Università di Pisa
Sant'Anna Scuola Universitaria Superiore Pisa
Consiglio Nazionale delle Ricerche
Università degli Studi di Cagliari

Book of Short Papers SIS 2021

SIS 2021
PISA, 21-25 GIUGNO 2021

SIS Società Italiana di Statistica

Editors: Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891927361

Contents

Cira Perna, Università degli Studi di Salerno (Chair of the Program Committee)
Nicola Salvati, Università di Pisa (Chair of the Local Organizing Committee)
Francesco Schirripa Spagnolo, Università di Pisa

Program Committee: Cira Perna (Chair), Massimiliano Caporin, Paola Cerchiello, Fabio Divino, Leonardo Egidi, Raffaele Guetto, Francesco Lagona, Silvia Loriga, Francesca Lotti, Mariagiulia Matteucci, Alessio Pollice, Maria Giovanna Ranalli, Nicola Salvati, Luca Secondi, Isabella Sulis, Luca Tardella, Donatella Vicari, Domenico Vistocco

Local Organizing Committee: Nicola Salvati (Chair), Gaia Bertarelli, Bruno Cheli, Paolo Frumento, Fosca Giannotti, Caterina Giusti, Piero Manfredi, Stefano Marchetti, Lucio Masserini, Vincenzo Mauro, Barbara Pacini, Dino Pedreschi, Francesco Schirripa Spagnolo, Chiara Seghieri.

Technical Committee: Luca Frigau (Chair), Silvia Columbu, Giulia Contu, Mara Manca, Marco Ortu, Maurizio Romano, Cristian Usala, Gianpaolo Zammarchi.

Organizers of Specialized and Solicited Sessions: Giada Adelfio, Emanuele Aliverti, Michela Battauz, Marco Bee, Mauro Bernardi, Luisa Bisaglia, Claudio Ceccarelli, Claudio Conversano, Marco Corazza, Giovanni Battista Dagnino, Antonella D'Agostino, Giovanni D'Alessio, Francesco Finazzi, Michela Gnaldi, Luigi Grossi, Roberto Impicciatore, Giovanna Jona Lasinio, Brunero Liseo, Orietta Luzi, Lucio Masserini, Letizia Mencarini, Stefania Mignani, Gianna Monti, Nazareno Panichella, Pierfrancesco Perri, Lea Petrella, Alessandra Petrucci, Paolo Righi, Elena Stanghellini, Venera Tomaselli, Cristina Tortora, Rosanna Verde, Paola Vicard, Maria Prosperina Vitale, Alberto Zazzaro.

Spatial modeling of childcare services in Lombardia

Modellazione spaziale dei servizi per l'infanzia in Lombardia

Emanuele Aliverti, Stefano Campostrini, Federico Caldura and Lucia Zanotto

Abstract We are interested in mapping the level of childcare services in Lombardia. As a first step, we focus on modeling the flows across municipalities, measuring the number of children that moves from a municipality to another one for using childcare services. Then, we model the coverage rate as the ratio between the number of childcare services and the number of children in a given municipality using a zero-inflated spatial Poisson regression model, providing a model-based map of the level of services in the region. Results allow to approximate the level of services in Lombardia, providing preliminary insights on how resources should be addressed to improve such an aspect.
Abstract *In questo articolo ci concentriamo sulla modellazione spaziale del livello di servizi per l'infanzia in Lombardia. In una fase preliminare, vengono modellati i flussi tra i diversi comuni, in modo da misurare il numero di bambini che si spostano da un comune ad un altro per utilizzare i servizi per l'infanzia. Successivamente, il tasso di copertura (inteso come il rapporto tra il numero di posti ed il numero di bambini) viene modellato tramite un modello di regressione spaziale con risposta Poisson sovra-dispersa, per fornire una stima del livello di copertura all'interno della regione. I risultati di questo approccio permettono di fare chiarezza sulla diffusione di questo fenomeno.*

Key words: Bayesian modeling; Childcare services; Spatial model; Poisson regression.

Emanuele Aliverti
Università Ca' Foscari di Venezia, e-mail: emanuele.aliverti@unive.it

Federico Caldura
Università Ca' Foscari di Venezia, e-mail: federico.caldura@unive.it

Stefano Campostrini
Università Ca' Foscari di Venezia, e-mail: stefano.campostrini@unive.it

Lucia Zanotto
Università Ca' Foscari di Venezia, e-mail: lucia.zanotto@unive.it

1 Introduction

Childcare services were firstly introduced in Italy during 1971 by Law 1044/1971 "servizi sociali di interesse pubblico". Their main aim was assisting parents — in particular women – during childcare, in order to facilitate their participation in the labor market and promote gender equality. Over the years, their role in infant education has been highlighted, since they contribute to cognitive, emotional and social development; in addition, childcare services can also reduce socio-economic inequalities, generating equal education opportunities for men and women. Even if their benefits are generally recognized and supported by national and local fundings, there are huge differences among areas: in northern regions, childcare network services are more developed, while southern ones still have some difficulties to implement them.

A useful index that can help to understand this phenomena is the level of coverage, which can be obtained as the ratio between the number of available childcare services and the number of children between 0-3 years old. In 2010, the European Union in the Barcelona European Council has fixed this parameter at 33% for all the European countries. In Italy, this goal in some municipalities is far exceeded, while in many others it is much lower and the differences are relevant also within regions [3]. One aspect regarding the estimation of the amount of coverage at a local level is that the raw division between available places and children in a single municipality generally underestimates the quantity of interest; for examples, in areas without childcare services, parents can decide to move to close kindergartens and municipalities (particularly those small in population) could decide to support the services of nearby areas, instead of opening a new one. In this work, we try to offer a better estimate of this quantity focusing on Lombardy region in 2018 and taking into account the possible flows between municipalities, relying on the survey "Asili nido e servizi integrativi per la prima infanzia" carried out by Istat. The dataset collects information about all kindergartens (public and private), including the spending of each municipality for childcare services.

2 Data pre-processing

We focus on data from 2018, measured at the municipal level and covering information on the number of children, the number of active childcare services and the overall municipal expenses for childcare services. Potentially, these data allow to measure the coverage rate by taking the ratio between the number of childcare places and the number of children in a given municipality. However, it is well known that several families bring their children to a different municipality, due to the lack of available places in the municipality of residence. We illustrate a simple procedure to estimate, at least partially, this phenomena. We define a municipality as an *out-taker* if it satisfies the following conditions:

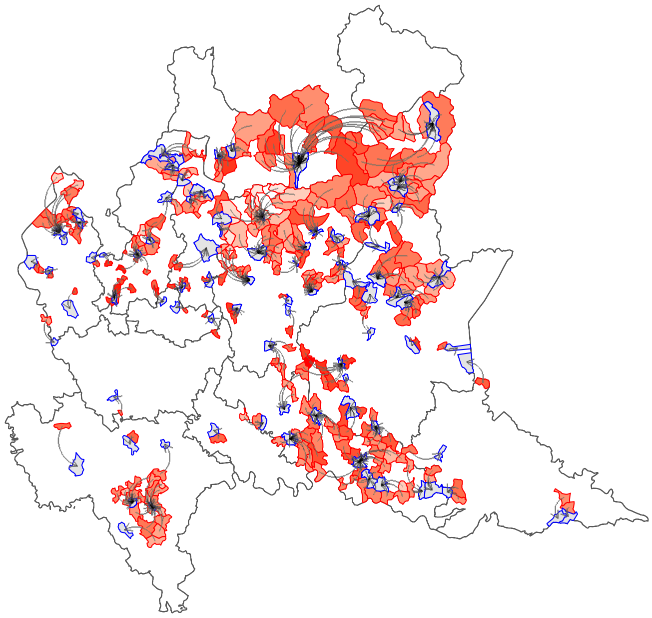

Fig. 1 Estimated childcare services flows.

- absence of childcare services within its boundary;
- it reports a positive number of children, and
- it spends more than 50 euro for each child.

These criteria provide a preliminary rule of thumb for detecting municipality from where there might be outgoing flows. We are aware that we are underestimating an important part of the phenomena, since there might be a flow without the financial contribution from the municipality. However, we found that these criteria are sufficiently simple to exclude a large number of non-interesting municipality, and eventually allow to introduce more refined criteria within a second step, described below.

As a further step, we need to determine where these flows might be directed to. We define as a potential "in-taker" a municipality that registers a ratio between the number of childcare places and the number of children above the 0.8 provincial quantile. Note that we define these municipalities as "potential" since, in order to be included in the analysis, they need to be matched with at least one *out-taker*. Therefore, it is reasonable to take a fairly conservative proportion of municipalities at this step, since many potential are likely to be discarded.

As a final step of this process, we match each *out-taker* with the closest potential *in-taker* in terms of temporal distance. These quantity has been measured in minutes from reaching the two municipalities, using the distance matrices provided by ISTAT.

Figure 1 shows the estimated movements for using childcare services. These are particularly evident in the north-east of the region, where the are some poles which cover the demand of neighboring territories. Also in the south-est of the map the movements are quite significant even if of minor intensity.

As a result, we obtain for each municipality $i = 1,\dots,n$ the number of active childcare places y_i and the number e_i of children referring to such municipality, adjusted through the procedure just outlined. This approach has some drawback, since it ignores the fact that often parents bring children to childcare while commuting to work, non necessary to the closest municipality. This issue could be mitigated including information on commuting into the analysis. Unfortunately, most recent data on this aspect refer to 2011, and in the last 10 years several municipalities have been merged, and the socio-economic landscape has definitely changed. As an alternative, the drawback of our procedure can be restricted considering the heterogeneity of labor market areas. In the next section, we follow this approach and use this information within a spatial model for the level of coverage.

3 Spatial modeling

We model the number of childcare places y_i through a Bayesian Zero-Inflated Poisson (ZIP) spatial model, introducing number of children e_i as an offset and considering the effect of the labor market area of each municipality. Specifically, we let

$$\begin{aligned} y_i &\sim \text{ZIP}(\lambda_i, \pi_0) \\ \log(\lambda_i) &= \alpha + u_i + v_i + x_i^T \beta + \log(e_i), \end{aligned} \tag{1}$$

where λ_i denotes the Poisson mean parameter, α and intercept term and u_i and v_i denote municipality-specific spatial and exchangeable random effects, respectively, while $\log(e_i)$ introduces the number of children as an offset; see [1] for a practical application of this model in epidemiology. In addition, we account for the heterogeneity of labor market areas (SLL) including a set of fixed effects $\beta = (\beta_1, \dots, \beta_p)$, where x_i denotes a p-dimensional indicator vector of the labor marker area for the i-th municipality.

The probability mass function for the response is given by

$$p(y_i \mid \lambda_i, \pi_0) = \pi_0 I(y_i = 0) + (1 - \pi_0) \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i},$$

where the additional parameter π_0 controls the amount of inflation. This specification allows to account for the large number of zero observations in our data, characterizing municipalities without childcare services.

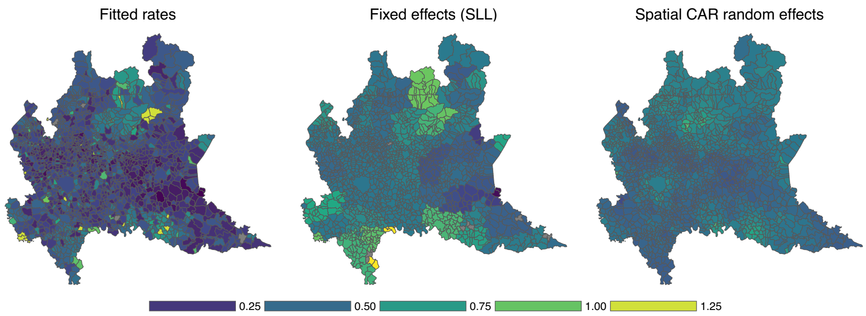

Fig. 2 Left plot: fitted rates of number of places over number of children. Middle and right plot: estimated fixed effects and smooth spatial components, transformed in the rate scales.

In order to account for the spatial dependence across observations, we follow a standard approach in modeling areal data and specify an intrinsic conditional autoregressive structure (iCAR) with precision τ_u on the random effects $\mathbf{u} = (u_i, \ldots, u_n)$; see, for example, [1, sec 6.1] for details. According to such a specification, each location i is modeled as conditionally independent from the others, given its neighbors (corresponding, in our settings, to the municipalities with whom i shares a border); the random effects v_i are instead assumed from a common Gaussian with precision τ_v. This random-effects specification allows to take into account the spatial structure of the data, borrowing information across municipalities. We further specify non informative Gaussian priors on α and non-informative log-Normal distributions on τ_u and τ_v.

We conduct approximate posterior inference trough the `R` package `INLA`, which performs an Integrated Nested Laplace Approximation of the posterior distribution of the model's parameters [2, 4]. We obtain estimates — via posterior mean — equal to $\hat{\pi}_0 = 0.327$, suggesting a modest amount of zero inflation, and for $\hat{\tau}_v = 7.58$ and $\hat{\tau}_u = 2.911$. Estimates for the random and fixed effects are reported in Figure 2, which also illustrates the fitted values for the expected rates $\hat{\lambda}_i/e_i$, as well as the fixed effects $\exp(\beta)$ and the spatial random effects in the rates scale $\exp(\hat{u}_i)$. Results indicate an interestingly heterogeneous coverage in Lombardy: its level is generally greater than 33% for most municipalities, except for the area located in north of Milan, the territories in south-est Mantova and especially around Brescia (20% or less). It seems particularity good in the Sondrio area and between Cremona and Mantova, where the childcare services are shared.

The spatial approach takes into account the proximity, and it has the advantage of providing an easier interpretation of the phenomena, but, of course, the special features of municipalities can be lost, such as the regional excellences and the major shortcomings. The estimation is very satisfactory for close homogeneous area, like Milan and surroundings, but probably less accurate where there are municipalities covering several neighboring territories, such as Sondrio areas, where the coverage levels are probably overestimated.

4 Discussion

In this article, we have focused on modeling the coverage of childcare services in Lombardia, using a simple spatial model. These analysis have some limitations, but, at the light of these first positive results, can be further improved. One important feature of the proposed work is taking into account parents movements for childcare services, which is essential to provide a more realistic view of the actual offer of services in a territory. Moreover, from a methodological perspective, it might be useful to consider a more elaborate spatial specification taking into account isolated peaks of coverage, characterizing hubs municipality. These aspects are currently under investigation.

5 Acknowledgments

This work has been possible and particularly supported by and in agreement between Dipartimento per le Politiche della Famiglia of the Presidenza del Consiglio dei Ministri, Istat and the Department of Economic at the University Ca' Foscari of Venice.

References

1 Marta Blangiardo and Michela Cameletti. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015.
2 Finn Lindgren and Håvard Rue. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25, 2015.
3 Dipartimento per le politiche della famiglia, Istat, Università Ca' Foscari, and MIPA. *Nidi e servizi educativi per la prima infanzia, stao dell'arte, criticità e sviluppi del sistema educativo integrato 0-6*. Presidenza del Consiglio dei Ministri, 2020.
4 Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B*, 71(2):319–392, 2009.