the best-fit line is Moran's I
- Four quadrants:
  - **High-High (HH):** high value surrounded by high neighbours — spatial cluster
  - **Low-Low (LL):** low value surrounded by low neighbours — spatial cluster
  - **High-Low (HL):** high value surrounded by low neighbours — spatial outlier
  - **Low-High (LH):** low value surrounded by high neighbours — spatial outlier

The scatterplot decomposes the global I into local contributions.

Computing Moran's I in R — Data

Computing W for Moran's I happens in two steps:

- Build a neighbours list from a polygon list
- Build a W matrix from a polygon list

```
# Build spatial weight matrix from contiguous neighbours
nb <- poly2nb(nl, queen = TRUE)                    # queen contiguity
W  <- nb2listw(nb, style = "W", zero.policy = TRUE) # row-standardise
```

Computing Moran's I in R — Test

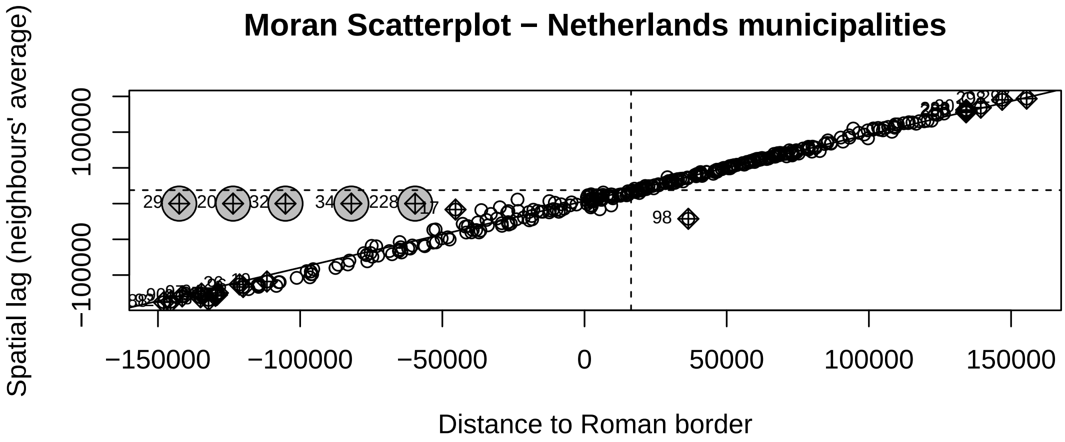
```r
# Test Moran's I on distance to Roman Empire border
moran.test(nl$distance_to_border, W, zero.policy = TRUE)
```

```

	Moran I test under randomisation

data:  nl$distance_to_border
weights: W
n reduced by no-neighbour observations

Moran I statistic standard deviate = 27.608, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.925703021      -0.002865330       0.001131214
```

Moran Scatterplot in R

```r
moran.plot(
  nl$distance_to_border,
  W,
  xlab = "Distance to Roman border",
  ylab = "Spatial lag (neighbours' average)",
  main = "Moran Scatterplot - Netherlands municipalities",
  zero.policy = TRUE
)
```

Moran Scatterplot — Plot



Implications of Moran's I

For variable selection:

- High I in y or x alone does not invalidate OLS — we care about I in **residuals**

For residual diagnostics:

- After fitting a model, compute Moran's I on residuals $\hat{u}_i$
- Significant $I \rightarrow$ spatial patterns remain unexplained $\rightarrow$ SEs unreliable

For interpretation:

- Moran's I is a **global** summary — it detects overall clustering
- It does **not** tell you where the clusters are (for that: LISA maps)
- A non-significant I does not rule out local spatial patterns

Moran's I on Regression Residuals

```r
# Simple OLS: does Roman Empire proximity predict municipal area?
model <- lm(area ~ distance_to_border, data = nl)
residuals_nl <- residuals(model)
# Moran's I on residuals
moran.test(residuals_nl, W, zero.policy = TRUE)
```

```
    Moran I test under randomisation

data:  residuals_nl
weights: W
n reduced by no-neighbour observations

Moran I statistic standard deviate = 5.0232, p-value = 2.54e-07
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
```

Other Spatial Diagnostics

Geary's C

An alternative to Moran's I, focusing on **local** differences:

$$C = \frac{(n-1)\sum_i \sum_j w_{ij}(y_i - y_j)^2}{2S_0 \sum_i (y_i - \bar{y})^2}$$

- $C \in (0, 2)$; expected value under null: $C = 1$
- $C < 1$: positive spatial autocorrelation
- $C > 1$: negative spatial autocorrelation
- More sensitive to **local** spatial autocorrelation than Moran's I
- Moran's I and Geary's C are approximately inversely related: $I \approx 1 - C$

Local Moran's I (LISA)

Local Indicators of Spatial Association (Anselin, 1995) decompose global I into unit-level contributions:

$$I_i = \frac{(y_i - \bar{y})}{m_2} \sum_j w_{ij}(y_j - \bar{y}), \quad m_2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2$$

Note: $I = \sum_i I_i / S_0$.

Uses of LISA:

- Identify specific **spatial clusters** (HH, LL) and **outliers** (HL, LH)
- Detect local non-stationarity — global $I \approx 0$ but local clusters exist
- Map significance of each unit's I_i after multiple-testing correction

Classical Correction: Conley Standard Errors

The Idea

- Allow errors to be correlated across units within a **distance bandwidth** d
- Beyond d: errors treated as uncorrelated
- A **HAC** (heteroskedasticity and autocorrelation consistent) estimator for spatial data

The sandwich variance estimator:

$$\widehat{\text{Var}}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{S}}(\mathbf{X}'\mathbf{X})^{-1}$$

$$\hat{\mathbf{S}} = \sum_i \sum_j K\left(\frac{d_{ij}}{d}\right) \hat{u}_i \hat{u}_j \mathbf{x}_i \mathbf{x}_j'$$

where $K(\cdot)$ is a kernel (e.g., Bartlett) that downweights distant pairs.

- Proposed by Conley (1999) — a standard tool in spatial econometrics

Limitations of Conley Standard Errors

Conley SEs are **not sufficient** (Conley & Kelly, 2025):

- Require choosing a bandwidth d — results sensitive to this choice
- Perform poorly when spatial correlation is **large** or **long-range**
- Assume the spatial correlation structure decays with distance — may not hold for trends
- At $\rho = 0.5$: rejection rate is **15%** vs nominal 5%

Root cause: Conley SEs correct the variance estimator but do not address **spatial trends** in the data. If x and y both have a north-south gradient, their correlation looks significant even if the gradient in y has nothing to do with x.

Conley & Kelly (2025): Spatial Basis Regression

The Core Idea

Problem: x and y may share common spatial trends that create spurious correlation.

Solution: control for those trends directly by adding **flexible spatial controls** to the regression.

$$y_i = \beta x_i + f(\mathsf{lon}_i, \mathsf{lat}_i) + \varepsilon_i$$

where f is a smooth function of coordinates. The function absorbs:

- North-south gradients
- East-west gradients
- Other spatial trends

After removing spatial trends from both y and x, the residual variation used to identify β is less spatially correlated.

Tensor Product B-Splines

Conley & Kelly use **tensor product B-splines** to estimate f:

$$f(\mathrm{lon}, \mathrm{lat}) = \sum_{j=1}^{J} \sum_{k=1}^{K} \tau_{jk}\, a_j(\mathrm{lon}) \cdot b_k(\mathrm{lat})$$

- $a_j(\cdot)$ and $b_k(\cdot)$ are **B-spline basis functions** (piecewise polynomial)
- The number of basis functions $J \times K$ controls flexibility
- **Basis dimension selected by BIC** — avoids overfitting while controlling for relevant trends

Interpretation: we are partitioning out a smooth spatial surface from both the outcome and the treatment variable. What remains is the within-surface variation.

Why Splines Work: An Analogy

Think of **regional fixed effects** — controlling for broad geographic areas:

- Absorbs common shocks within regions
- But regions are discrete; spatial correlation is continuous

B-splines generalise this:

- Absorb spatial trends **continuously** across the map
- More flexible than dummies, less arbitrary than polygons
- BIC ensures we don't over-absorb (which would consume the treatment's variation)

This is similar to adding **time trends** in a panel model to control for secular trends that affect all units.

Implementation in R

```r
library(spatInfer)
# Extract coordinates
coords <- st_centroid(nl) |> st_coordinates()
nl$Y <- coords[, 1]; nl$X <- coords[, 2]

# Spatial basis regression via spatInfer
model_ols   <- lm(area ~ distance_to_border, data = nl)
model_basis <- basis_regression(area ~ distance_to_border,
  splines=3,
  pc_num=5,
  clusters=11,
  df = nl |> st_drop_geometry())
```

Implementation in R (Cont.)

```
# Compare Moran's I on residuals
cat("Moran's I - without basis:",
    round(moran.test(residuals(model_ols),   W, zero.policy = TRUE)$statis
```

```
Moran's I - without basis: 5.023
```

```
cat("Moran's I - with basis:    ",
    round(moran.test(residuals(model_basis), W, zero.policy = TRUE)$statis
```

```
Moran's I - with basis:    1.033
```

What the Spatial Basis Achieves

After adding basis terms, Conley & Kelly (2025) find:

- **Moran statistics drop substantially** in residuals of most persistence studies
- t-statistics fall from, e.g., 5.25 to 1.5 (Becker & Woessmann study)
- The dramatic significance of many findings largely disappears

Key insight: the historical variable and the modern outcome often share the **same spatial surface** — a confound, not a causal mechanism. The spatial basis absorbs this confound.

Large Cluster Inference

BCH: Bester, Conley & Hansen (2011)

Idea: divide the n observations into k large **spatial clusters**, then apply cluster-robust inference.

- Use k-**medoids clustering** on coordinates to form geographic clusters
- Apply the cluster-robust variance estimator (Liang & Zeger, 1986) treating each cluster as a unit
- Each cluster acts like a "super-observation" — within-cluster correlation is unrestricted

Requirements:

- $k \geq 6$ clusters for reliable inference (need enough "units" for asymptotics)
- Clusters should be large enough to capture spatial correlation within them
- Tradeoff: more clusters $\rightarrow$ less conservative but more assumptions

IM: Ibragimov & Müller (2010)

A different approach: estimate $\hat{\beta}$ separately within each cluster, then test $H_0 : \beta = 0$ using the **distribution of cluster-specific estimates**.

1. Divide observations into k spatial clusters
2. Estimate $\hat{\beta}_1, \dots, \hat{\beta}_k$ within each cluster
3. Test $H_0 : \mu = 0$ where $\mu = E[\hat{\beta}_c]$ using a two-sided t-test with $k - 1$ degrees of freedom

Advantage: valid under very general within-cluster dependence. No assumptions about the correlation structure.

Advantage: intuitive — if the treatment effect is real, it should appear in most clusters. If only some clusters drive the result, that is informative.

Comparing the Approaches

Method	Key idea	Pros	Cons
HC SEs	Robust to heteroskedasticity	Simple	Ignores spatial correlation
Conley SEs	HAC with distance kernel	Well-known	Bandwidth choice; insufficient for trends
BCH	Large geographic clusters	No correlation assumptions	Needs $k \geq 6$; BIC choice
IM	Cluster-level t-test	Intuitive; very robust	Low power with few clusters
Spatial basis	Remove spatial trends + HC	Addresses root cause	May reduce power; BIC choice

Recommendation: use spatial basis regression as the default; corroborate with BCH or IM.

Diagnostic Tests

The Placebo Treatment Test

Core idea: if the finding is real, shuffling the treatment's spatial locations should destroy the result.

Procedure: 1. Estimate the model, record $\hat{\beta}_{\text{obs}}$ 2. Randomly permute the coordinates (or the treatment variable) across units 3. Re-estimate the model on the permuted data, record $\hat{\beta}_{\text{perm}}$ 4. Repeat many times (e.g., 1000) to build the **empirical null distribution** 5. The **placebo p-value** is the share of permutations where $|\hat{\beta}_{\text{perm}}| \geq |\hat{\beta}_{\text{obs}}|$

Why it works: permutation destroys the spatial structure of the treatment while preserving the spatial structure of the outcome. If the coefficient survives permutation, the spatial trend in x is correlated with the spatial trend in y even when we break the unit-level link.

The Synthetic Outcome Test

Core idea: generate fake outcomes that mimic the spatial structure of the true outcome, then check whether the observed $\hat{\beta}$ is unusual.

Procedure: 1. Estimate a **spatial noise model** for y: $\hat{y}_i = \hat{f}(\text{lon}_i, \text{lat}_i) + \hat{\varepsilon}_i$ 2. Simulate many synthetic outcomes y^* from this model (matching the spatial trend and noise structure) 3. Regress each y^* on the actual treatment $x \rightarrow$ synthetic coefficient distribution 4. Check where $\hat{\beta}_{\text{obs}}$ falls in this distribution

Advantage over placebo test: tests against the appropriate null (data-generating process), not just spatial randomisation.

Conley & Kelly Apply This to 30 Papers

They apply spatial basis regression, BCH, IM, and placebo tests to **30 seminal historical persistence papers**:

Correction method	Studies losing significance at 5%
None (original)	0 / 30
Conley SEs (generous bandwidth)	~5 / 30
Spatial basis + HC	~15 / 30
Placebo test	~15 / 30
BCH ($k = 4$)	~12 / 30
IM ($k = 4$)	~10 / 30

Approximate figures from Table 1 and Figures 3–4 of Conley & Kelly (2025).

Takeaway: roughly half of the literature's findings become insignificant under appropriate correction.

Interpreting the Placebo Results

- A high placebo p-value does **not** prove the effect is spurious
- It means: the spatial structure of x is consistent with producing the observed $\hat{\beta}$ by chance
- A low placebo p-value provides stronger evidence for a genuine effect

Why the persistence literature is particularly vulnerable:

- Historical variables (Roman roads, colonial institutions, settler mortality) are **strongly spatially structured**
- Modern outcomes (income, trust, gender norms) are also **strongly spatially structured**
- Both share climate, geography, and historical trajectories — a rich confound structure

Practical Guidance

A Checklist for Spatial Regressions

Step 1 — Visualise: - Map your outcome variable and key regressors - Map OLS residuals — look for spatial patterns

Step 2 — Test: - Compute Moran's I on residuals with a reasonable weight matrix - If I is significant, spatial correlation is present

Step 3 — Correct: - Add spatial basis terms (polynomial or B-splines in lon/lat) selected by BIC - Report results with and without spatial basis

Step 4 — Validate: - Run BCH or IM as a robustness check - Run placebo treatment test - If results are robust to all corrections, you have stronger evidence

What “Correcting” Does and Does Not Do

Does:

- Provide valid inference in the presence of spatial correlation
- Reduce inflated t-statistics to appropriate levels
- Distinguish genuine effects from spurious spatial correlations

Does not:

- Solve endogeneity (use IV for that)
- Guarantee a causal interpretation
- Replace the need for a credible identification strategy

Spatial basis regression absorbs the confound of shared spatial trends. It is not a substitute for — but a complement to — good research design.

When to Worry Most

Spatial correlation is a bigger concern when:

- The treatment variable has strong **geographic clustering** (colonial-era variables, climate zones, borders)
- The sample covers a **large geographic area** with potential for long-range trends
- There are few natural **discontinuities** to exploit (RDD-based designs are more robust)
- The outcome is a modern aggregation of processes that were themselves spatially structured

Reassuring cases: geographic RDD designs with tight bandwidths near a border naturally control for spatial trends — the basis regression is less critical.

Conclusion

Summary

The problem:

- Spatially structured data violates the i.i.d. assumption
- OLS standard errors are too small $\rightarrow$ inflated t-statistics
- Effect is quantitatively large: nominal 5% tests reject at 27–45% in simulations

The diagnostic:

- Moran's I on residuals detects spatial correlation
- Moran scatterplot reveals the nature of spatial dependence

Summary (Cont.)

The solutions (Conley & Kelly, 2025):

1. **Spatial basis regression** — remove shared spatial trends with flexible controls
2. **Large cluster inference** (BCH, IM) — treat geography as the clustering variable
3. **Placebo and synthetic outcome tests** — validate the spatial robustness of findings

Key References

- Conley, T.G. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*.
- Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*.
- Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*.
- Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*.
- Bester, C.A., Conley, T.G. & Hansen, C.B. (2011). Inference with Dependent Data Using Cluster Covariance Estimators. *Journal of Econometrics*.
- Ibragimov, R. & Müller, U.K. (2010). t-statistic Based Correlation and Heterogeneity Robust Inference. *Journal of Business & Economic Statistics*.
- **Conley, T.G. & Kelly, M. (2025). Standard Errors of Persistence. *Journal of International Economics*.**

Application: Chapman (2026)

Context: Democracy and Redistribution

Chapman, J. (2026). "Democracy, Redistribution, and Inequality: Evidence from the English Poor Law." *Journal of the European Economic Association*.

The question: Does income inequality shape how much democratisation increases redistribution?

- The **Meltzer-Richard (1981) model**: the median voter, poorer than the mean, demands more redistribution as inequality rises
- But testing this empirically is hard: inequality, democracy, and redistribution are all endogenous

The setting: 19th-century England. Local poor law unions (boards of guardians) controlled spending on poor relief – the main safety net before the modern welfare state. Funded by a proportional tax on rental values: the mechanism in Meltzer-Richard holds almost exactly.

The Identification Strategy

The exogenous shock: The 1894 Local Government Act (LGA) democratised all poor law union boards simultaneously, removing four elite protections in one stroke:

1. Property qualifications to stand as guardian
2. A graduated franchise (wealthy voters held up to 12 votes)
3. Unelected *ex officio* magistrates (local landowners) on boards
4. The absence of a secret ballot

Interpretation: Because all boards were treated at the same time and the reform was driven by national political logic – not local spending trends – the LGA is as good as exogenous to each union.

Empirical specification (difference-in-differences with heterogeneous treatment):

$$y_{it} = \beta \cdot \text{Inequality}_i \times \text{post1894}_t + \gamma_0 X'_{it} + T_t + \alpha_i + \varepsilon_{it}$$

$\beta > 0$ means more unequal unions increased spending more after democratisation.

Inequality Measure and Key Variables

Measuring the Meltzer-Richard mechanism directly:

$$\text{meanMedianRatio}_i = \frac{\text{Mean rental value}}{\text{Median wage}}$$

- Directly captures the relative tax burden of rich vs. poor (the Meltzer-Richard channel)
- Tax was levied on rental values, so this ratio maps precisely to the model
- Measured pre-reform to avoid endogeneity

Other inequality dimensions tested (Table 3):

Channel	Measure	Predicted sign
Meltzer-Richard	Mean/median ratio	$+$
Elite de facto power	Ex officio chairman	$-$
Fairness concerns	Land concentration, large farms	$+$
Economic need	Cereal suitability, % agricultural laborers	$\pm$

Addressing Spatial Correlation

Income inequality is **geographically clustered** in this data – arable agriculture (high inequality) dominates the east, pastoral (low inequality) the west. This is a textbook case for spatial concern.

Chapman's approach, following Conley & Kelly (2025):

1. **Moran's I diagnostic:** Computed for the inequality measure and for changes in relief spending. After controlling for agriculture type and other observables, *little evidence of spatial autocorrelation remains*.
2. **Conley standard errors:** All regressions use SEs that allow for arbitrary spatial correlation within a **100 km bandwidth** (Colella et al., 2019). Results are robust across a range of bandwidths.
3. **Flexible spatial time trends:** Specification (7) in Table 2 includes interactions of quartic time trends with longitude and latitude – equivalent to a spatial basis regression – and results are unchanged.

Main Results

Support for the Meltzer-Richard hypothesis (Table 2):

- $\hat{\beta} \approx 0.11$ (SD units) across all specifications, significant at 1%
- At the 90th percentile of inequality: spending rose by $ $13% relative to 1893 levels
- Effect is concentrated on **outdoor relief** (direct cash/kind transfers), not indoor (workhouse) – consistent with the redistributive mechanism
- Robust to crop controls, demographic controls, financial controls, union-specific time trends, and spatial corrections

Elite de facto power (Table 3):

- Unions with an unelected (ex officio) chairman before reform saw **smaller** spending increases ($\hat{\beta} \approx -0.13$)
- Elites used local institutional control to blunt democratisation's redistributive impact
- Effect mainly on less publicly salient spending items – harder for voters to monitor

Conclusion: the paper provides clean empirical support for Meltzer-Richard in a