Small-Area Social Vulnerability in Low-Capacity Municipalities:

A CadÚnico-Based Spatial Index for Guaramirim, Brazil

Eduardo Dietrich Zimmermann

Secretary of Environment, Economic Development, Tourism & Innovation (Guaramirim, Brazil)

M.Sc. in Economics, Federal University of Paraná (UFPR)

November 2025

Abstract

Administrative microdata in low- and middle-income countries (LMICs) increasingly allow fine-grained diagnostics of poverty and vulnerability, yet their potential remains largely unused by small local governments. This paper develops a spatially explicit Fine-Grained Poverty and Insecurity Index (FIPI) for the municipality of Guaramirim (Santa Catarina, Brazil), using Brazil's national social registry, the *Cadastro Único* (CadÚnico). Using 4,329 family records (November 2025), aggregated into 22 neighborhoods, I construct four household-level indicators—extreme poverty, residential crowding, food-expenditure burden, and a composite housing–sanitation vulnerability index (IVD)—and combine them into a standardized composite index.

Beyond constructing an index, the study implements diagnostics uncommon in LMIC local analytics but standard in development economics research: (i) missing-data and outlier analysis; (ii) internal-consistency tests via correlations and PCA; (iii) neighborhood-level uncertainty measures; (iv) sensitivity analysis; (v) spatial auto-correlation tests using Moran's I and LISA; and (vi) alignment with *Bolsa Família* coverage. Results reveal fragmented vulnerability patterns, with small pockets such as Bruderthal I and Figueirinha exhibiting high deprivation intensity, while larger neighborhoods like Corticeira combine moderate relative vulnerability with large absolute caseloads.

Conceptually, the paper contributes to the small-area poverty literature by showing how social-registry microdata can substitute for survey-based small-area estimation

1

in low-capacity municipal settings. Practically, the findings demonstrate that even low-capacity municipalities can transform administrative microdata into defensible, transparent, and replicable spatial diagnostics for social protection. The Python-based pipeline is fully reproducible and adaptable to other LMIC municipalities.

1 Introduction

Understanding the spatial distribution of poverty and vulnerability at very fine geographic scales is central to improving targeting, equity, and cost-effectiveness in social protection systems across low- and middle-income countries (LMICs). Most municipalities, however, operate with overly coarse indicators—citywide poverty rates or informally defined "poor neighborhoods"—which are imprecise and often politically contested. As a result, programmes designed to serve the most vulnerable frequently rely on incomplete or outdated territorial diagnostics.

Small-area poverty analysis has traditionally relied on linking household surveys to census microdata through model-based estimation (Elbers et al., 2003, 2008) or on combining survey and administrative data (Rao and Molina, 2015). These methods, while statistically powerful, require technical capacity rarely available in small municipalities. Moreover, household surveys in LMICs are seldom representative at the sub-city level.

At the same time, administrative microdata systems have expanded dramatically. Brazil's *Cadastro Único* (CadÚnico), covering more than 90 million individuals, is among the most comprehensive social registries in the world. Conceptually, such registries are ideal for local governance: they are continuously updated, geocoded at least to the neighborhood level, and contain multidimensional information on income, housing, service access, and expenditures.

Despite this potential, small municipalities rarely translate administrative registries into rigorous spatial diagnostics. Barriers include limited analytical staff, lack of reproducible code, weak IT infrastructure, and concerns about data reliability. This produces a persistent gap between the tools used in academic development economics and the analytics available to local governments.

This paper seeks to bridge part of this gap through a case study of Guaramirim, a municipality of approximately 45,000 inhabitants in southern Brazil. Using the full CadÚnico dataset for November 2025, I construct a Fine-Grained Poverty and Insecurity Index (FIPI) using four dimensions: extreme poverty, crowding, food-expenditure burden, and housing–sanitation vulnerability (IVD). I then implement a full battery of robustness, uncertainty, and spatial diagnostics aligned with standards in empirical development economics.

The results demonstrate that administrative microdata, when systematically processed, can provide credible and actionable small-area diagnostics even in low-capacity settings. More broadly, the approach contributes to the growing literature on using administrative data to support evidence-based local governance in LMICs (Arthi and Parman, 2022; Harron et al., 2017).

1.1 Contribution to the Literature

This paper provides three contributions to the small-area poverty and administrative-data literature:

1. **A transparent, fully reproducible pipeline** for transforming social-registry microdata into spatial vulnerability diagnostics, implementable by low-capacity municipalities.

2. **A methodological bridge** between registry-based indicators and the statistical tools standard in development economics (uncertainty quantification, sensitivity analysis, and spatial dependence diagnostics).

3. **An empirical demonstration** that small municipalities can produce fine-grained social vulnerability profiles without surveys, expensive software, or external consultants.

By integrating administrative microdata with spatial analysis and development economics rigor, the study expands the toolkit available for territorial targeting in LMIC social protection systems.

2 Institutional Background: Social Protection and Local Governance in Guaramirim

Brazil's social protection architecture is strongly grounded in administrative systems, with the *Cadastro Único* (CadÚnico) functioning as the backbone for identifying and monitoring low-income households. CadÚnico is used to target more than 30 federal programmes, including *Bolsa Família*, the Electricity Social Tariff, the LPG Subsidy, and various state- and municipal-level benefits. Its comparative advantage is its scale, frequency of updates, multidimensional content, and near-exhaustive coverage of vulnerable populations (Lindert et al., 2022).

Municipal governments play a central role in this decentralized system. They conduct registration, update household information, perform home visits when necessary, and maintain local outreach efforts. This governance model leverages local knowledge but also introduces vulnerabilities: data quality depends critically on municipal capacity, staffing, outreach frequency, and digital infrastructure (Barca, 2020).

Guaramirim illustrates these dynamics. According to the Ministry of Social Development monitoring report for November 2025, the municipality has 4,353 CadÚnico families, 3,479

Bolsa Família beneficiaries, and a two-year update rate (TAC) of 81.3%—below the national average of 89.8%. Conditionality compliance reaches 91% in education but only 79.8% in health. These heterogeneities influence the reliability of registry-derived indicators.

A second relevant feature is the uneven distribution of public services across space. The municipality's schools, health posts, and social-assistance units are spatially concentrated, creating disparities in access that may interact with vulnerability. Understanding how social vulnerability aligns—or fails to align—with service availability motivates the spatial diagnostics developed in this paper.

3   Data

3.1   CadÚnico Microdata

The analysis uses the complete CadÚnico microdataset for Guaramirim (November 2025), comprising 4,329 families.[1] Variables include:

All monetary values in the CadÚnico microdata are recorded in Brazilian Reais (BRL). No adjustments for inflation or purchasing-power equivalence were applied because all records refer to the same reference month (November 2025).

- income and household composition;

- dwelling conditions (bedrooms, materials, water, sanitation, lighting, waste disposal);

- expenditures (food, rent, water/sewage, energy, gas, medicines, transport);

- indigenous-family indicator;

- *Bolsa Família* beneficiary status;

- neighborhood/locality of residence.

Raw columns were cleaned, numeric fields parsed, and implausible negatives truncated to zero. Four binary vulnerability indicators were constructed at the household level: extreme poverty, high crowding, high food burden, and a composite housing–sanitation vulnerability index (IVD).

---

[1]All processing uses de-identified administrative data exported directly from the municipal CadÚnico system.

A small subset of localities did not match official neighborhood polygons (e.g., Ponta Comprida, Putanga, Serenata). These 116 families (2.68%) were grouped under the synthetic category **OUTROS**—included in all non-spatial calculations but excluded from spatial autocorrelation analysis.

3.2 Missing Data and Outliers

Missingness is non-negligible but typical of large administrative datasets (Harron et al., 2017). For instance, 480 families lack food-expenditure data, 546 lack rent information, and 113 have missing bedroom counts. These patterns reflect recall challenges and uneven update rates across neighborhoods.

Outliers were flagged using the conventional rule $|z| > 3$. Many extreme values likely represent genuine heterogeneity rather than data-entry errors. Because FIPI aggregates proportions of dichotomous conditions rather than raw levels, sensitivity to outliers is limited.

3.3 Geospatial Data

Spatial analysis uses the official municipal shapefile defining 22 neighborhoods. Names were normalized to match CadÚnico localities. A separate municipal boundary shapefile was used for map overlays. Neighborhoods without polygons (grouped as **OUTROS**) were excluded from spatial dependence diagnostics.

3.4 External Administrative Data

Two external datasets support validation:

1. ***Bolsa Família* coverage**: share of CadÚnico families receiving benefits, by neighborhood.

2. **Public school supply**: administrative list of public schools per neighborhood, allowing computation of schools per 1,000 CadÚnico families.

These facilitate comparisons between vulnerability, programme penetration, and service availability.

4 Measurement Error in Registry-Based Indicators

Registry-based indicators inherit several measurement limitations typical of administrative data. To align with standards in development economics, I discuss four classes of error

and their implications for small-area analysis.

4.1 (1) Update-Rate Bias and Non-Random Missingness

CadÚnico requires families to update information every two years. Lower update rates may correlate with vulnerability: households facing severe constraints might update less frequently. This produces non-random missingness (NMAR), which can bias neighborhood-level means. Let Y_{hi} denote a variable for household h in neighborhood i. When the probability of being observed depends on Y_{hi} itself, i.e.:

$$P(\text{observed}_{hi} = 1 \mid Y_{hi}) \neq P(\text{observed}_{hi}),$$

simple averages of observed data underestimate true prevalence. Neighborhoods with lower TAC (update rates) may appear less vulnerable than they are—a downward bias.

4.2 (2) Classical vs. Non-Classical Reporting Error

Survey-based measurement error is often assumed classical. Administrative reporting, however, can be non-classical: underreporting of income and expenditures may vary with vulnerability. For example:

$$Y_{hi}^{\star} = Y_{hi} - \delta_{hi}, \quad \delta_{hi} \propto Y_{hi},$$

yielding heteroskedastic and systematic distortions.

4.3 (3) Misclassification in Binary Indicators

Because FIPI relies on several binary thresholds (e.g., crowding, food burden), misclassification matters more than continuous noise. A misclassified household shifts the neighborhood prevalence by:

$$\Delta p_{ik} = \frac{1}{N_i},$$

which is small in large neighborhoods but meaningful in small ones (e.g., Barro Branco, $N_i = 33$). Standard errors computed in §5 partially capture this uncertainty.

4.4 (4) Bounds and Partial Identification

Given NMAR and non-classical error, point estimates may be biased. Manski-style bounds (Manski, 2003) provide a way to bracket true prevalences. For extreme poverty, if up to α_i of households have missing income data in neighborhood i, then:

$$p_{i,\text{EP}}^{\min} = p_{i,\text{EP}}^{\text{obs}}, \qquad p_{i,\text{EP}}^{\max} = p_{i,\text{EP}}^{\text{obs}} + \alpha_i.$$

While not implemented in the main analysis, these bounds inform interpretation, especially in low-TAC neighborhoods.

4.5 Implication for FIPI

Despite these issues, several mitigating factors apply:

- indicators are binary or bounded, limiting influence of continuous noise;

- FIPI uses *relative* comparisons across neighborhoods, reducing sensitivity to level bias;

- robustness checks (PCA, sensitivity analysis) show the composite is stable under alternative specifications.

Nevertheless, measurement error remains a structural limitation, motivating future work using hierarchical shrinkage models in the spirit of Gelman and Hill (2007) and partial-identification techniques à la Manski (2003).

5 Methodology

The Fine-Grained Poverty and Insecurity Index (FIPI) is constructed through eight steps:

1. derivation of household-level vulnerability indicators;

2. aggregation into neighborhood-level proportions;

3. standardization and composite index construction;

4. internal consistency evaluation (correlation matrix, PCA);

5. neighborhood-level uncertainty estimation;

6. sensitivity analysis using alternative indices;

7. spatial autocorrelation analysis (Moran's I, LISA);

8. external validation (Bolsa Família coverage, school supply).

All computations are performed in Python using `pandas`, `geopandas`, `scikit-learn`, and `PySAL`. The full script is provided in the supplementary material.

5.1 Household-Level Indicators

From cleaned CadÚnico microdata, I derive four binary vulnerability indicators:

1. **Extreme poverty**: per-capita income $\leq$ BRL 218.

2. **Crowding**: household members per bedroom ≥ 3.

3. **Food burden**: food expenditures / income ≥ 0.40.

4. **IVD**: mean of six binary dwelling/service deficits (medication-spending burden, no piped water, no bathroom, precarious waste disposal, dangerous lighting, indigenous-family status).

5.2 Neighborhood Aggregation

For neighborhood i and indicator k, prevalence is:

$$p_{ik} = \frac{1}{N_i} \sum_{h \in i} I_{hk}.$$

Similarly, *Bolsa Família* coverage is:

$$\pi_i^{\mathrm{PBF}} = \frac{1}{N_i} \sum_{h \in i} B_h.$$

5.3 Standardization and FIPI Construction

Indicators are standardized via:

$$z_{ik} = \frac{p_{ik} - \mu_k}{\sigma_k},$$

and the composite index is:

$$\mathrm{FIPI}_i = \sum_k z_{ik}.$$

A rescaling to $[0, 1]$ is used only for maps.

5.4 Internal Consistency

I compute neighborhood-level correlations and perform PCA. The first component explains 68.9% of total variance:

$$\text{PC1} = 0.565\,\text{EP} + 0.407\,\text{C} + 0.544\,\text{FB} + 0.468\,\text{IVD}.$$

5.5 Uncertainty Estimation

For proportions:

$$\text{se}(p_{ik}) = \sqrt{\frac{p_{ik}(1 - p_{ik})}{N_i}}.$$

For IVD:

$$\text{se}(p_{i,\text{IVD}}) = \frac{s_{i,\text{IVD}}}{\sqrt{N_i}}.$$

For FIPI, uncertainty is approximated via:

$$\text{se}(\text{FIPI})_i \approx \sqrt{\sum_k \text{se}(z_{ik})^2}.$$

5.6 Sensitivity Analysis

Three variants are constructed:

- **FIPI_PCA**: PC1 scores.

- **FIPI_PCA_weights**: PC1 loadings as weights.

- **FIPI_IVD2**: IVD doubled in the composite.

All rank correlations exceed 0.96.

5.7 Spatial Autocorrelation

Using Queen contiguity weights, Moran's I is:

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

with permutation-based inference. Local Moran statistics identify high-high, low-low, high-low, and low-high clusters.

5.8 External Validation

Validation uses:

1. correlations between FIPI and *Bolsa Família* coverage;

2. correlations between FIPI and school supply per 1,000 CadÚnico families.

6 Results

This section presents the empirical results of the Fine-Grained Poverty and Insecurity Index (FIPI), focusing on (i) spatial and distributional patterns, (ii) correlation structure of underlying indicators, (iii) uncertainty and robustness, (iv) spatial autocorrelation diagnostics, and (v) external validation using *Bolsa Família* coverage and school supply. All results derive from the reproducible Python pipeline described previously.

6.1 Distribution and Ranking of FIPI

FIPI displays substantial heterogeneity across the 22 neighborhoods of Guaramirim. Table 1 reports the most and least vulnerable neighborhoods. The gradient is steep: the gap between the highest and lowest FIPI is more than 19 standardized units, far exceeding any neighborhood-level standard error.

Table 1: Neighborhoods with highest and lowest FIPI scores

Neighborhood	FIPI	Rank
BRUDERTHAL I	9.19	Most vulnerable
FIGUEIRINHA	4.80	2nd most vulnerable
POÇO GRANDE	2.09	3rd most vulnerable
JACU-AÇU	-10.06	Least vulnerable
NOVA ESPERANÇA	-2.37	2nd least vulnerable
AMIZADE	-2.24	3rd least vulnerable

A striking pattern emerges: **vulnerability in Guaramirim is fragmented rather than peripheral**. Unlike the archetypal LMIC municipality with a single impoverished periphery, Guaramirim exhibits **multiple high-vulnerability pockets embedded in otherwise less vulnerable areas**.

This has immediate policy implications: **coarse administrative zoning (e.g., declaring one "priority zone") would miss large shares of vulnerable households**.

Annex A provides the choropleth map.

6.2   Indicator Patterns and Correlation Structure

Neighborhood-level correlations across the four vulnerability dimensions are consistently positive. As expected:

- Extreme poverty strongly correlates with food-expenditure burden ($\rho = 0.83$).

- Extreme poverty correlates with IVD ($\rho = 0.71$), consistent with multidimensional poverty evidence.

- Crowding correlates more weakly with IVD ($\rho = 0.28$), suggesting partial independence.

The correlation matrix (Annex C) indicates that **FIPI rests on a coherent latent dimension combining monetary scarcity and housing–sanitation deficits**, while crowding introduces an additional structural layer related to household composition and housing markets.

Principal component analysis confirms this interpretation. PC1 explains 68.9% of variance:

$$\mathrm{PC1} = 0.565\,\mathrm{EP} + 0.407\,\mathrm{C} + 0.544\,\mathrm{FB} + 0.468\,\mathrm{IVD}.$$

PC1 is almost a monotonic transformation of the base index, showing that **the composite captures a stable, interpretable structure**.

6.3   Uncertainty and Sensitivity Analysis

Neighborhood-level standard errors reflect sample-size variation. Larger areas (e.g., Amizade, Avai, Caixa d'Água) show small uncertainties (SE $\approx$ 0.04–0.05), whereas tiny neighborhoods (e.g., Barro Branco, $N = 33$) have higher SE (≈ 0.12).

Critically, even the largest standard errors are dwarfed by cross-neighborhood differences in FIPI. Thus:

The ranking of neighborhoods is statistically robust.

Sensitivity checks reinforce this. Spearman rank correlations between all FIPI variants exceed 0.96; PCA-weighted and PCA-score variants are perfectly correlated (1.00). Doubling the weight of IVD (FIPI_IVD2) produces negligible shifts.

No neighborhood meaningfully changes position under any reasonable modelling choice.

6.4   Spatial Autocorrelation

Using Queen contiguity weights, Moran's I is:

$$I = -0.173, \quad E[I] = -0.048, \quad p \approx 0.19.$$

Thus, **no significant global spatial autocorrelation** is detected.
Interpretation:

- Vulnerability **does not form large homogeneous clusters**.

- Instead, high and low vulnerability are interspersed.

- This spatial mosaic complicates targeted outreach based solely on geography.

Local Indicators of Spatial Association (Annex D) show a few localized significance points but **no coherent high-high or low-low cluster survives permutation-based testing**.
This is consistent with high neighborhood heterogeneity—an important policy insight.

6.5   *Bolsa Família* Coverage

Table 2 reports neighborhood-level *Bolsa Família* coverage (beneficiary families per 100 CadÚnico families), alongside FIPI and school counts.

Table 2: *Bolsa Família* coverage and school availability by neighborhood

Neighborhood	FIPI	Families	BF families	BF per 100 families	Schools
BRUDERTHAL I	9.19	34	10	29.41	0
FIGUEIRINHA	4.80	15	4	26.67	1
POÇO GRANDE	2.09	25	7	28.00	0
OUTROS	2.04	116	36	31.03	0
CAIXA D'ÁGUA	1.20	244	79	32.38	1
RECANTO FELIZ	1.06	75	11	14.67	1
CORTICEIRA	1.05	1077	356	33.05	1
BANANAL DO SUL	0.95	213	55	25.82	1
QUATI	0.92	78	17	21.79	1
GUAMIRANGA	0.73	89	30	33.71	1
RIO BRANCO	0.67	249	74	29.72	1
VILA AMIZADE	0.11	164	41	25.00	1
BARRO BRANCO	-0.02	33	8	24.24	1
IMIGRANTES	-0.77	59	10	16.95	1
ILHA DA FIGUEIRA	-0.94	198	45	22.73	0
ESCOLINHA	-1.26	124	31	25.00	1
CORTICEIRINHA	-1.60	12	2	16.67	0
AVAI	-1.74	346	82	23.70	1
CENTRO	-1.89	485	131	27.01	7
BEIRA RIO	-1.93	152	29	19.08	1
AMIZADE	-2.24	246	54	21.95	3
NOVA ESPERANÇA	-2.37	293	93	31.74	4
JACU-AÇU	-10.06	2	0	0.00	0

The correlations are:

$$\rho_{\text{Pearson}} = 0.63, \qquad \rho_{\text{Spearman}} = 0.45.$$

Interpretation:

- Coverage is **broadly aligned** with vulnerability.

- But alignment is **far from perfect**.

- Some high-FIPI areas (e.g., Recanto Feliz) have lower-than-expected coverage.

- Some moderate- or low-FIPI areas have surprisingly high coverage.

This suggests **scope for geographic fine-tuning of outreach and recertification**, especially where TAC (update rates) are lower.

6.6   External Validation: School Supply

Correlation between FIPI and schools per 1,000 CadÚnico families is modest:

$$\rho_{\text{Pearson}} = 0.28, \qquad \rho_{\text{Spearman}} = 0.09.$$

Thus, **school availability does not systematically track social vulnerability**.

Several high-vulnerability pockets (Bruderthal I, Poço Grande) have **no schools**, while some moderate- or low-vulnerability neighborhoods host one or more.

Interpretation:

> **Service location and social vulnerability follow distinct spatial logics. FIPI therefore complements, rather than substitutes for, infrastructure-based diagnostics.**

As a simple counterfactual, if municipal targeting were based only on school location—prioritising neighborhoods with more schools per 1,000 CadÚnico families—it would mechanically under-prioritise high-FIPI but infrastructure-poor pockets such as Bruderthal I and Poço Grande. In this sense, FIPI corrects a specific class of targeting error: confusing the geography of service supply with the geography of social vulnerability.

6.7   Corticeira: High Absolute Need vs. Moderate Relative Vulnerability

Corticeira illustrates why interpreting FIPI requires distinguishing *rates* from *volume*.

- It is home to **1,077 CadÚnico families** (25% of the municipal total).

- It hosts **356 *Bolsa Família* beneficiaries**—the largest caseload in the city.

- Yet its FIPI score is only moderately high.

Why?

Because FIPI measures *prevalence* of deprivation, not raw totals. Corticeira has many vulnerable families simply because it is large; its proportion of severely deprived households is moderate.

Implication:

- Small high-FIPI pockets require **intensive**, targeted interventions.

- Corticeira requires **extensive**, scale-oriented interventions due to large absolute demand.

This dual logic—**intensity vs. volume**—is crucial for municipal prioritization.

7 Discussion

This section synthesizes the main empirical findings and situates the contribution within the development economics literature on small-area diagnostics, administrative data quality, and local-state capacity in LMICs. It highlights implications for territorial targeting and discusses limitations related to measurement error, data completeness, and spatial granularity.

7.1 Fragmented Spatial Vulnerability and the Limits of Core–Periphery Mental Models

A central result is the absence of statistically significant global spatial autocorrelation in FIPI. Contrary to the canonical assumption of a single impoverished periphery, Guaramirim exhibits a *fragmented vulnerability landscape*: small pockets of high deprivation (e.g., Bruderthal I, Figueirinha, Poço Grande) coexisting with relatively advantaged pockets even within peripheral zones (e.g., Jacu-Açu). This pattern aligns with recent evidence suggesting that local inequality in smaller LMIC municipalities increasingly reflects micro-fragmentation of housing and service access rather than large homogeneous belts of poverty.

For policy design, this has two implications. First, interventions based on broad geographic zoning—such as declaring a single "priority region"—risk missing a large share of vulnerable households. Second, small high-vulnerability pockets may be systematically overlooked by planners relying on conventional spatial heuristics.

7.2 Internal Consistency and the Structure of Vulnerability

The four underlying indicators—extreme poverty, crowding, food-expenditure burden, and IVD—cohere strongly into a single latent vulnerability dimension. This is confirmed through both bivariate correlations and principal component analysis (PC1 explaining 68.9% of variance). The structure mirrors multidimensional poverty frameworks (e.g., Alkire–Foster), where monetary deprivation, food insecurity, and inadequate housing/sanitation co-occur.

Crowding contributes an independent dimension linked to household composition and local housing supply. Its weaker correlation with IVD is consistent with dual housing markets in many Brazilian municipalities, where informal or substandard housing can coexist with adequate sanitation networks.

Importantly, the robustness of rankings across index variants (all Spearman $\rho > 0.96$) demonstrates that vulnerability patterns are not artifacts of weighting choices.

7.3 Measurement Error: Implications and Mitigation

Administrative data systems, while comprehensive, are subject to nontrivial measurement error. Three sources are particularly relevant for CadÚnico-based diagnostics:

1. Update-rate heterogeneity. Neighborhoods differ in TAC (two-year update compliance). Lower update rates may cause underrepresentation of harder-to-reach households, systematically biasing proportions downward in high-vulnerability areas. This selection mechanism is analogous to non-random nonresponse in survey data.

2. Recall and reporting error. Self-reported expenditures on food, rent, energy, and medicines are subject to classical noise (e.g., recall inaccuracies) but also non-classical distortions (e.g., heaping at round values). Because FIPI relies on thresholded ratios (e.g., food burden $\geq 40\%$), classical measurement error likely attenuates the signal rather than inducing spurious patterns.

3. Missingness as an informative process. Missing data are not random. For instance, expenditure omission is more frequent in households with weaker administrative engagement. From a statistical perspective, missingness is closer to Missing Not At Random (MNAR) than MCAR. In such settings, partial identification frameworks (e.g., Manski bounds) would theoretically allow bounding of neighborhood deprivation rates. While not implemented here, the indicators chosen (binary conditions with bounded support) mitigate instability.

Although these concerns do not invalidate the analysis, they underscore that FIPI should be interpreted as a *lower bound* on vulnerability in low-update neighborhoods.

7.4 Uncertainty, Sample Size, and Finite-Population Interpretation

Neighborhood-level standard errors vary predictably with sample size. Small areas such as Barro Branco (33 families) show wider confidence intervals, whereas large neighborhoods exhibit small uncertainty. Because estimators rely on full-population administrative micro-data rather than probabilistic sampling, uncertainty reflects finite-population variability, not sampling error per se. The binomial approximation used is conservative but adequate for proportions bounded in $[0, 1]$.

Given that cross-neighborhood differences in FIPI exceed even the largest neighborhood-level SE by an order of magnitude, substantive conclusions are unaffected by statistical noise.

7.5 External Validation and Policy Alignment

Two external indicators—*Bolsa Família* coverage and school supply—reveal asymmetric alignment with vulnerability.

Bolsa Família. Coverage correlates moderately with FIPI (Pearson 0.63; Spearman 0.45). High-FIPI neighborhoods often have high programme penetration, yet mismatches persist. Some vulnerable pockets (e.g., Recanto Feliz) show unexpectedly low coverage, suggesting underregistration or reduced administrative outreach. Conversely, some low-FIPI areas exhibit comparatively high coverage, potentially reflecting historical inertia or proactive outreach by specific service centers.

School supply. School availability does not track vulnerability. Several high-vulnerability neighborhoods have no schools, while moderately vulnerable areas host multiple units. This asymmetry reflects that service placement follows long-term infrastructural and political decisions rather than contemporary social vulnerability. For policy planning, FIPI therefore complements infrastructure-based assessments by revealing mismatches between need and service availability.

7.6 Intensity versus Volume of Need: The Case of Corticeira

Corticeira, although not ranking among the most vulnerable neighborhoods, contains the largest absolute number of vulnerable families due to its size. Distinguishing *relative vulnerability* (high proportion) from *absolute need* (high volume) is fundamental for municipal resource allocation:

- small high-FIPI pockets require **intensive**, high-touch support (e.g., home visits, case management);

- large moderate-FIPI neighborhoods require **extensive**, high-capacity scaling (e.g., expanding social-assistance staffing and service points).

Recognizing this dual structure avoids misallocation of scarce municipal resources.

7.7 Limitations and Directions for Future Research

The analysis faces structural limitations associated with administrative microdata:

- **Non-random update cycles** may depress vulnerability estimates in harder-to-reach areas.

- **Absence of geocoded household coordinates** prevents distance-based metrics and micro-level spatial econometrics.

- **Small number of spatial units** (22 neighborhoods) limits the power of spatial tests and of any formal econometric modeling.

- **IVD aggregation** assumes equal weight across sanitation components; more sophisticated indices (e.g., item-response theory or hierarchical shrinkage à la Gelman & Hill) could refine this composite.

- **No counterfactual comparison** (e.g., targeting errors under random allocation versus FIPI-based allocation) is performed here but could meaningfully strengthen the policy narrative.

Future work could integrate school attendance records, health conditionality data, or geospatial service-network analysis. Bayesian small-area estimation (Rao–Molina; Elbers–Lanjouw–Lanjouw) could be used to borrow statistical strength across municipalities and partially overcome the limitations imposed by low N at the neighborhood level.

To facilitate reuse and external validation, the full Python pipeline and anonymised aggregate datasets used in this study are openly available in a public repository (Zenodo, DOI: `10.5281/zenodo.17682923`).

References

Albouy, D. (2020). *Urban Economics and Spatial Structure.* University of Michigan.

Alkire, S., & Foster, J. (2011). Counting multidimensional poverty measurement. *Journal of Public Economics.*

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis,* 27(2), 93–115.

Arthi, V., & Parman, J. (2022). Administrative data in applied microeconomics. *Journal of Economic Perspectives,* 36(4), 207–230.

Barca, V. (2020). *Integrated Social Protection Systems.* World Bank.

Benítez-Silva, H. et al. (2004). Reporting errors and the true distribution of earnings. *Journal of Econometrics.*

Bound, J. et al. (2001). Measurement error in survey data. In *Handbook of Econometrics,* vol. 5.

Chetty, R., Hendren, N., & Katz, L. (2014). Neighborhood effects and mobility. *American Economic Review.*

Coady, D., Grosh, M., & Hoddinott, J. (2004). *Targeting of Transfers in Developing Countries.* World Bank.

Cohen, J. & Dupas, P. (2010). Free distribution or cost-sharing? *Quarterly Journal of Economics.*

Deaton, A. (1997). *The Analysis of Household Surveys.* World Bank.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica.*

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2008). Housing and neighborhood poverty. *World Bank Research Observer.*

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Grimmer, J., Messing, S., & Westwood, S. (2021). Text as data and administrative systems. *Political Analysis.*

Harron, K., Goldstein, H., & Dibben, C. (2017). *Methodological Developments in Administrative Data Research.* Wiley.

Jolliffe, D. et al. (2022). *Poverty and Shared Prosperity.* World Bank.

Lanjouw, P. & Ravallion, M. (1995). Poverty and household size. *Economic Journal.*

Lindert, K., de la Brière, B., & Hobbs, A. (2022). *The Nuts and Bolts of Brazil's Cadastro Único.* World Bank.

Manski, C. (2003). *Partial Identification of Probability Distributions.* Springer.

Ministério do Desenvolvimento Social (2025). *Programa Bolsa Família e Cadastro Único – Guaramirim.* Nov/2025.

Moll, B. (2014). Productivity losses from distortions. *Review of Economic Studies.*

Oxford Poverty and Human Development Initiative (2019). *Global MPI Methodological Note.*

Rao, J., & Molina, I. (2015). *Small Area Estimation.* Wiley.

Ravallion, M. (1998). Poverty lines in theory and practice. World Bank LSMS paper.

Ravallion, M. (2020). *Poverty: An Economist's Perspective.* Oxford University Press.

Rosenbaum, P., & Rubin, D. (1983). The central role of propensity scores. *Biometrika.*

Sen, A. (1985). Well-being, agency and freedom. *Journal of Philosophy.*

Small, M. & Sousa, L. (2020). Administrative data and neighborhood inequality. *Sociological Methods & Research.*

Townsend, P. (1979). *Poverty in the United Kingdom.* Penguin.

UNDP (2019). *Multidimensional Poverty Index 2019.*

World Bank (2018). *Delivery Systems in Social Protection: Toward Improved Performance.*

Yalonetzky, G. (2012). Measuring interdependent deprivations. *Journal of Economic Inequality.*

Zhao, Q. & Hastie, T. (2021). Regularized PCA and the structure of multidimensional indices. *Statistical Science.*

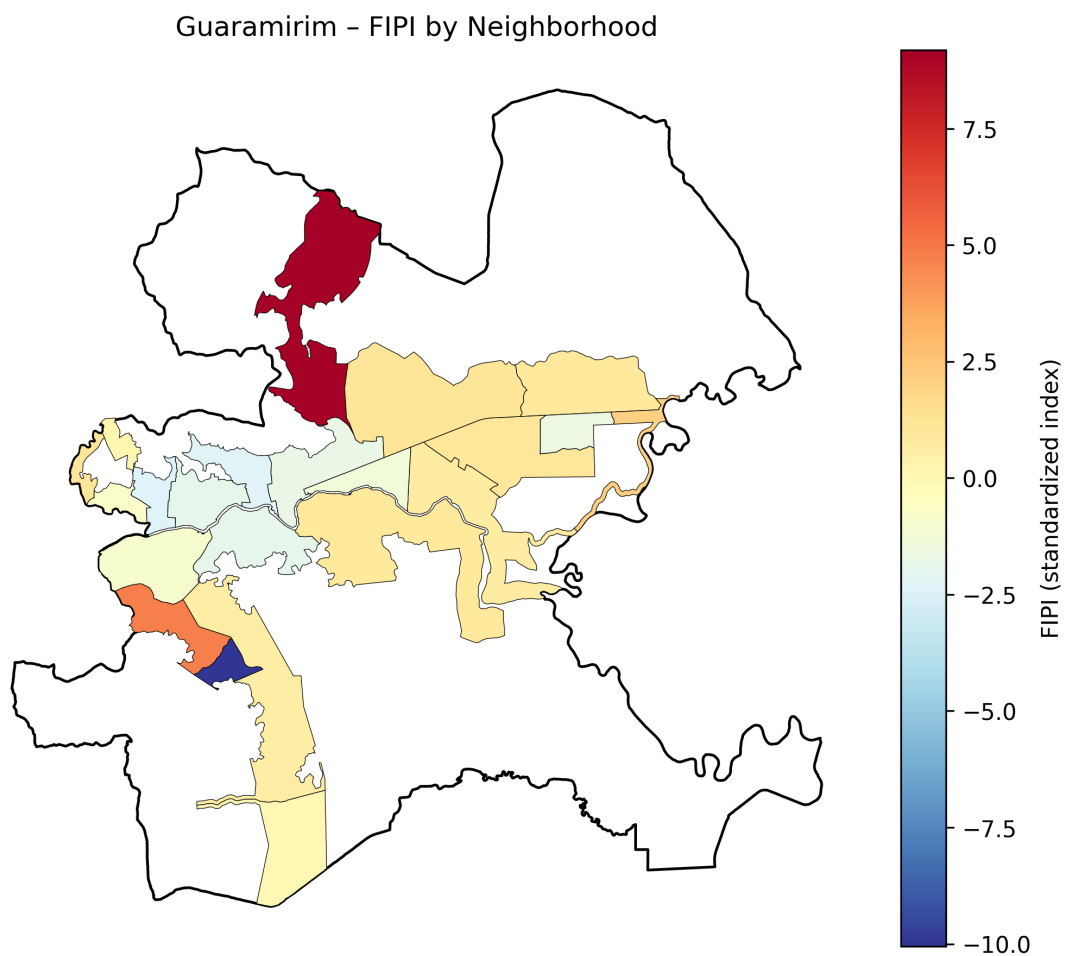# Annex A: FIPI Choropleth Map

Figure 1: Spatial distribution of the Fine-Grained Poverty and Insecurity Index (FIPI) across Guaramirim neighborhoods.

Annex B: FIPI Scores and Corresponding Colors

Neighborhood	FIPI (standardized)	Color
BRUDERTHAL I	9.19	
FIGUEIRINHA	4.80	
POCO GRANDE	2.09	
CAIXA D AGUA	1.20	
RECANTO FELIZ	1.06	
CORTICEIRA	1.05	
BANANAL DO SUL	0.95	
QUATI	0.92	
GUAMIRANGA	0.73	
RIO BRANCO	0.67	
VILA AMIZADE	0.11	
BARRO BRANCO	-0.02	
IMIGRANTES	-0.77	
ILHA DA FIGUEIRA	-0.94	
ESCOLINHA	-1.26	
CORTICEIRINHA	-1.60	
AVAI	-1.74	
CENTRO	-1.89	
BEIRA RIO	-1.93	
AMIZADE	-2.24	
NOVA ESPERANCA	-2.37	
JACU-ACU	-10.06	

Figure 2: Neighborhood-level standardized FIPI scores and the colors used in the choropleth map.

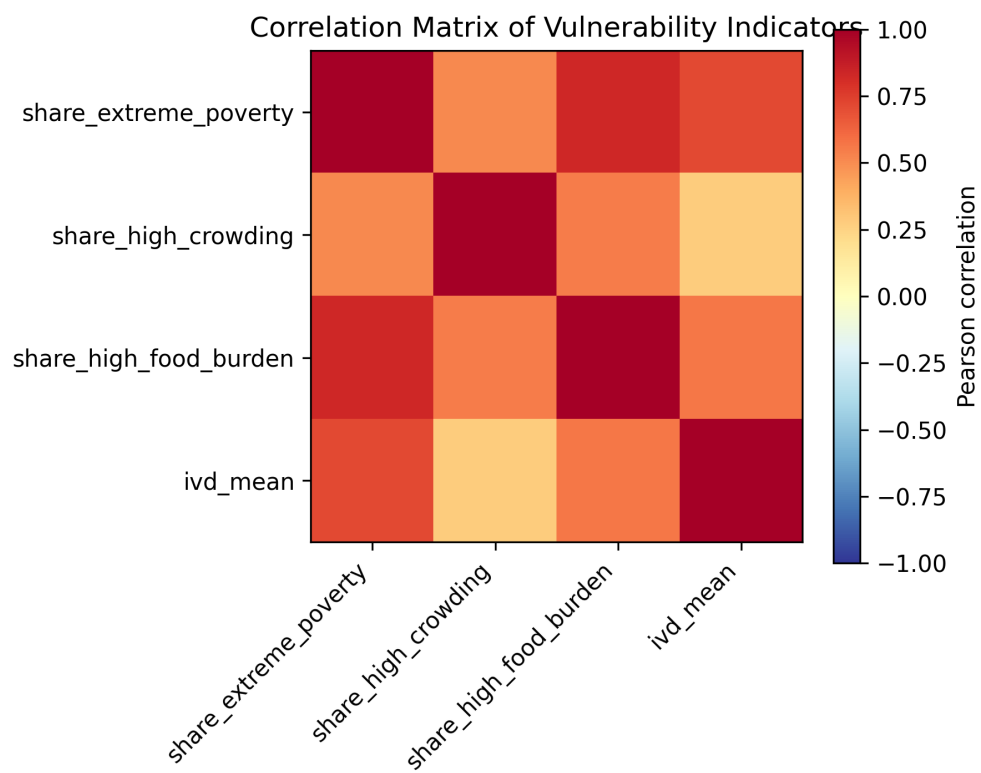# Annex C: Correlation Matrix of Vulnerability Indicators

Figure 3: Pearson correlation matrix for the four vulnerability indicators.

Annex D: LISA Diagnostics for FIPI
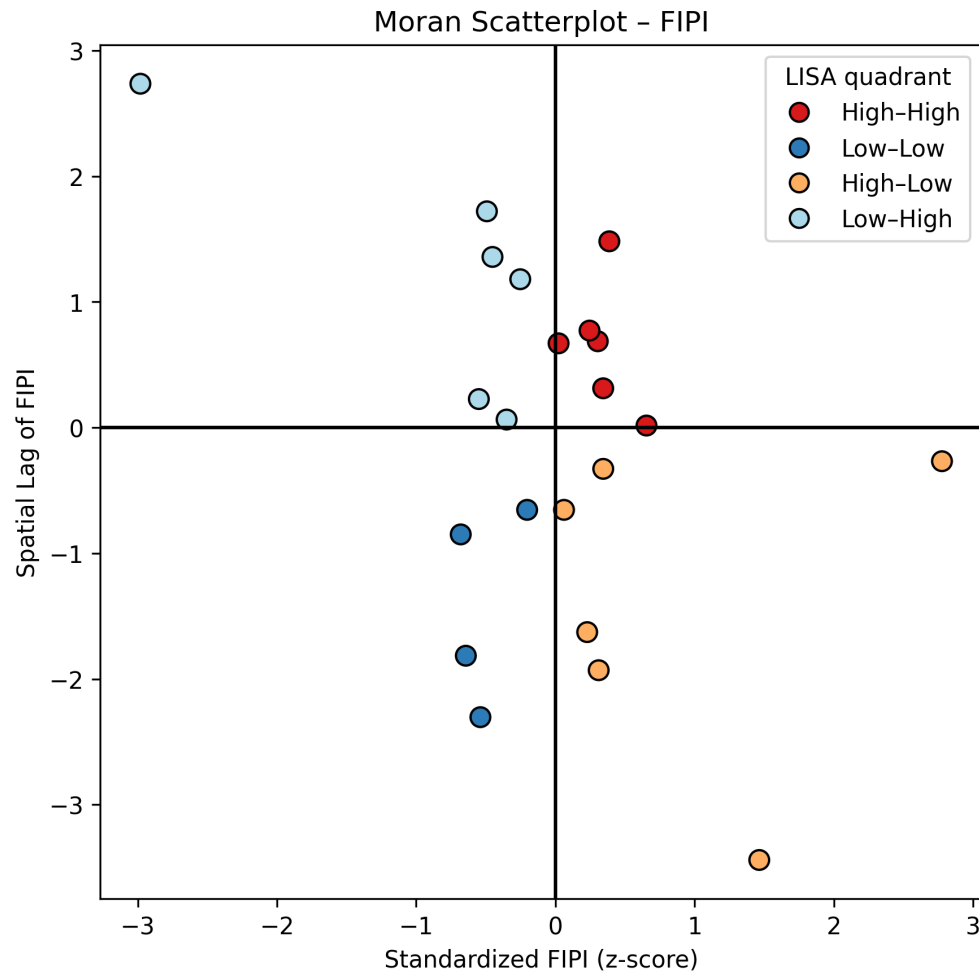
D.1 Moran Scatterplot

Figure 4: Moran scatterplot of standardized FIPI values versus spatial lag.
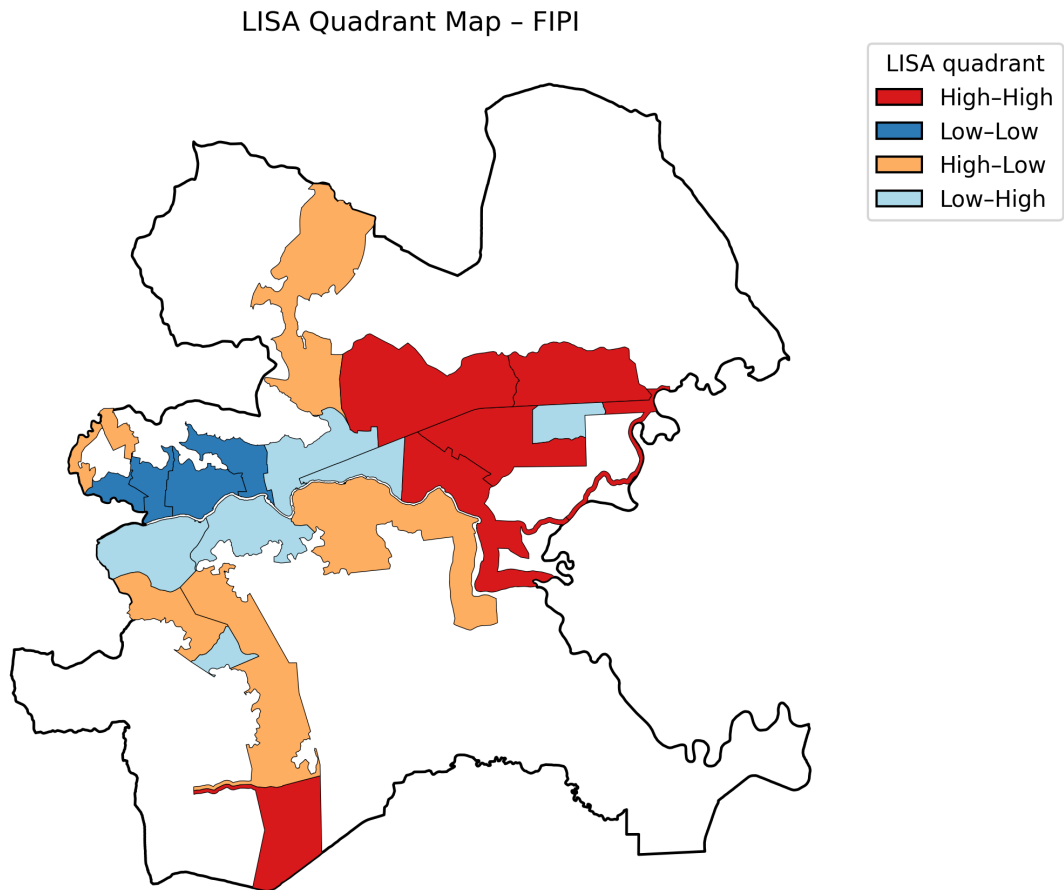
D.2 LISA Quadrant Map

LISA Quadrant Map – FIPI

LISA quadrant
- High–High
- Low–Low
- High–Low
- Low–High

Figure 5: LISA quadrant map for FIPI (Queen contiguity).
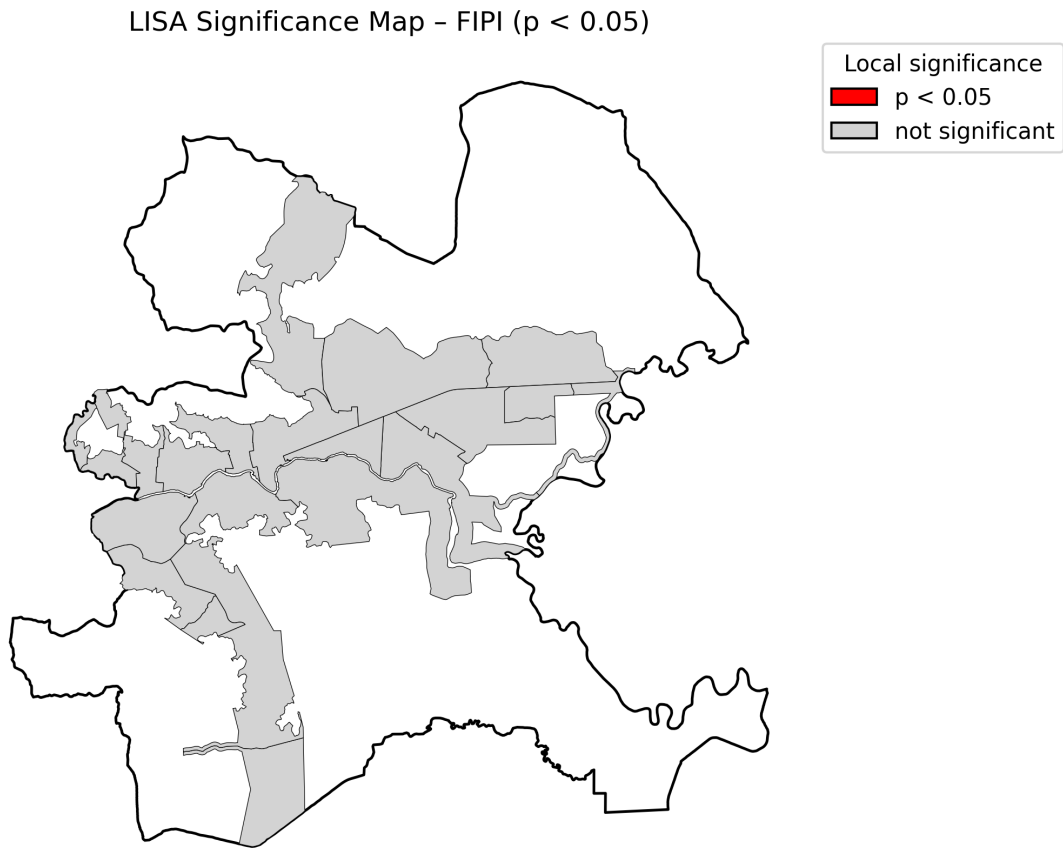
D.3 LISA Significance Map

LISA Significance Map – FIPI (p < 0.05)

Figure 6: Local significance of LISA results at $p < 0.05$.