Indirect Fertility Estimates in Historical Italy: a Bayesian Approach

Michail Raftakis, Riccardo Omenti, and Nicola Barban

Department of Statistical Sciences, University of Bologna

Short Abstract

This paper examines Italy's comparatively late demographic transition, marked by substantial subnational disparities, by developing the first long-run, province-level series of overall fertility that explicitly incorporates male fertility. Using the rich but under-digitized historical censuses of ISTAT, we employ large language models to transcribe and harmonize tabulations and apply a Bayesian hierarchical framework to generate indirect fertility indicators for men and women across 1861–1961. These estimates allow us to chart the evolution and spatial dispersion of fertility trajectories and to quantify systematic sex differences in timing and intensity, yielding a more granular portrait of Italy's transition and a replicable approach to subnational demographic reconstruction. Preliminary results show that subnational fertility trajectories were far from uniform: the more urbanized and industrial North-West entered decline first, while in the broader North and Center the First World War and the 1918–19 influenza pandemic marked an inflection after which fertility either continued its pre-war slide or stabilized at lower levels. By contrast, several Southern provinces had reverted to pre-war fertility by 1931 as wartime disruptions receded and traditional family formation resumed. Over time, these divergent paths widened dispersion and produced a pronounced North–South gradient, yielding an increasingly polarized fertility map. Throughout, male fertility levels exceeded female levels, consistent with older paternal ages, remarriage, and a wider male reproductive span.

Extended Abstract

1 Introduction

Italy's fertility transition unfolded later than in much of Western Europe and with marked spatial heterogeneity (Coale and Watkins, 1986). A key reason for this delayed start was the persistence of high mortality and the country's particular vulnerability to shocks and crises through the 1870s. Under these conditions, fertility limitation, already practiced in small pockets and social groups earlier in the nineteenth century, did not diffuse widely until the century's end. Once underway, however, limitation spread along clear spatial and social gradients: from the Centre–North toward the South, from west to east, and from urban to rural contexts. The consequence was a pronounced peak in variability during the 1930s, when provinces and social categories with very low fertility coexisted alongside areas where natural fertility still prevailed (Breschi et al., 2003, 2014; Livi Bacci, 1977; Livi Bacci and Breschi, 1990; Rettaroli et al., 2017).

These features make Italy a powerful setting for disentangling how local institutions, economic change, and cultural norms shaped reproductive behavior. Prior country-level narratives risk masking this subnational complexity; our study instead emphasizes the geography and timing of fertility change, moving beyond classic summary measures such as Coale's index of overall fertility to develop indicators tailored to provincial dynamics (Coale and Watkins, 1986; Livi Bacci, 1977). Moreover, we place particular emphasis on male fertility, heeding calls to analyze men and women in parallel rather than inferring couple behavior from female series alone (Breschi et al., 2020).

Empirically, we exploit the rich but under-digitized historical censuses by the Italian National Institute of Statistics (ISTAT), using large language models (LLMs) to transcribe and harmonize tabulations and modern statistical methods to confront sparse and noisy signals. Within this broader historical frame, we implement a Bayesian modeling approach to produce provincial-level indirect estimates of overall fertility for both men and women over 1861-1961, enabling us to trace the evolution and dispersion of fertility trajectories across space and to assess systematic sex differences in timing and intensity. Together, these contributions yield a more granular overview of Italy's fertility transition and a replicable framework for subnational demographic reconstruction. Although the results presented in this abstract constrained to 1911-1931, the analysis of this paper will be extended to capture the full trajectory of the Italian fertility transition from 1861 to 1961.

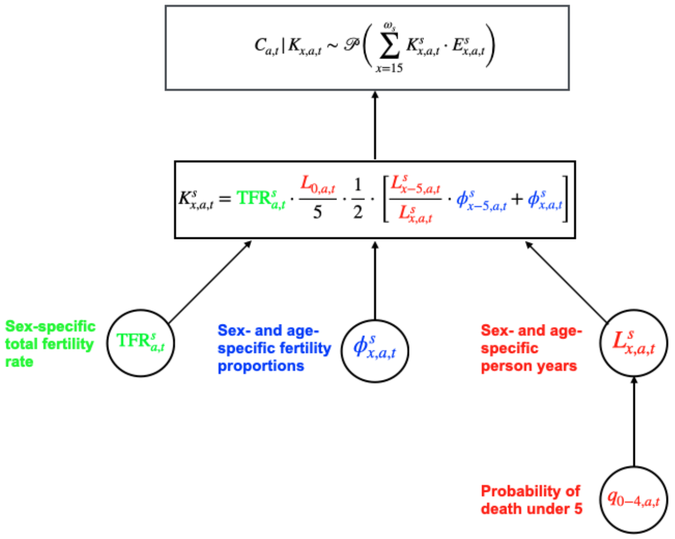
2 Data and Methods

2.1 Method

We employ the Bayesian indirect estimation approach introduced by Schmertmann and Hauer (2019) and further refined by Omenti et al. (2025) to estimate male and female fertility at the subnational level. This method enables the indirect estimation of the total fertility rate (TFR) for both sexes using age- and sex-specific population counts as main data input, without requiring data on births by parental age. The method is particularly suited to reconstructing fertility patterns at the provincial level in Italy during the study period, when detailed vital statistics on live births by parental age at such detailed geographical level were unavailable.

As illustrated in Figure 1, the model assumes that the number of children, $C_{a,t}$, in province a and year t follows a Poisson distribution. The mean of this distribution represents the expected number of children under age five either to woman aged 15-49 or to man aged 15–59. This expected value is expressed as a sum over five-year reproductive age groups, where each term, $K^s_{x,a,t}$, denotes the expected number of children per individual (male or female) in age group x. The number of women and men are present as offset. Fertility is assumed to be zero outside the age ranges $[15, 49]$ for women and $[15, 59]$ for men.

Figure 1: The figure provides an overview of the Bayesian modeling approach by **?**



Note: The figure summarizes the Bayesian model used to obtain indirect total fertility rate estimates. The upper panel represents the main data model, where the number of children is assumed to follow a Poisson distribution. The lower panel shows the expression for the expected number of children ($K_{x,a,t}$), which decomposes fertility into sex-specific total fertility rates (green), sex- and age-specific fertility proportions (blue), and exposure and mortality components (red).

Each $K^s_{x,a,t}$ depends on parameters with clear demographic interpretations. The total fertility rate, $TFR^s_{a,t}$, represents the fertility level of men or women in province a at time t and is modeled as a normal distribution centered on the Italian national TFR. The age- and sex-specific fertility proportions, $\phi_{a,t,s}$, are estimated using a principal component regression model applied to standard fertility curves, following an approach similar to the Lee–Carter model Lee and Carter (1992). When modeling $\phi_{a,t,s}$, temporal and spatial smoothing components are incorporated to allow fertility patterns to evolve smoothly over time and across provinces. Finally, the person-year parameters, $L^s_{x,a,t}$, are derived using the log-quadratic mortality model of Wilmoth et al. (2012), applied to regional estimates of under-five mortality, $q_{0-4,a,t}$.

2.2 Data

The analysis draws on provincial-level population counts disaggregated by age and sex from the decennial Italian censuses of 1911, 1921, and 1931. These data are obtained by ISTAT and provide complete enumerations of the resident population at ten-year intervals. Scanned images of printed census volumes were transcribed using LLMs. The LLM-assisted transcription process was followed by extensive manual verification to ensure consistency with the original tabulations.

To generate annual population estimates by age, sex, and province, we interpolated between census years. Specifically, using national population counts by year, age, and sex from the Human Mortality Database, we assumed that provincial population growth rates followed the corresponding national trends. This assumption allows us to account for historical shocks, such as the demographic effects of the First World War and the 1918-19 Influenza Pandemic, that linear interpolation alone would fail to capture.

To calibrate the child mortality component of the model, we calculated regional estimates of the probability of dying before age five ($q_{0-4,a,t}$) using death counts from annual aggregate statistics of the movement of the population obtained by ISTAT. The parameter representing total fertility ($TFR_{a,t}$) is similarly calibrated using national-level female total fertility rate (TFR) values reported by ISTAT for the corresponding years.

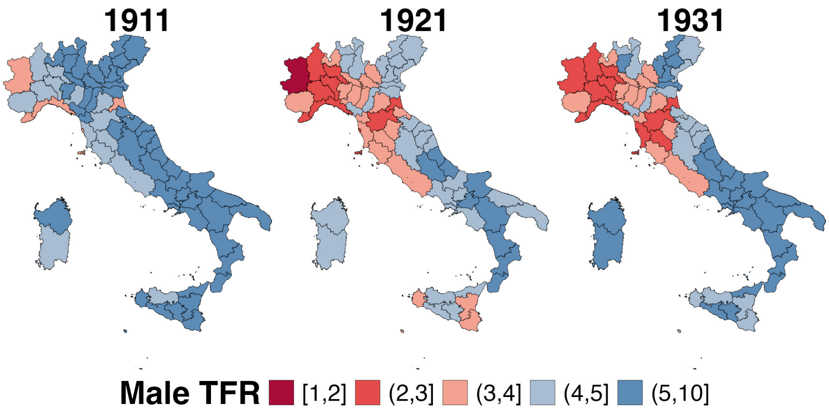
3 Preliminary Findings

Figure 2 displays the geographical distribution of the male and female TFR estimates in Italy in three subsequent census years. Across the period, subnational trajectories were far from uniform. The North-Western provinces, more urbanized and industrial, entered the fertility downturn first. In the broader North and Centre, the First World War and the 1918-19 influenza pandemic seem to may have marked a clear inflection: for many provinces overall fertility either continued its pre-war slide or flattened at lower levels, consistent with the diffusion of parity control and delayed marriage as described by Livi Bacci (1977). By contrast, several Southern provinces reverted to pre-war fertility by 1931, as temporary wartime shocks, mobilization, nuptiality disruption, and influenza, receded and traditional family formation resumed. Over time these divergent paths widened the dispersion of provincial fertility levels: early-declining Northern areas pulled further below the national mean while the South lagged, producing a pronounced North–South gradient by the interwar years. The result is an increasingly polarized fertility map, with timing and tempo differences, rather than a uniform national decline,driving the rise in subnational inequality. Throughout, male fertility indicators remained above female ones, reflecting older male ages at reproduction, remarriage patterns, and the wider age span over which men father children (Figure 3).

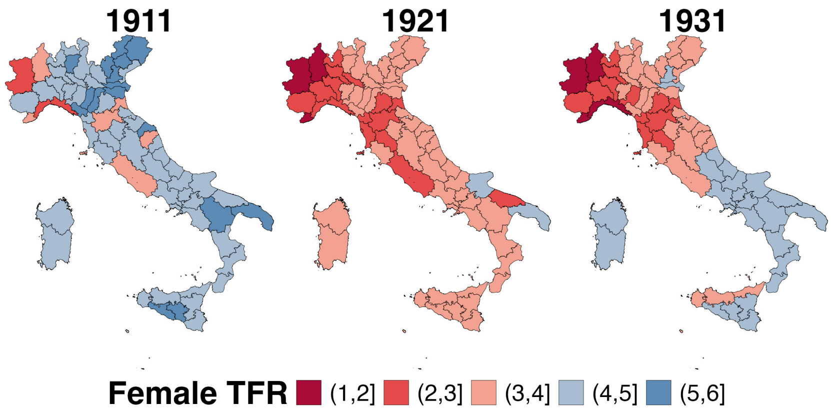
4 Next Steps

In the coming months, we will extend the time horizon to 1861–1961 by systematically and automatically transcribing additional census tables, a step that is essential for capturing the full trajectory of Italy's fertility transition. This expanded window will allow us to pinpoint the timing of fertility declines at a finer geographical scale and to compare North–South patterns more precisely. With harmonized series for both men and women, we will examine how fertility trajectories evolved between provinces, assessing whether and when convergence or divergence predominated.

Figure 2: Male and female TFR estimates in three selected years (1911, 1921, and 1931).

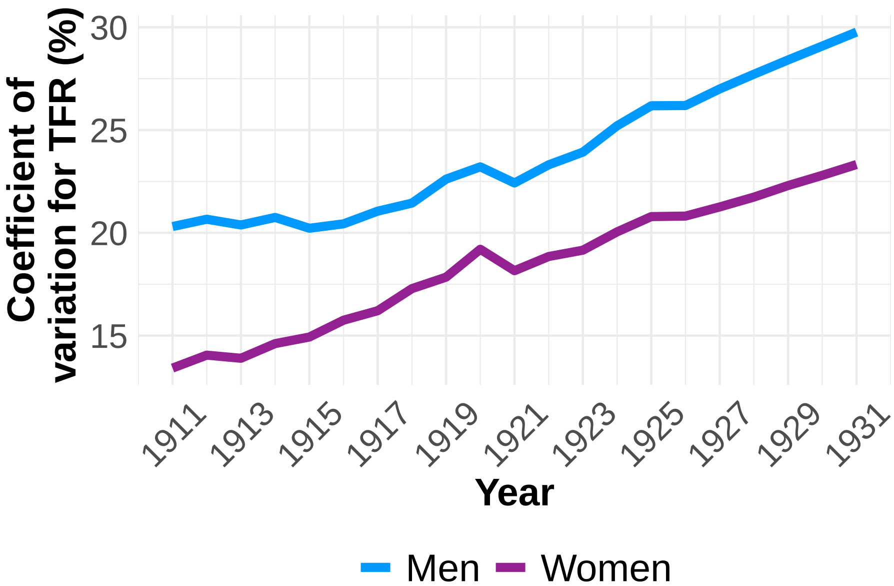


(a) Men



(b) Women

Figure 3: Coefficient of Variation in male and female TFR estimates across Italian provinces from 1911 to 1931.



References

Breschi, M., Esposito, M., Mazzoni, S., and Pozzi, L. (2014). Fertility transition and social stratification in the town of alghero, sardinia (1866–1935). *Demographic Research*, 30:823–852.

Breschi, M., Fornasin, A., Gonano, G., Mazzoni, S., and Manfredini, M. (2020). Male fertility between biology and the socioeconomic context: News from the past (alghero, 1866–1935). *Economics & Human Biology*, 37:100833.

Breschi, M., Kurosu, S., and Michel, O., editors (2003). *The Own-Children Method of Fertility Estimation: Applications in Historical Demography.* Editrice Universitaria Udine, Udine.

Coale, A. J. and Watkins, S. C., editors (1986). *The Decline of Fertility in Europe: The Revised Proceedings of a Conference on the Princeton European Fertility Project.* Princeton University Press, Princeton, NJ.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Livi Bacci, M. (1977). *A History of Italian Fertility During the Last Two Centuries.* Princeton University Press, Princeton, NJ.

Livi Bacci, M. and Breschi, M. (1990). Italian fertility: An historical account. *Journal of Family History*, 15(4):385–408.

Omenti, R., Alexander, M., and Barban, N. (2025). A bayesian model to estimate male and female fertility patterns at a subnational level. In *Proceedings of the 30th International Population Conference (IUSSP 2025)*, Brisbane, Australia. Session 174: New Approaches and Tools in Fertility Research.

Rettaroli, R., Samoggia, A., and Scalone, F. (2017). Does socioeconomic status matter? the fertility transition in a northern italian village (marriage cohorts 1900–1940). *Demographic Research*, 37:455–492.

Schmertmann, C. P. and Hauer, M. E. (2019). Bayesian estimation of total fertility from a population's age–sex structure. *Statistical Modelling*, 19(3):225–247.

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M., and Sawyer, C. (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies*, 66(1):1–28.