Matrix-Free Spatial Autoregressive Modeling Using the EquiPop Neighborhood Structure: A Computational Alternative to Sparse-Matrix Estimation

Extended Abstract

1. Motivation

Spatial autoregressive (SAR) and Spatial Durbin (SDM) models are central tools in spatial econometrics. Their estimation traditionally relies on explicit construction of sparse spatial weight matrices and evaluation of the log-determinant term

$$\log \ |I - \rho W|$$

which is computationally demanding for large datasets.

At the same time, EquiPop — introduced by John Östh (2016) — provides an alternative representation of neighborhood relationships based on population-balanced nearest-neighbor structures. EquiPop produces structured neighbor tables that can be used directly without forming full sparse matrices.

This study asks:

Can we estimate SLX/SAR/SDM models in a fully matrix-free way using the EquiPop structure while preserving the theoretical foundations of spatial econometrics?

---

2. Methodological Framework

2.1 EquiPop Neighborhood Representation

Instead of constructing an explicit $n \times n$spatial weight matrix W, we use:

- Neighbor index lists
- Row-standardized weights
- Precomputed spatial lag variables

Spatial multiplication is performed through iterative neighbor aggregation:

$$\left(Wy\right)_i = \sum_{j\in\mathcal{N}(i)} w_{ij}\, y_j$$

This avoids sparse matrix storage entirely.

2.2 SAR and SDM Specifications

We estimate:

SAR model:

$$y = \rho Wy + X\beta + \varepsilon$$

SDM model:

$$y = \rho Wy + X\beta + WX\theta + \varepsilon$$

Estimation proceeds via grid search over ρ, using:

- Transformed regression:

$$y - \rho Wy = X\beta + \varepsilon$$

- OLS estimation conditional on ρ

2.3 Log-Determinant Approximation (Pace–Barry)

To evaluate the likelihood, we approximate:

$$\log\ \mid I - \rho W \mid$$

using the stochastic trace estimator of Pace and Barry (1999):

$$\log\ \mid I - \rho W \mid = -\sum_{m=1}^{M} \frac{\rho^m}{m} \operatorname{tr}\left(W^m\right)$$

where traces are approximated via random probe vectors.
Importantly, this is done without forming W, using only iterative neighbor multiplication.

2.4 Global Feedback and Neumann Series

The spatial multiplier:

$$(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 + \cdots$$

represents global spillovers.
Higher-order terms capture:

- First-order neighbors
- Neighbors-of-neighbors
- System-wide feedback loops

We investigate how matrix-free approximations handle these global feedback effects.

---

3. Empirical Validation

The framework is validated in two stages:

Stage 1: Anselin's Columbus Dataset

- Benchmark spatial dataset
- Comparison of:
    - Coefficient estimates
    - Spatial autoregressive parameter ρ
    - Log-likelihood values
    - Residual Moran's I

Results show close correspondence between:

- `spatialreg::lagsarlm`
- EquiPop matrix-free implementation

Stage 2: Large-Scale Airbnb Data

We then apply the method to Airbnb listing data, where computational scalability becomes crucial.

Findings:

- Coefficient magnitudes broadly align with traditional models
- Spatial parameter estimates are comparable
- Matrix-free implementation dramatically reduces memory requirements
- Computational time is substantially lower when spatial lags are precomputed

Residual spatial diagnostics reveal that:

- Local neighbor averaging captures substantial dependence
- Proper treatment of global feedback remains critical in SAR settings

---

4. Computational Comparison

Step	Traditional Model	EquiPop Matrix-Free Model
Construct W	Sparse matrix (n×n)	Neighbor lists only
Compute Wy	Sparse matrix multiplication	Iterative neighbor aggregation
Log-determinant	Sparse Cholesky / trace approx	Stochastic trace (Pace–Barry)
Memory	O(n·k) storage	O(n·k) but no matrix object
Scalability	Limited by sparse matrix ops	Fully matrix-free

The EquiPop approach avoids explicit matrix algebra while maintaining theoretical consistency.

---

5. Key Contributions

1. Demonstrates feasibility of matrix-free SAR/SDM estimation.
2. Integrates EquiPop neighborhood structures into spatial econometrics.
3. Implements Pace–Barry trace approximation without constructing W.
4. Shows scalability advantages for large spatial datasets.
5. Highlights importance of correctly approximating global feedback effects.

---

6. Future Research Directions

The next research steps include:

- Developing improved Neumann-series approximations within the EquiPop framework.
- Implementing Bayesian spatial autoregressive estimation.
- Extending the approach to origin–destination (OD) flow models.
- Investigating convergence properties of higher-order spillover approximations.

---

7. Conclusion

This study shows that spatial autoregressive models can be estimated without constructing explicit sparse spatial weight matrices. By combining EquiPop neighborhood structures with matrix-free trace approximation methods, we provide a computationally efficient and scalable alternative to traditional sparse-matrix spatial econometric estimation.

The framework opens the door to large-scale spatial modeling where matrix construction becomes the primary bottleneck.