Sapienza University of Rome, 22 June 2026

Satellite meeting of the Joint Meeting SIS-FENStatS 2026
100 YEARS OF METHODOLOGICAL EVOLUTION: OFFICIAL STATISTICS BETWEEN SCIENTIFIC RIGOR AND DIGITAL INNOVATION

Small Area Estimation and Official Statistics

Maria Rosaria Ferrante
University of Bologna

Relevance of Small Area Estimations and challenges

Demand for granular, small domain statistics: policymakers and society need detailed local data to design and evaluate targeted policies on poverty, inequality, economic development, environment

Traditional design-based estimators: generally unbiased but with high variance in small domains due to small sample sizes

Small Area Estimation (SAE) models by combining survey data with auxiliary data

- can lead to large variance reduction, even if this gain comes at the cost of reliance on the model
- model assumptions can introduce bias

Importance of Model Specification to obtain estimators MSPE reduction

Reducing mismatch between models and socio-economic data reinforces official small area estimates

Our research path: moving beyond Normality

Key starting points

- Many Small Area Estimation models (Fay–Herriot or Battese-Harter-Fuller models) assume Normality
- Unfortunately, economic variables or estimators are often skewed, with heavy tails or with bounded support
- In situations where the normality assumption is violated by real-world data, SAE Normal models may lead to biased or inefficient estimates and inaccurate uncertainty measures
- Over the last two decades, the SAE literature has expanded beyond the classical Gaussian linear mixed model framework by adopting transformations, Generalized Linear Mixed Models for binary or count data, M-quantile approaches, semiparametric and nonparametric models

Extending SAE Framework

Research focuses on flexible models capable of dealing with violations of the normality assumption. Our contribution is not simply to relax normality, but to introduce distributional assumptions tailored to the specific support and shape of the target indicators/variables

Improving Statistical Relevance

Aligning models with real phenomena enhances robustness, interpretability, and policy relevance of official statistics, and then the validity of estimates in the institutional context

Main contributions

Estimated parameter	distributional assumption proposed	survey	Journal + year	Co-authors
Firm averages of value added and labour cost	**Multivariate Skew Normal**	SME	JRSS-A 2017	Pacei
Firm Total Value Added	Log-normal and Variance-gamma	SME	JRSS-C 2018	Fabrizi, Trivisano
relative median poverty gap	Generalized Beta of the second kind (GB2)	EU-SILC	JRSS-A 2020	Fabrizi, Trivisano
Inequality Gini, Relative Theil and Atkinson indexes	**Flexible Beta**	EU-SILC	JRSS-A 2024	De Nicolò, Pacei
Per-capita consumption	Generalized Beta of the second kind (GB2) within GAMLSS	HBS	JSSAM 2025	Mori
Per capita Carbon Foot Print	**Generalized Beta of the second kind (GB2) within GAMLSS**	HBS	JRSS-C 2025	Mori
Inequality Gini, Relative Theil and Atkinson indexes	Generalized Beta of the second kind (GB2) and LogNormal within GAMLSS	HBS	JRSS-A 2025	Mori
Gini index on gross and disposable income	Beta	EU-SILC	RIW 2025	De Nicolò

1926 | 2026 Istat

Phase 1: Small Area Estimation with skewed distributions (1)

M. R. Ferrante, S. Pacei, (2017) Small Domain Estimation of Business Statistics by Using Multivariate Skew Normal Models, JOURNAL OF THE ROYAL STATISTICAL SOCIETY. SERIES A. STATISTICS IN SOCIETY, 180, 4

Goal: to obtain reliable estimates of per-capita Value Added (VA) and Labour Cost (LC)

Domains: macro-regions x economic sectors (NACE) x firm size classes, about 400 small domains where direct estimator is unstable

Data: ISTAT SME survey

Key features:

- explicitly account for the **positive skewness** typical of business data
- exploit the strong **correlation** between VA and LC estimators
- paper carried out in the BLUE-ETS EU project, in collaboration with ISTAT

Phase 1: Small Area Estimation with skewed distributions (2)

Sampling Model $\widehat{\boldsymbol{\theta}}_i \sim SN_K(\boldsymbol{\theta}_i^*, \Omega_i, \boldsymbol{\lambda}_i)$ *K=VA,LC* *i: small domain*

Linking Model $\boldsymbol{\theta}_i^* \sim SN_K(\boldsymbol{\mu}_i, \Omega_\nu, \boldsymbol{\lambda}_\nu)$ $\mu_{ik} = \mathbf{x}_i^T \boldsymbol{\beta}_k$

Inference is conducted in a **Hierarchical Bayesian framework**

Main Empirical Results

- SAE-SkewNormal model outperforms SAE-Normal model in terms of better fit (bias reduction and strong efficiency gains)
- Multivariate model further improves performance

Phase 2: Beyond averages - distributional modelling of inequality (1)

De Nicolò, S., Ferrante, M.R., Pacei, S. (2024). Small area estimation of inequality measures using mixtures of Beta. JOURNAL OF THE ROYAL STATISTICAL SOCIETY. SERIES A: STATISTICS IN SOCIETY, 187 (1)

Goal: to obtain reliable estimates of inequality measures (Gini, Relative Theil, Atkinson indices)

Domains: NUTS-3 regions

Data: EU-SILC survey

Key features:

- Outcomes defined in (0,1) interval
- Skewed and heavy-tailed distributions of estimators
- The Gini index design-based estimator may suffer from underestimation bias in small samples

Research conducted in “GRINS -Growing Resilient, INclusive and Sustainable project” (PNRR), with the participation of ISTAT

Phase 2: Beyond averages – distributional modelling of inequality (2)

Sampling model $y_d \mid \lambda_{1d}, \lambda_{2d}, \phi_d, p \sim FB(\lambda_{1d}, \lambda_{2d}, \phi_d, p)$ $\forall d$ *d: area*

$$f_{FB}\,(\lambda_{1d}, \lambda_{2d}, \phi_d, p) = p\, f_B(y_d; \lambda_{1d}, \phi_d) + (1-p)\, f_B(y_d; \lambda_{2d}, \phi_d)$$

Area-specific mean inequality $\theta_d = \mathbb{E}[\, y_d \mid \lambda_{1d}, \lambda_{2d}, \phi_d, p\,] = \lambda_{2d} + p\,(\lambda_{1d} - \lambda_{2d})$

λ_{2d} location parameter $p(\lambda_{1d} - \lambda_{2d})$ skewness adjustment

Linking model $\text{logit}(\lambda_{2d}) = x_d^{\top}\beta + v_d$ $v_d \sim \mathcal{N}(0,\ \sigma_v^2)$

Main results

- The SAE-Flexible Beta model improves fit, given its four parameters structure, increases reliability and reduces bias compared to standard SAE-Beta model
- Significant improvements are for Relative Theil and Atkinson indices estimators due to their skewed and heavy-tailed distributio

Phase 3: GAMLSS as a flexible SAE framework (1)

Mori L., Ferrante M.R. (2026), Estimating the consumption-based carbon footprint: a small area model as a tool for place-based policies, *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C: APPLIED STATISTICS*, 75,1.

Mori L., Ferrante M.R. (2025), Small area estimation of economic inequality indices using GAMLSS in the absence of covariates, JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A: STATISTICS IN SOCIETY,

Goals:
1) estimation of the per-capita Carbon Footprint (CFP) - amount of carbon dioxide (CO_2) emissions produced directly and indirectly by human activities 2) inequality indices for foreigners living in Italy

Domains: 1) Provinces (NUTS3) 2) Regions (NUTS2) x urban, peri-urban, and rural areas

Data: Household Budget Survey (HBS)

Key Features
- Selection of best-fitting distribution for CFP and foreigners consumption, both skewed and with heavy tails
- SAE Model specification in absence of covariates

Research conducted in “GRINS -Growing Resilient, INclusive and Sustainable project” (PNRR), with the participation of ISTAT

Phase 3: GAMLSS as a flexible SAE framework (2)

Unit level model: SAE-GAMLSS specification

- Each distribution parameter (location, dispersion, scale, and shape) is modelled separately
- GAMLSS admit more than 100 different distributions

$$g_\mu(\mu_{ij}) = x_{ij}^{\mu}\beta_\mu + \gamma_j^{\mu} \qquad \gamma_{\mu j} \sim N(0,\sigma^2_\mu)$$

$$g_\sigma(\sigma_{ij}) = x_{ij}^{\sigma}\beta_\sigma + \gamma_j^{\sigma} \qquad \gamma_{\sigma j} \sim N(0,\sigma^2_\sigma)$$

i: unit, j: area

$$g_\nu(\nu_{ij}) = x_{ij}^{\nu}\beta_\nu + \gamma_j^{\nu} \qquad \gamma_{\nu j} \sim N(0,\sigma^2_\nu)$$

$$g_\tau(\tau_{ij}) = x_{ij}^{\tau}\beta_\tau + \gamma_j^{\tau} \qquad \gamma_{\tau j} \sim N(0,\sigma^2_\tau).$$

- GB2 (four parameter distributions) is the best-fitting distribution for CFP
- LogNormal is the best-fitting distribution for foreigners' consumption

$CFP_{ij} \sim GB2(\mu_{ij}, \sigma_{ij}, \nu_{ij}, \tau_{ij})$ $\quad$ $consumption_{ij} \sim LogN(\mu_{ij}, \sigma_{ij})$ without covariates

Phase 3: GAMLSS as a flexible SAE Framework (3)

Main Results

- SAE-GAMLSS leads to large reduction in MSPE and better fit with respect to SAE-Normal models, even in absence of covariates
- Strong heterogeneity between NUTS3 regions in CFP, overcoming the north-south divide
- Large differences, in the territorial heterogeneity of Gini index, among foreigners and natives

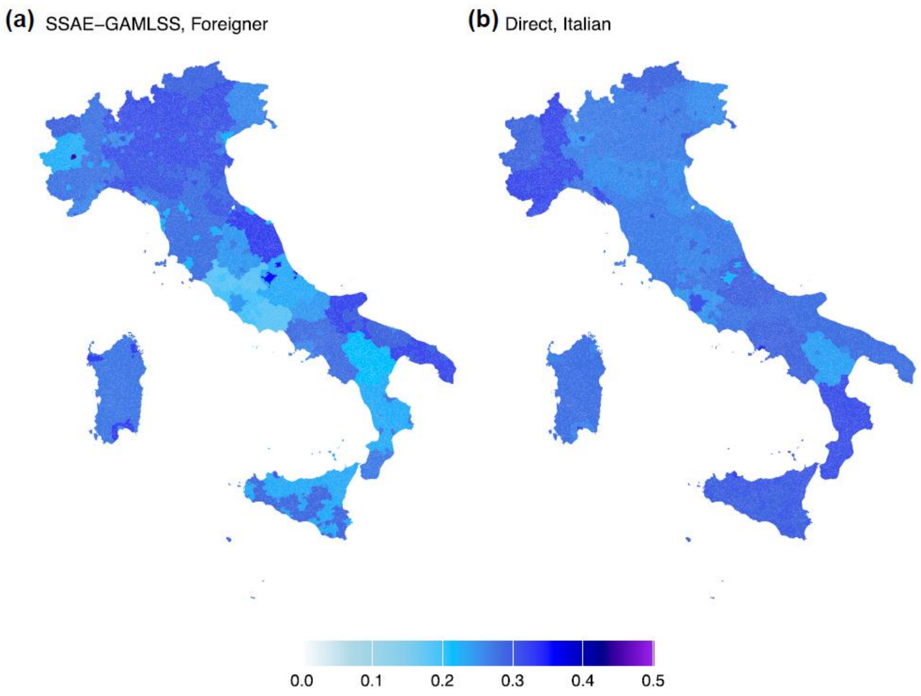

Figure 7. Estimates of the Gini index for Foreigners and Italians.

Conclusions and future directions in SAE for academia and Official Statistics

Distributional Small Area Estimation is a strategic tool in official statistics
- to obtain more accurate and policy-relevant data
- to detect disparities that remain hidden with Normal models

Future research objectives involve ISTAT and some Universities that host SAE researchers (UNIPI, UCSC Piacenza, UNIPG, UNIBO) in a joint research project:

- constructing a multidimensional framework that integrates traditional and emerging deprivation dimensions (a.e., health poverty, energy poverty, resilience)
- building a Territorial Multidimensional Deprivation Index (TMDI)
- integrating SAE and causal inference method, for both ex-post policy evaluation and ex-ante scenario simulation

Synergies between academia and ISTAT researchers can contribute to the mission of official statistics in making available
timely, credible, informative, policy-relevant estimates

grazie

MARIA ROSARIA FERRANTE| maria.ferrante@unibo.it

1926 | 2026

Istat | Istituto Nazionale di Statistica