

# Municipalities' size estimation correcting administrative data for coverage errors and misclassification: A Bayesian approach

V. Ballerini<sup>1</sup>, M. Di Zio<sup>2</sup>, B. Liseo<sup>3</sup> and S. Toti<sup>2</sup>

<sup>1</sup> University of Florence, Italy  
e-mail: [veronica.ballerini@unifi.it](mailto:veronica.ballerini@unifi.it)

<sup>2</sup> Istat, Italy  
e-mail: [dizio@istat.it](mailto:dizio@istat.it), [toti@istat.it](mailto:toti@istat.it)

<sup>3</sup> Sapienza University of Rome, Italy  
e-mail: [brunero.liseo@uniroma1.it](mailto:brunero.liseo@uniroma1.it)

## Abstract

In the process of maintaining the Italian permanent Population Census, the Italian National Statistics Institute (Istat) relies on the *Base Register of Individuals* (BRI, hereafter) to compute population sizes at different levels of aggregation, correcting the administrative data for under- and/or over-coverage. To evaluate the probabilities of under- and over-coverage, Istat conducts two surveys and the *adjusted* population counts estimates are currently obtained by weighting the BRI counts with the ratio of such probabilities. However, this process only produces point estimates for the population sizes, and to obtain an uncertainty quantification of such estimates, a complex bootstrap procedure must be performed, which must consider the complex sampling planes of the two above-mentioned surveys. To overcome such complexities, we approach the problem in a fully Bayesian way by treating the observed BRI counts and the number of under-covered individuals as realizations of random quantities, whose distributions' parameters are functions of the probabilities of over- and under-coverage. The proposed approach makes the model more flexible and able to incorporate uncertainty from different sources. Indeed, we also allow for the concrete possibility that BRI units might be misclassified with respect to some of the individual characteristics, such as citizenship.

We produce an estimate of the posterior distribution of the Italian population size at a municipal level using Markov Chain Monte Carlo methods, overcoming difficulties associated with uncertainty quantification. We illustrate the procedure using the subset of municipalities of less than 18,000 residents (Non-Auto Representative municipalities) in 2018. We also analyze the sensitivity of the results to different prior specifications, demonstrating the robustness of our method.

**Keywords:** Bayesian inference, permanent census, population size estimation, coverage error, misclassification error.

## References

Mancini, L. and Toti, S. (2014) Dalla popolazione residente a quella abitualmente dimorante: modelli di previsione a confronto sui dati del censimento 2011. *Technical report, Istat Working Paper*.