A two-step approach to analyze the inclusivity distribution in childcare services in Italy

Angela Andreella[1], Gaia Bertarelli[1], Federico Caldura[1], and Stefano Campostrini[1]

Ca' Foscari University of Venice, Department of Economics, Italy,
gaia.bertarelli@unive.it

Abstract. This paper studies inclusiveness in Early Childhood Education and Care (ECEC) services in Italy. The Italian legislative decree n. 65/2017 emphasizes the inclusiveness of ECEC services, defined as openness to all children. We propose a two-step model to analyze this concept. In the first step, the inclusiveness is estimated as a latent variable through an item response theory model. Then, the empirical best prediction (EBP) approach provides a reliable aggregate estimate of the phenomenon at the regional (NUTS2) level, which are unplanned domains for the considered data. In the second step, a mixed quantile model analyzes the distribution of this new inclusivity index, showing the disparities between public and private ECEC services and identifying virtuous regional scenarios. The results underline the need to standardize the accessibility features of ECEC services across Italy.

Keywords: Inclusiveness, ECEC, IRT model, mixed quantile regression

1 Introduction

Early Childhood Education and Care (ECEC) services are crucial in cultivating well-rounded economic and social conditions within countries [1], promoting gender employment equality [2], addressing demographic challenges [3] and improving the cognitive development of children [4]. In Italy, these services were introduced in 1971 and expanded in the 1990s. Italy generally still lacks ECEC services, with an overall coverage falling below the European threshold set at 33% and distant from the new target set by EU for the 2023 of 45%. This lack is unequal across Italy, highlighting regional disparities. Addressing these disparities is crucial for improving ECEC service coverage in Italy [5, 6]. Moreover, a pivotal aspect for policymakers is understanding how much of the limited and variable supply is truly accessible to the population. Even with a considerable number of available places, if they are not easily accessible, the overall situation for the country will not improve. The Italian legislative decree n. 65/2017 defines ECEC services as inclusive if open to all children, respecting the individuality, culture, and religion of the child and their family.

The Italian National Statistical Institute (Istat) [7, 8] has analyzed preliminary data on this inclusivity concept, underscoring a notable territorial fragmentation. Public ECEC services are uneven across the country [7], more prevalent in

the north and large cities. This is important since most ECEC services providing financial aid for families facing socio-economic circumstances (e.g., low-income, foreign, children with disabilities) are public [8]. Italy is characterized by various experiences and local choices affecting management models, fee reduction mechanisms, ranking criteria, and accommodation of children with disabilities. The fragmented landscape raises concerns about neglecting inclusiveness as a harmonized and systemic attribute. It is crucial to expand ECEC services, ensuring not only an increase in terms of available places but also in terms of inclusivity and accessibility for families, regardless of their geographical, economic, and social situations. Analyzing the concept of inclusivity requires novel data and proper statistical approaches that can synthesize its multidimensionality and highlight the ECEC services features.

This work proposes a two-step model to define and analyze the concept of inclusivity. Based on the available data, we assume that two dimensions can define an ECEC service as inclusive. The first describes the social aspect, while the second is the economic aspect. Then, in the first step, we estimate an index of inclusivity as a latent trait using an Item Response Theory model (IRT) [9] and return its estimates at the regional level using the empirical best prediction (EBP) approach [10] since regions are unplanned domains of the considered survey. To consider the within-region variability and the bimodal distribution of the inclusivity index, in the second step, we estimate a mixed quantile model [11]. Thanks to this two-step model, we found some virtuous regions where private childcare services are also characterized by a high level of inclusivity.

2 Data

The proposed analysis is based on the 2022 Italian sample survey on ECEC services [8], which recollects information such as the families' demand, accommodation capacity, occupancy rates, and quality of the educational offer for children interviewing ECEC services. This paper focuses on the survey section devoted to measuring the accessibility in terms of inclusiveness of the educational offer. Universally defining this concept may lead to discussions and debates on social, philosophical, and economic ideas, which is beyond this paper's scope. We practically define the concept of inclusiveness by two dimensions: a social and an economic one, starting from the available data and arguing that the selected variables described below can represent only one aspect of the multidimensional concept of inclusiveness.

Table 1 describes the variables used in our analysis. As mentioned earlier, we define inclusiveness by exploring the social (i.e., first 5 variables) and economic (i.e., last 7 variables) aspects of Italian ECEC services. As seen from Table 1, the questionnaire responses are thus simplified as dummy variables, losing some of the information but being able to handle the strong data variability. Moreover, the "foreigners" variable was constructed to consider the actual presence of foreigners in each territory. Let define by c_k the number of foreign children enrolled in the kindergarten k, where $k = 1, \ldots, 1323$ having 1323 childcare services sur-

veyed, and by f_p where $p \in \{1, \ldots, 110\}$ the number of resident foreigners aged 0 to 2 years at the province level (i.e., sub-regional territory). The provincial foreign child enrolled rate is then defined as $F_p = \sum_{k \in \mathcal{P}_p} c_k / f_p$, where $\mathcal{P}_p$ is the set containing the childcare services index in the province p.

The foreign variable at level λ for the kindgarten $k \in \mathcal{P}_p$ is then defined as:

$$\text{Foreign } \lambda = \begin{cases} 1 & \text{if } F_p \geq \mathcal{Q}_F(\lambda) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{Q}_F(\lambda)$ is the quantile of $F = \{f_1, \ldots, f_{110}\}$ at level $\lambda \in \{0.25, 0.5, 0.75\}$. Therefore, the variable Foreign λ is equal for every $k \in \mathcal{P}_p$.

Variable	**Description**
Foreign 0.25	1 if the $F_p \geq \mathcal{Q}_F(0.25)$, 0 otherwise
Foreign 0.5	1 if the $F_p \geq \mathcal{Q}_F(0.5)$, 0 otherwise
Foreign 0.75	1 if the $F_p \geq \mathcal{Q}_F(0.75)$, 0 otherwise
Foreign 1	1 if the $F_p \geq \mathcal{Q}_F(1)$, 0 otherwise
Disability	1 If at least one disabled person registered, 0 otherwise
Meal	1 If meal is included in the fee, 0 otherwise
Fee	1 If no entry fee, 0 otherwise
Disability fee	1 a reduction for child with a disability exists, 0 otherwise
Full Fee	1 If full tuition exemption is granted, 0 otherwise
ISEE	1 If there is a reduction for ISEE, 0 otherwise
Child	1 If there is a reduction for other children (enrolled or not), 0 otherwise
Social Services	1 If there is a reduction for social services indications, 0 otherwise
Family	1 If there is a reduction for another family condition, 0 otherwise

Table 1. Description of the variables analyzed coming from [8].

3 Inclusivity Index

In the first step of our analysis, an IRT model is employed to estimate the concept of inclusivity using the binary variables defined in Table 1.

Consider the manifest binary variable X_i where $i = 1, \ldots, 12$ following a Bernoulli distribution with expected value $\pi_i(\mathbf{z})$. Since this paper is interested in defining an indicator of inclusiveness, we assume that the concept of inclusivity is described by only one latent variable $\mathbf{z}$. The IRT model is defined by:

$$\log(\pi_i(\mathbf{z})/(1 - \pi_i(\mathbf{z}))) = \beta_{0i} + \beta_{1i}\mathbf{z} \tag{1}$$

where β_{0i} can be interpreted as the prevalence of the manifest variable i, and β_{1i} as the effect of the manifest variable i in the inclusivity definition.

The sampling weights at ECEC service level are used inside the likelihood-based estimation process of the parameters β_{0i}, β_{1i} and $\mathbf{z}$ [12].

Figure 1 shows the estimated probabilities $\hat{\pi}(\mathbf{z})$ across different values of the estimated latent variable, i.e., the inclusivity index. The probabilities $\hat{\pi}_i(\mathbf{z})$ are estimated for each variable i, i.e., the binary variables described in Table 1.

The most prevalent inclusive feature is the the meal included in the fee and the less prevalent one is the reduction for disabled children. The most impactful characteristic in defining an ECEC service inclusive is the ISEE reduction, while the least important is the meal's presence within the tuition.

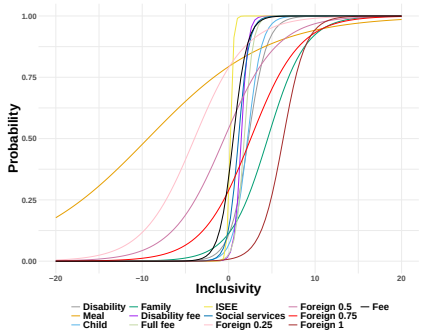

Fig. 1. Estimated probabilities $\hat{\pi}(\mathbf{z})$ across $\mathbf{z}$ for each covariate i defined in Table 1.

We scale $\hat{z}_k$ for each ECEC service k: $\tilde{z}_k = \hat{z}_k - \min_k \hat{z}_k / \max_k \hat{z}_k - \min_k \hat{z}_k$.

Two aspects of $\tilde{\mathbf{z}}$ are pointed out by the two figures below: (i) bimodality, (ii) high within-region variability. Figure 2 shows the frequency distribution of $\tilde{\mathbf{z}}$ by titularity. Figure 3 displays the median and interquartile range (IQR) values of $\tilde{\mathbf{z}}$ at the province level. These aspects lead us to study the features that define the bimodal distribution of the phenomenon with mixed quantile regression. Also, the high variability requires using a small area estimation approach to provide estimates of our inclusiveness index, at least at the regional level.

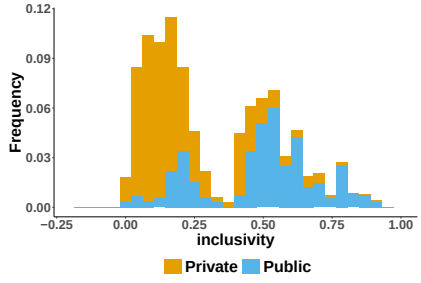

Fig. 2. Relative frequency distribution of the scaled inclusivity index estimated by the IRT model defined in Eq. (1) divided by type of service (i.e., private and public ones).

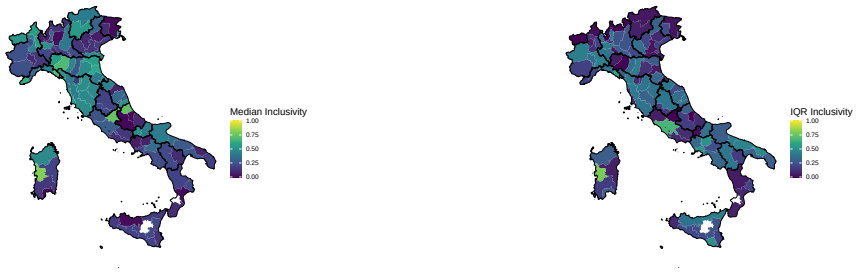

Fig. 3. The median and IQR of $\tilde{\mathbf{z}}$ estimated by the IRT model defined in Eq. (1) for each Italian province where the black lines represent the regional boundaries.

4 Virtuos scenarios

Let $\tilde{\mathbf{z}} = [\tilde{z}_{ij}] \in \mathbb{R}^k$, where $i = 1, \ldots, n_j$ is the single ECEC service, and $j = 1, \ldots, J$ is the region level such that $\sum_{j=1}^{J} n_j = k$. Let $\mathbf{X} \in \mathbb{R}^{k \times P}$ be the design matrix and $\gamma_\tau \in \mathbb{R}^{1 \times P}$ the unknown fixed parameter of interest where $\tau \in [0, 1]$ indicates the quantile level. The mixed quantile model [11] estimates the τ-quantile of the conditional distribution function of $\tilde{\mathbf{z}}$ as

$$Q_{\tilde{\mathbf{z}}|\mathbf{u}_\tau}(\tau \mid \mathbf{X}, \mathbf{u}_\tau) = \mathbf{X}\gamma_\tau^\top + \mathbf{u}_\tau$$

where $\mathbf{u}_\tau \sim \mathcal{N}(0, \mathbf{\Sigma}_\tau)$ denotes the vector of region-specific random effects at τ level quantile. Weights at regional level (i.e., proportions of the Italian ECEC services in the j region) are inserted in the log-likelihood function.

Fixing $\tau = 0.75$, the estimated parameter for privates ECEC equals 0.25 while the public one 0.616 with p-values < 0.0001. After validating the effect of titularity, we now analyze the regional variability. Figure 4 shows the conditional quantile predictions with 0.95 confidence interval. We can note that some virtuous Italian region exist, i.e., Liguria, Trentino Alto Adige, Emilia-Romagna, and Lazio, where also private childcare services are defined as inclusive.

5 Discussion

The most important limitation in our work is that data used in our analysis can only capture partial aspects of inclusiveness, due to the use for which they were originally collected. Despite this, the use of advanced statistical methods can help to capture complex, multidimensional and not directly observable phenomena and to obtain useful information for policy makers at a finer geographical level. For example, private childcare services are not always characterized by less inclusiveness. Although this is generally true, in some regions private services are also virtuous. Moreover, the high heterogeneity of the observed phenomenon within the regions is also evident. This is a sign of the need for more uniform regulations on childcare services not only at a national but also at regional level.

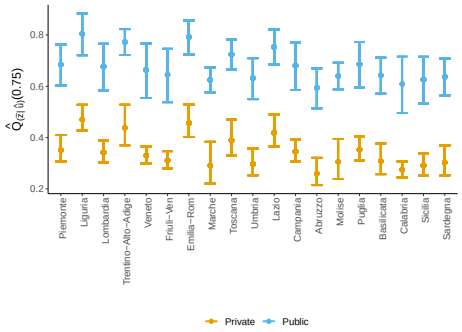

Fig. 4. Conditional quantile predictions at level $\tau = 0.75$ divided by type of ECEC services with 0.95 confidence interval where standard errors are calculated by bootstrap.

References

1. Plantenga, J., Remery, C., & Camilleri-Cassar, F. (2009). The provision of childcare services: a comparative review of 30 European countries. Office for Official Publications of the European Communities, Luxembourg.
2. Landivar, L. C., Ruppanner, L., & Scarborough, W. J. (2021). Are states created equal? Moving to a state with more expensive childcare reduces mothers' odds of employment. Demography, 58(2), 451-470.
3. Luci-Greulich, A. & Thevenon, O. (2013). The impact of family policies on fertility trends in developed countries: L'influence des politiques familiales sur les tendances de la fécondité des pays développés. European Journal of Population/Revue européenne de Démographie, 29:387–416
4. Felfe, C., & Lalive, R. (2018). Does early child care affect children's development?. Journal of Public Economics, 159, 33-53.
5. Andreella, A., Aliverti, E., Caldura, F., & Campostrini, S. (2024). Spatial clusters for demand and supply of childcare services in Italy. arXiv:2401.07600 [stat.AP].
6. Andreella, A. & Campostrini, S. (2023). Suitable statistical approaches for novel policies: spatial clusters of childcare's services in Veneto, Italy. CARMA, 205–212. Editorial Universitat Politecnica de Valencia. 10.4995/CARMA2023.2023.16439
7. Dipartimento per le politiche della famiglia, ISTAT, Università Ca' Foscari, & MIPA (2020). Nidi e servizi educativi per la prima infanzia, stato dell'arte,criticità e sviluppi del sistema educativo integrato 0-6. https://istat.it/it/archivio/244158.
8. Dipartimento per le politiche della famiglia, ISTAT, & Università Ca' Foscari (2022). I servizi educativi per l'infanzia in un'epoca di profondi cambiamenti. https://istat.it/it/archivio/287748.
9. Darrell Bock, R. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. Psychometrika, 35(2), 179-197.
10. Molina, I. & Rao, J. (2010). Small area estimation of poverty indicators. The Canadian Journal of Statistics, 38(3), 369–385.
11. Geraci, M. & Bottai, M. (2014). Linear quantile mixed models. Statistics and computing, 24:461–479.
12. Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural equation modeling, 12(3):411–434.