[Journal of Demography](https://digital.car.chula.ac.th/jdm)

Volume 39
Issue 2 *December 2023*

Article 1

January 2024

Subnational probabilistic projections of fertility: rethinking from Latin America

Lucia Andreozzi
Universidad Nacional de Rosario - CONICET, andreozzi.lu@gmail.com

Journal of Demography Volume 39 Number 2 December 2023

Subnational probabilistic projections of fertility: rethinking from Latin America

Lucia Andreozzi[1]

Abstract

General trends in fertility, mortality, and migration can be discerned and projected into the future with reasonable results. However, considerable uncertainty is attached to each specific trend in a particular country or region. Thus, subnational projections represent a special chapter within demographic projections. After introducing subnational projections and its close relationship with fertility, this work proposes as its main objective to project fertility rates at the subnational level for Argentina using a probabilistic method, i.e., the Bayesian hierarchical model (BHM), and then compare the results with the point estimates of deterministic projections published by its national agency for statistics. Bayesian hierarchical modelling is a statistical model written at multiple levels that estimates the parameters of the posterior distribution using the Bayesian method. In the second stage, forecasts were obtained from two models, one including all the countries available in the World Population Prospects (WPP) and a second model based only on a subgroup of countries (Argentina, Colombia, Chile, Cuba and Uruguay), both of which are based on data from 1980 to 2010. This set of countries was selected using the transition to identify a subset of countries with similar patterns that includes Argentina, but showing different patterns compared with the rest of the region. A time period is selected to include years for which these countries have achieved adequate data quality. BHM for subnational projections is an extremely useful and flexible method, presenting many advantages over classical methods. The Bayesian framework is a powerful scheme to generate national and subnational projections for mortality, fertility, and population. This research aims to address the limitations and uncertainties in existing subnational population projections and reinforce the use of probabilistic models: specifically, BHM, which respects data and avoids forcing it to get caught up in a mathematical assumption.

Keyword: Argentine, Bayesian Hierarchical Models, Probabilistic Projections, Subnational Projections, Total Fertility Rate, World Population Prospects.

---

[1] Lucia Andreozzi, Universidad Nacional de Rosario, andreozzi.lu@gmail.com
Received: 20-Aug-2023; Revised: 15-Nov-2023; Accepted: 15-Dec-2023

Introduction

While demographers and non-demographers commonly use population projections in their work, only in recent years have such forecasts become a main subject of study among population scientists. General trends in fertility, mortality, and migration can be discerned and projected into the future with reasonable results. However, considerable uncertainty is attached to each specific trend in a particular country or region. As a higher degree of disaggregation (spatial, temporal, subgroups, etc.) is introduced into any statistical estimation, greater methodological challenges arise, and demographic projections are no exception; thus, subnational projections represent a special chapter within demographic projections. The quantification of uncertainty, i.e., the probability associated with a forecast, becomes a key point for users of projections since they consider the different implications of a result. In the most widely used forecasts, uncertainty is expressed through alternative scenarios based on the variation of the trajectory for fertility, but rarely for mortality and migration. High and low scenarios are used to cover a range of possible futures. However, no specific probability measure is attached to this range, and its meaning is therefore ambiguous. Each component's projections are presented as inputs for the projection of the population, and they are given no interest or analysis outside of this forecast. On the other hand, the subnational projections produced by most national statistical offices are based on non-probabilistic methods, as is the case with projections at the national level.

To construct a population projection, mortality and fertility are considered as main inputs, while migration, given the complexity in its measurement and registration, presents a smaller number of associated methodological developments. The study of fertility has been and continues to be a key point for understanding a population's dynamics. Biological and social factors affect the number of sons or daughters one has (intensity) and the moment in which one has them (calendar). Given the diversity of factors that affect fertility, projecting it into the future is a complex task. According to some studies by Chackiel and Schkolnik (2004), based on the estimates and projections produced by national statistical offices and the Latin American Demographic Center (CELADE) in Latin America, the period from 1950-55 to 1985-90 experienced a significant decline in fertility. For the five-year period 1950-55, Argentina and Cuba registered medium-low fertility, while Uruguay presented low fertility, ranking among the countries with the lowest global fertility rates in the region. In particular, the case of fertility transition in Argentina is interesting because it occurred earlier than in many Latin American countries; the evolution of mortality and fertility bears little resemblance to the well-known "classical" shape of the transitional model (Pantelides, 1983).

United Nations projections for "developing"[1] countries assume that the total fertility rate will eventually reach replacement level and then fall slightly. For regions that have already achieved a fertility level higher than the replacement level, a small increase is expected. Subnational projections are based on the extrapolation of the global fertility rates according to the expected evolution for each country, considering the relative weights of the provinces and based on mathematical functions such as logistics. Méndez Borges (2015) presents one of the first applications of subnational projections with a probabilistic model for Brazil, and states that in recent years, national statistical offices have been including uncertainty in their official population projections (Abel et al., 2010), although they continue to use projections of deterministic variants to provide uncertainty (Lutz and Goldstein, 2004). While probabilistic approaches receive criticism linked to the difficulty of interpreting the results (Lee, 1998), at the same time they appear as promising and innovative methods to project fertility, especially when considering a Bayesian approach (Abel et. al., 2010).

After introducing subnational projections and their close relationship with fertility, this work proposes as its main objective to project fertility rates at the subnational level for Argentina using a probabilistic method, i.e., the Bayesian hierarchical model (BHM), and then compare the results with the point estimates of deterministic projections published by INDEC (2013).

---

[1] The distinction between developed and underdeveloped countries, although attributed to Wilfred Benson in a 1942 text, has been used discreetly. In his inaugural speech in 1949, U.S. president Harry S. Truman used the word "underdevelopment" to identify a specific calamity affecting most human beings, thus drawing the line that separates developed from underdeveloped countries and labeling a good part of the world with this second concept (Esteva, 2009).
Attempts have been made to soften and even blur the duality between developed and underdeveloped countries, which have been assumed to be dichotomous realities, with terms such as "Highly Developed Countries", "Developing Countries", and "Less Developed Countries" in World Bank and UN documents. These new terms diminish the pejorative load (no one would be proud to live in an "underdeveloped" country), while keeping alive the optimism that the promised development will one day be achieved (Ramirez, 2008).

This subset of subnational probabilistic projections is derived from the national projections presented at the XVI Argentine Conference on Population Studies and at the XIV Latin American Congress of Statistical Societies.

The remainder of this paper is structured as follows. The next section describes the method and data used in the analysis, i.e., the Bayesian hierarchical model for subnational projections of fertility and data sources, followed by an overview of fertility in Argentina and the Latin American region. Section 3 presents a selection of results generated by two models, one including all the countries available in WPP (Model 1) and the second based on a subgroup of countries: Argentina, Colombia, Chile, Cuba, and Uruguay (Model 2). Both are based on data from 1980 to 2010 and applied to generate results for Argentina but are in principle applicable to other countries as well. These results are compared with the official projections of fertility. The paper concludes with a discussion of the results, limitations, and an outlook regarding opportunities for further research.

Methods

Subnational projections are the byproduct of a Bayesian projection model proposed by Alkema et al. (2019) based on five-year estimates of the fertility rate from 1950-1955 to 2015-2020. Its implementation is available in R through the bayesTFR package (***Ševčíková***, H, 2011). The model's basic approach states that the evolution of the total fertility rate (TFR) divides into three broad phases: a pre-transitional phase of high fertility; a second phase of fertility transition in which the TFR declines from high fertility levels to replacement fertility level or below it; and a post-fertility transition phase, which includes the recovery of fertility to replacement level and oscillations around that level. The BHM does not include Phase I since it is assumed that in the following period, fertility will reach Phase II and, on this basis, the TFR projections are calculated. For Phase II, fertility is modeled using a random walk with drift given by,

$$f_{c,t+1} = f_{c,t} - d_{c,t} + \varepsilon_{c,t}, \quad for\ \tau_c \leq t < \lambda_c, \tag{1}$$

where $f_{c,t}$ is the TFR for the period $(t, t+5)$ in country c, $d_{c,t}$ is the decline that represents the systematic decrease in fertility during the transition, $\varepsilon_{c,t}$ is the random error, and τ_c and λ_c are the initial and end moments of the transition stage. The error distribution is given by

$$\varepsilon_{c,t} \sim \begin{cases} N(m_\tau, s_t^2) & for\ \ t = \tau_c \\ N\left(0, \sigma(f_{c,t})^2\right) & otherwise \end{cases} \tag{2}$$

where m_τ is the mean and s_t^2 the standard deviation at the start period of the phase, τ_c and $\sigma(f_{c,t})$ is the deviation for the remaining values of t. Finally, $d_{c,t}$ is modeled as a function of the TFR level and the vector $\boldsymbol{\theta}_c$ as follows:

$$d_{c,t} = d(\boldsymbol{\theta}_c, \lambda_c, \tau_c, f_{c,t}) = \begin{cases} g(\boldsymbol{\theta}_c, f_{c,t}) & for\ f_{c,t} > 1 \\ 0 & otherwise \end{cases} \tag{3}$$

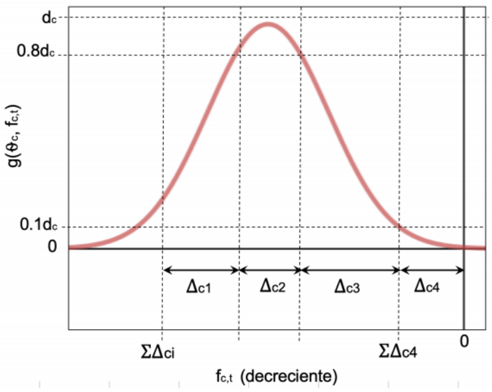
where $g(.,.)$ is a parametric diminution function. The decline function is the sum of two logistic functions, and the same for a specific country is given by

$$\frac{-d_c}{1+exp\left(-\frac{2\ln(p_1)}{\Delta_{c1}}(f_{c,t} - \sum_i \Delta_{ci} + 0{,}5\Delta_{c1}\right)} + \frac{d_c}{1+exp\left(-\frac{2\ln(p_2)}{\Delta_{c3}}(f_{c,t} - \Delta_{c4} + 0{,}5\Delta_{c3}\right)}, \tag{4}$$

where $\boldsymbol{\theta} = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$ is the vector of country-specific parameters; d_c is the maximum decline; $p_1 = p_1 = 9$ are constants; the Δ_{ci} describe the TFR ranges between which the rate of decline changes; and $U_c = \sum_{i=1}^{4} \Delta_{ci}$ is the starting level of fertility decline.

The decline parameters are estimated for each country, and for the countries in which the beginning of Phase II is within the observed data period, U_c is set equal to the value of the TFR in that period, that is, $U_c = f_{c,\tau_c}$. For countries in which Phase II began prior to the observation period, the initial level is added as another parameter to the model.

Fig. 2: Decline function



Given the start period, U_c, the five parameters that determine the pace of fertility decline and the time that the transition takes in country are c; Δ_{c4}, $\left\{\frac{\Delta_{ci}}{U_c-\Delta_{c4}}: i = 1,2,3\right\}$ and d_c. The parameters in the model in this phase are estimated in a Bayesian framework:

$$d_c^* = log\left(\frac{d_c-0,25}{2,5-d_c}\right), \quad (5)$$

$$d_c^* \sim N(\chi, \psi^2), \quad (6)$$

$$\Delta_{c4}^* = log\left(\frac{\Delta_{c4}-1}{2,5-\Delta_{c4}}\right), \quad (7)$$

$$\Delta_{c4}^* \sim N(\Delta_{c4}, \delta_4^2), \quad (8)$$

$$p_{ci} = \frac{\Delta_{ci}}{U_c-\Delta_{c4}} \; for \; i = 1,2,3 \;, \quad (9)$$

$$p_{ci} = \frac{\exp(\gamma_{ci})}{\sum_{j=1}^{3}\exp(\gamma_{cj})}, \quad (10)$$

$$\gamma_{ci} \sim N(\alpha_i, \delta_i^2), \quad (11)$$

where $\{\chi, \psi^2, \Delta_4, \delta_4, \boldsymbol{\alpha}, \boldsymbol{\delta}\}$ are the mean and variance of parameters.

The TFR projections for Phase II countries are based on the posterior distribution of the model parameters, and the median is used as the projection given its robustness and simplicity. In Phase III or post-transition, the change in TFR is modeled through an AR(1) process with mean $\mu = 2.1$, which is an approximation of the TFR for fertility at replacement level.

$$f_{c,t+1} \sim N\left(\mu + \rho\left(f_{c,t} - \mu\right), s^2\right) \; for \; t \leq \lambda_c \;, \qquad (12)$$

where ρ, which is the autoregressive parameter, with $|\rho| < 1$, and s is the standard deviation of the random errors, both estimated through maximum likelihood.

For the TFR projection through the Bayesian hierarchical model, the start times of phases II and III are estimated for each country. An a-posteriori sample of the model parameters is then obtained using a MCMC (Markov chain Monte Carlo) algorithm, and from these we generate the TFR's future trajectories.

National projections are the basis for subnational projections. Initially, this paper proposes two methods: the first is related to the general approach for world and country data, giving the country the role of the world and the subregions/provinces the role of the countries. The second option, based on Watkins (1990, 1991), states that the intra-country variation of the TFR decreased during the demographic transition period between 1870 and 1960. Although it is based on data from European countries, Watkins relates this phenomenon to the greater integration of national markets, the expansion of the role of the state, and the construction of nations based on linguistic standardization. Delving deeper into her hypothesis, it holds that after fertility transition begins (1870 in most European countries), intra-country demographic variability both in marital fertility and the proportion of marriages increases, but at the end of the period, around 1960, intra-country variability is less than that observed in 1870. This reduction suggests the importance of the national community in modern demographic behavior. In her analysis, Watkins speculates on the role of changes in the social structure, particularly the integration of national markets and the expansion of the role of state processes, which Tilly (1981) set as the two dominant processes of the modern era and how this influenced demographic uniformity. On the other hand, Méndes Borges (2017) discusses the hypothesis of convergence and divergence in fertility and mortality at the subnational level in Latin America and then applies a set of indicators to Brazilian data. Fertility patterns in Latin

America have varied widely from those observed in European countries, with extreme socioeconomic inequalities such as educational levels generating differentials among countries. Examining the data at subnational levels, Méndes Borges concludes that indicators like the Coefficient of Quartile Deviation and Variation Coefficient show a similar trend, reaching maximum dispersion in 1980 and 1991 with a sharp reduction in 2000 and 2010. These conclusions must be considered when evaluating model assumptions.

Finally, Alkema et al. (2019) show that the second approach, based on the Watkins hypothesis, presents better results, compared with the one that changes roles in the hierarchy, given that the method proposed by **Ševčíková** et. al (2018) produces correlations like those observed in the data, and is simple and easy to implement. Based on each national probabilistic projection of the TFR, they are "scaled" through an autoregressive factor of order 1, AR(1), producing a set of future TFR trajectories at the subnational level; furthermore, it is even possible to convert each trajectory into age-specific fertility rates. The analysis is based on the modelling of 47 countries for which information is available, which cover 1,092 regions. The rationale behind the chosen study period (1980 to 2010) is, for the beginning of the period, data quality and, for the end of the period, the last available census. Argentina began having acceptable data quality in 1980, with the last available census taking place in 2010.

The first option consists of applying a region-specific scale factor to each national trajectory.

$$f^{(R)}_{r_c,t,i} = \alpha_{r_c} \cdot f^{(C)}_{c,t,i} \quad (13)$$

$$\alpha_{r_c} = f^{(R)}_{r_c,t=P} / f^{(C)}_{c,t=P} \quad (14)$$

where $f^{(R)}_{r_c,t,i}$ is the TFR for region r_c of country c at time t in the i-th trajectory; $f^{(C)}_{c,t,i}$ is the TFR projection at the national level of country c at time t in the i-th trajectory, that is, the product of the estimation of the BHM; α_{r_c} is the scale factor (invariant in time); and P is the last observed period. However, this first approach produces perfectly correlated probabilistic trajectories, and does not allow crossings between the trajectories obtained for each region, whose occurrence is implausible. Therefore, Alkema et al. (2019) proposes a second option of allowing the scale factor to change slowly over time.

$$f_{r_c,t,i}^{(R)} = \alpha_{r_c,t} . f_{c,t,i}^{(C)} \quad (15)$$

where

$$\alpha_{r_c,t} - 1 = \emptyset\left(\alpha_{r_c,t-1} - 1\right) + \varepsilon_{r_c,t} \quad (16)$$

with

$$\varepsilon_{r_c,t} \sim^{iid} N(0, \sigma_c^2) \quad (17)$$

and $\alpha_{r_c,t}$ converges to a distribution centered on 1. These estimates are obtained by using data from 47 countries (see estimation process in **Ševčíková**et. al., 2018).

$$\emptyset = 0.925 \text{ and } \sigma = 0.045 \quad (18)$$

$$\sigma_c^2 = min\left\{\sigma^2, (1 - \emptyset^2) Var_{r \in R_c}(\alpha_{r,t=P})\right\} \quad (19)$$

where P again denotes the present time and R_c denotes the set of regions of country c. The restriction in (19) ensures that the variation in the scale factor across regions is not greater than the variation in the last observed time, in line with the hypothesis of Watkins (1990, 1991) and the observed long-term data. The last option has multiple advantages over the others. First, it is a simple probabilistic method, based on the national BHM, and captures the correlation present in the data. In addition, it has a reasonable out-of-sample validation, is in line with Watkins's (1990, 1991) hypothesis that TFRs within countries converge in response to country-specific factors, and is available in the bayesTFR R package.

According to the world population projections developed by the United Nations (2019), the fertility level in Argentina is below the world average but above the regional average. In the five-year period 2015-2020, the Global Fertility Rate was 2.5 children per woman worldwide, 2.0 in Latin America and the Caribbean, and 1.9 in South America, while in Argentina it was 2.3 children per woman.

In a previous work presented at the XVI Argentine Conference on Population Studies and at the XIV Latin American Congress of Statistical Societies, national projections were developed for Argentina. Pantelides (1983) analyzed the case of the Argentinian fertility transition; it is interesting because it occurred early in the Latin American context; the evolution of mortality and fertility bears little resemblance to the well-known "classical" pattern of the transitional model. Although the start of the decline in fertility in Argentina occurred early compared with that in most Latin American countries, the slowdown in its decline over the past few decades has led to Argentina finding itself in a situation like that in many countries whose decline began later, even with a higher total fertility rate than several of them.

The transitional model allows the identification of a subset of countries with similar patterns among them, including Argentina, but shows different patterns compared with the rest of the region. Chackiel and Schkolnik (2004) state that according to estimates and projections produced by national agencies and the Latin American Demographic Center (CELADE), Latin America experienced a significant decline in fertility between 1950-55 and 1985-90. For the five-year period 1950-55, Argentina and Cuba had medium-low fertility and Uruguay low fertility, ranking among the countries with the lowest global fertility rates in the region.

Tab. 1: Latin America, countries according to fertility level 1950-1955 and 1985-1990

Fertility level 1950-1955	Fertility level 1985-1990			
	High	Medium-High	Medium-Low	Low
High	Guatemala Honduras Nicaragua	Bolivia El Salvador Haití Paraguay	Brazil Costa Rica Ecuador México Panamá Perú Dominica Venezuela	Colombia
Medium-High				Chile
Medium-Low				Argentina Cuba
Low				Uruguay

Source: Chackiel & Schkolnik 2004

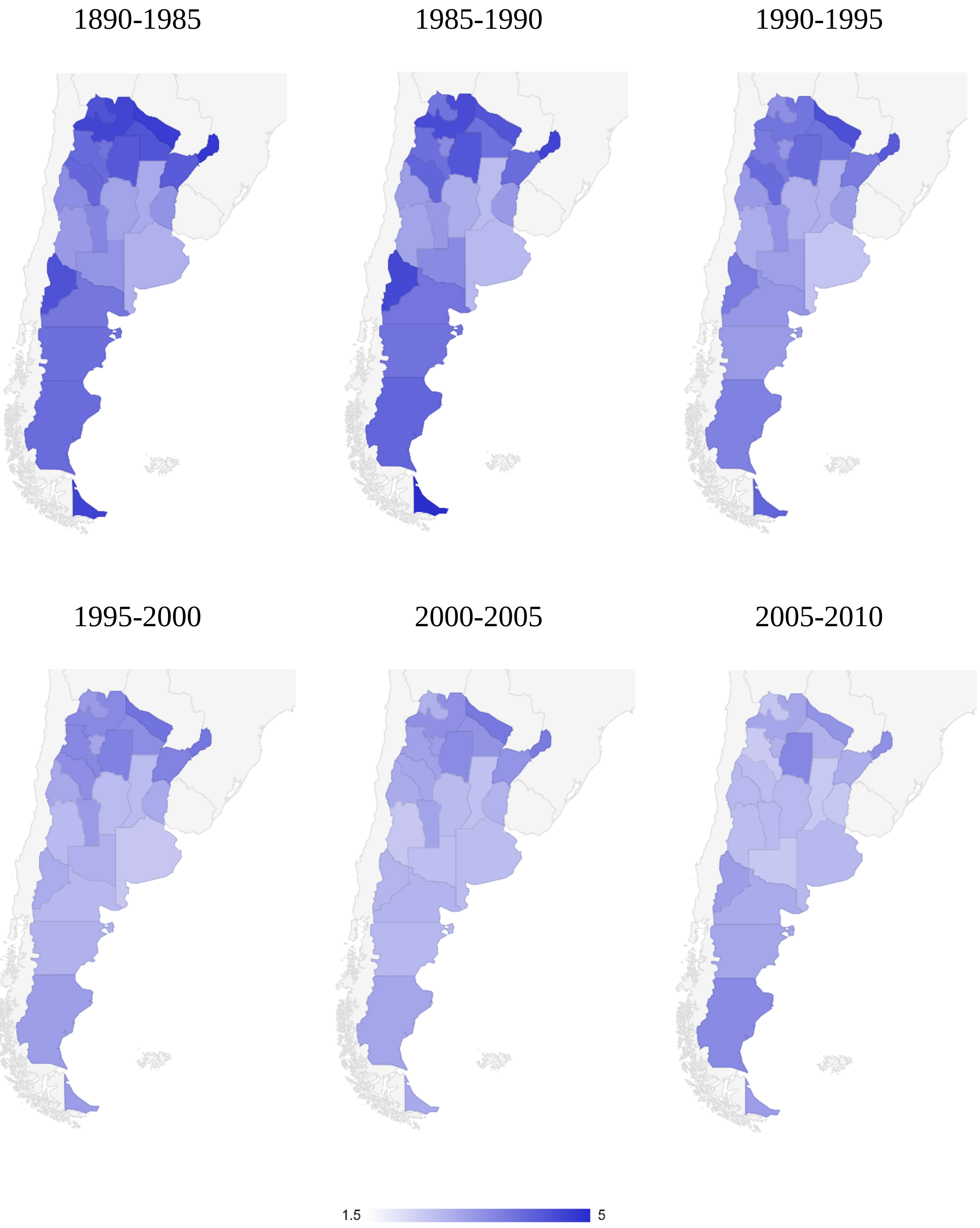
Based on WPP data used in the model, four indicators were constructed to evaluate fertility convergence through 1980-2010. Table 2 shows that convergence is a plausible hypothesis since range and variance decrease consistently towards the end of the period. Moreover, Figure 1 displays the maps that show how provinces reached similar colors in the final years.

Tab. 2: Total fertility rate, descriptive statistics by provinces, Argentina (1980-2010)

	Max	Min	Variance	Range
1980-1985	4,78	1,99	0,49	2,79
1985-1990	4,94	2,01	0,52	2,93
1990-1995	4,39	1,73	0,36	2,66
1995-2000	3,79	1,74	0,22	2,06
2000-2005	3,71	1,80	0,18	1,92
2005-2010	3,47	1,81	0,15	1,66

Source: Author's elaboration based on data from WPP, 2019

Fig. 1: Total fertility rate by province, Argentina 1890-2010



Source: Author's elaboration based on data from WPP, 2019

Finally, the model developed by **Ševčíková** et. al. (2018) is estimated based on the data set of subnational total fertility rates compiled in 2012 by Patrick Gerland (with contributions from Julien Thillard, Thomas Spoorenberg, and Danan Gu) and based on information provided by the official data sources of each country and academic demographic studies. The reliability of these data varies among countries, but for most, the fertility estimates are based on birth registration data. Data sources for Argentina are:

• Pantelides, Edith Alejandra (2006). The Fertility Transition in Argentina 1869-1947. CENEP, Center for Population Studies, CENEP Notebook No. 54

• Pantelides, Edith Alejandra (1989). Argentine fertility since the mid-twentieth century. CENEP, Center for Population Studies, CENEP Notebook No. 41

• National Institute of Statistics and Census, INDEC (2012). Dynamics and structure of the population. Crude birth rate by province, 1980 – 2009

To explore the data, TFRs are mapped for all provinces from 1980 to 2010 and descriptive statistics are calculated. Table 1 shows the decrease in the range of fertility rates in Argentina over the years, with a decrease in variance of around 70%; in the same sense, the choropleth maps show the homogenization in the rates for an entire territory, supporting the Watkins hypothesis that gives rise to the model for subnational projections under the Bayesian approach.

Results

Forecasts were obtained from two models, one including all the countries available in WPP (Model 1) and a second model based only on a subgroup of countries: Argentina, Colombia, Chile, Cuba and Uruguay (Model 2). Both models are based on data from 1980 to 2010. The selection criteria for the countries included in the comparative analysis relied on the analysis of demographic transition for Latin America made by Chackiel and the corresponding classification observed in Table 1. These nations were chosen because of specific similarities in their demographic patterns. Moreover, BHM "borrows" information from other countries instead of the official UN projects that use all countries, so it is possible to select countries whose data could contribute more due to similarities. There is not a region "logic"; thus, a geographic region can contain big differences between countries. A time period is selected to include years for which these countries achieve an adequate data quality. Finally, based on these models, two sets of subnational projections were developed (Table 3).

Tab. 3: TFR projections by provinces, Argentina 2020-2100

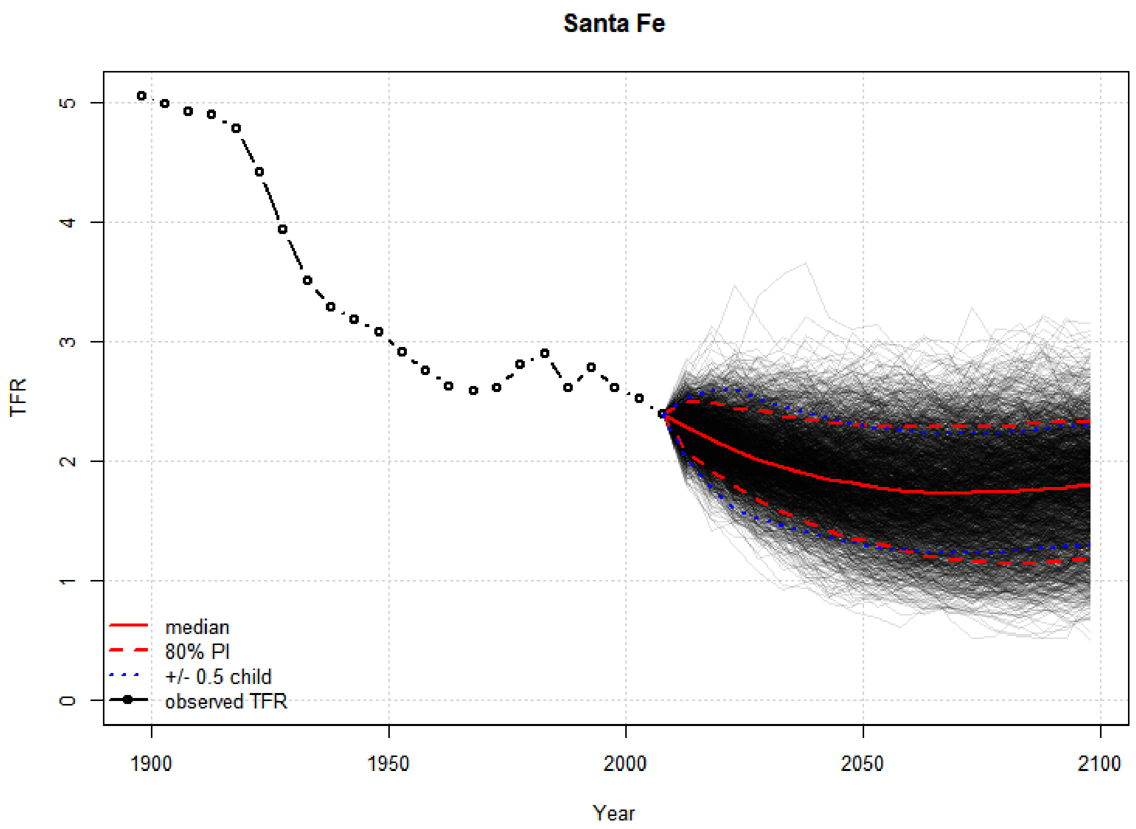
	Model 1			Model 2		
	2020-2025	2050-2055	2095-2100	2020-2025	2050-2055	2095-2100
Buenos Aires	2.31	1.93	1.86	2.30	1.87	1.86
Catamarca	2.09	1.81	1.80	2.09	1.76	1.79
Chaco	2.37	1.97	1.88	2.36	1.92	1.87
Chubut	2.51	2.04	1.92	2.51	1.98	1.92
Buenos Aires City	1.70	1.61	1.70	1.69	1.56	1.69
Córdoba	2.29	1.93	1.86	2.29	1.88	1.87
Corrientes	2.43	2.00	1.90	2.41	1.95	1.89
Entre Ríos	2.11	1.83	1.81	2.11	1.78	1.81
Formosa	2.70	2.16	1.98	2.71	2.10	1.97
Jujuy	2.15	1.84	1.83	2.15	1.79	1.81
La Pampa	2.11	1.83	1.82	2.11	1.78	1.81
La Rioja	2.24	1.90	1.85	2.23	1.84	1.83
Mendoza	2.24	1.90	1.85	2.23	1.84	1.84
Misiones	2.77	2.20	1.98	2.76	2.13	1.99
Neuquén	2.60	2.09	1.94	2.59	2.04	1.94
Río Negro	2.45	2.01	1.89	2.44	1.95	1.91
Salta	2.50	2.05	1.92	2.50	1.99	1.91
San Juan	2.29	1.92	1.85	2.29	1.88	1.86
San Luis	2.28	1.92	1.85	2.26	1.86	1.84
Santa Cruz	2.83	2.23	2.00	2.82	2.16	1.99
Santa Fe	2.11	1.84	1.81	2.09	1.76	1.80
Santiago del Estero	2.87	2.24	2.01	2.86	2.18	2.02
Tierra del Fuego	2.61	2.11	1.95	2.60	2.06	1.95
Tucumán	2.38	1.98	1.88	2.37	1.91	1.87

Source: Author's elaboration based on data from WPP, 2019

Among highlights of the results of the projections, Buenos Aires City shows a recovery in TFR after a slight decrease. This is interesting since, given the model's characteristics, it allows estimated increases preceded by decreases. Santa Cruz and Santiago del Estero present the highest levels at the beginning of the forecast period, but then decrease towards replacement level by the end of the centennial, at which no province exceeds 2.02 children per woman.

On the other hand, the quantification of uncertainty is one of the most outstanding properties presented by probabilistic models; Table 4 include forecasts with their intervals and plus or minus a half of a child, as an analogy of how the United Nations constructs its high and low variants. Each forecast is reported with an associated interval and a confidence coefficient. This coefficient, which is set at 95% in many cases, is often set at 80% when we are working with demographic data, given that social indicators with rare exceptions present greater variability than do experimental data, whose large quantity of factors that are not of interest in a study can be controlled and measured. Figure 3 shows the TFR future distribution (in black) from which the median (forecast) and the percentiles (forecast interval limits) are estimated.

Fig. 3: TFR projections, Province of Santa Fe, Argentina, 2010-2100



Source: Author’s elaboration based on data from WPP, 2019

Tab. 4: Forecasts 2050-2055 (Model 1) with 80% confidence intervals and variant ±0.5 children, by province

Province	Forecast	Lower limit	Upper limit	-0.5 child	+0.5 child
Buenos Aires	1.93	1.39	2.46	1.43	2.43
Catamarca	1.81	1.28	2.32	1.31	2.31
Chaco	1.97	1.42	2.52	1.47	2.47
Chubut	2.04	1.46	2.62	1.54	2.54
Buenos Aires City	1.61	1.14	2.08	1.11	2.11
Córdoba	1.93	1.38	2.45	1.43	2.43
Corrientes	2.00	1.43	2.56	1.50	2.50
Entre Ríos	1.83	1.30	2.35	1.33	2.33
Formosa	2.16	1.54	2.74	1.66	2.66
Jujuy	1.84	1.29	2.34	1.33	2.33
La Pampa	1.83	1.29	2.34	1.33	2.33
La Rioja	1.90	1.36	2.42	1.40	2.40
Mendoza	1.90	1.36	2.41	1.40	2.40
Misiones	2.20	1.55	2.77	1.70	2.70
Neuquén	2.09	1.51	2.66	1.59	2.59
Río Negro	2.01	1.44	2.57	1.51	2.51
Salta	2.05	1.46	2.61	1.55	2.55
San Juan	1.92	1.38	2.46	1.42	2.42
San Luis	1.92	1.36	2.44	1.42	2.42
Santa Cruz	2.23	1.60	2.81	1.73	2.73
Santa Fe	1.84	1.30	2.36	1.34	2.34
Santiago del Estero	2.24	1.60	2.82	1.74	2.74
Tierra del Fuego	2.11	1.51	2.68	1.61	2.61
Tucumán	1.98	1.42	2.51	1.48	2.48

Source: Author's elaboration based on data from WPP, 2019

Finally, it is interesting to compare forecasts from the model with those from the National Institute of Statistics and Censuses (INDEC), namely, the official subnational projections. INDEC subnational projections (Table 5) were developed setting initial levels for the TFR, the specific fertility rates, and the global fertility rate for the year 2010. To estimate these levels, data were used from births from vital statistics corrected by late registration, with the female population from the 2010 Census corrected by a demographic evaluation process. For the projection of fertility, the global fertility rates were extrapolated according to the expected evolution for country level, a similar procedure to that used to project life expectancies by considering the relative weights of the provinces. For country level, the TFR trend is set based on historical data and adjusting a logistic function to estimate future values for the period 2010-2040.

Tab. 5: INDEC official projections of the TFR 2025-2040

Province	Total fertility rate			
	2025	2030	2035	2040
Buenos Aires	2.11	2.05	2.01	1.98
Catamarca	2.08	2.04	2.00	1.97
Chaco	2.18	2.11	2.06	2.02
Chubut	2.10	2.05	2.01	1.99
Buenos Aires City	1.85	1.85	1.86	1.86
Córdoba	2.04	2.00	1.97	1.95
Corrientes	2.17	2.10	2.06	2.03
Entre Ríos	2.08	2.03	2.00	1.98
Formosa	2.25	2.16	2.11	2.06
Jujuy	2.12	2.07	2.03	2.00
La Pampa	2.06	2.02	1.99	1.97
La Rioja	2.01	1.99	1.97	1.96
Mendoza	2.15	2.08	2.05	2.02
Misiones	2.28	2.19	2.12	2.07
Neuquén	2.15	2.09	2.04	2.01
Río Negro	2.08	2.04	2.01	1.99
Salta	2.27	2.18	2.11	2.06
San Juan	2.21	2.14	2.08	2.04
San Luis	2.07	2.02	1.99	1.96
Santa Cruz	2.18	2.10	2.05	2.00
Santa Fe	1.99	1.96	1.95	1.94
Santiago del Estero	2.21	2.15	2.11	2.07
Tierra del Fuego	2.07	2.00	1.97	1.95
Tucumán	2.16	2.10	2.06	2.02

Source: INDEC. Projections made based on the 2010 National Population Census.

The differences between these methods and the classical or mathematical ones lie in the input data: the mathematical model uses one country's information, while BHM borrows strength from the use of many countries (a subset or all of them). Moreover, the stability (or the lower variability) observed in the demographic components for Argentina generates results with high certainty and small prediction intervals. The choice of using Bayesian hierarchical models (BHM) over other methods is a pivotal aspect of the research; this is the official method selected by the United Nations to elaborate probabilistic projections.

The selection criteria for the countries included in the comparative analysis depends on the analysis of demographic transition for Latin America made by Chackiel and his given classification. These nations were chosen because of specific similarities in their demographic patterns.

Given that BHM borrows information from other countries instead of the official UN projections that use all countries, it is thus possible to select countries whose data could contribute more due to similarities; even so, this does not imply that the subset corresponds to a geographical area. An established region can contain big differences between countries.

The most remarkable result is the negative net difference for Buenos Aires City; Bayesian model projections are smaller than INDEC ones, with a slight difference in trend. Although it is interesting that some regions such as CABA show bigger differences than others, this could be due to INDEC methodology, which does not use the correlation between provinces as a part of its assumptions and extrapolates country level trends to provinces, while BHM allows each province to have its own trend. On the other hand, for Santa Cruz, Santiago and Tierra del Fuego, Bayesian projections are greater than those of INDEC, but differences decrease through the years. A third group of provinces present moderate differences in Misiones, Río Negro, and Formosa. The rest of the country’s subregions present similar forecasts in the two models. The sources of these differences are many, including that the first starting level for the INDEC projection is fixed, while BHM considers the time series trend. There are many reasons to consider the superiority of BHM over the classical mathematical method used by INDEC.

Tab. 6: Differences between the official projections of INDEC and the BHM for the TFR 2025-2040

Province	Differences			
	2025	2030	2035	2040
Buenos Aires	0.16	0.12	0.08	0.05
Catamarca	-0.02	-0.04	-0.06	-0.08
Chaco	0.15	0.12	0.09	0.07
Chubut	0.23	0.18	0.14	0.10
Buenos Aires City	-0.17	-0.19	-0.22	-0.24
Córdoba	0.21	0.16	0.12	0.08
Corrientes	0.21	0.18	0.14	0.10
Entre Ríos	0.00	-0.02	-0.05	-0.07
Formosa	0.40	0.36	0.30	0.26
Jujuy	0.00	-0.03	-0.06	-0.08
La Pampa	0.02	-0.01	-0.04	-0.07
La Rioja	0.20	0.13	0.08	0.03
Mendoza	0.05	0.03	-0.01	-0.04
Misiones	0.43	0.38	0.34	0.29
Neuquén	0.39	0.33	0.28	0.23
Río Negro	0.33	0.26	0.21	0.15
Salta	0.19	0.17	0.14	0.11
San Juan	0.04	0.02	0.01	-0.01
San Luis	0.16	0.12	0.08	0.05
Santa Cruz	0.59	0.52	0.45	0.40
Santa Fe	0.09	0.05	0.00	-0.03
Santiago del Estero	0.60	0.51	0.42	0.36
Tierra del Fuego	0.49	0.43	0.36	0.29
Tucumán	0.18	0.13	0.10	0.07

Source: Author’s elaboration based on data from WPP, 2019, and 2010 National Census

Conclusion

BHM for subnational projections is an extremely useful and flexible method, presenting many advantages over the classical methods. The Bayesian framework is a powerful scheme to generate national and subnational projections for mortality, fertility, and population. Its application for Argentina sees small differences between both models: one including all the countries available in WPP (Model 1), and the second model based only on a subgroup of countries: Argentina, Colombia, Chile, Cuba and Uruguay (Model 2). Both models are based on data from 1980 to 2010 and show the robustness of this approach. There are no remarkable differences between the models and, in the long-term, the projections are almost identical. This fact supports the model that includes all countries. The regional selection does not introduce more or additional information to the estimation stage. The differences appear between BHM and INDEC estimations; the BHM model allows probabilistic results, while the other model is based on a mathematical function; it is a deterministic model by which the alternative scenarios are the projected TFR plus or minus a half of a child. This paper reinforces the use of probabilistic models. Moreover, BHM respects data; it doesn't force it to get caught up in a mathematical assumption, though it is important to remark that BHM is very complex for most users. Even so, this new approach should be beneficial for demographic analysts, policymakers, and society. It offers robustness and considers uncertainty by offering prediction intervals, thus providing extreme scenarios that would be extremely useful in practice because it provides bandwidth with an associated probability.

This study has limitations. First, it tested one country, so it would be interesting to test the same logic (all countries and a subset) with countries at different stages in their demographic transition. Second, it would be interesting to test it versus classical time series models and not only the mathematical approach. Future research must be done with mortality estimates to finally integrate them into the most important product, namely, subnational population projections.

References

Abel, G. J., Bijak, J., Raymer, J. (2010). A comparison of official population projections with Bayesian time series forecasts for England and Wales. *Popul Trends.* 141, 92-111. doi: 10.1057/pt.2010.23.

Lee, R. D. (1998). Probabilistic approaches to population forecasting. *Population and Development Review*, 24 (supplement), 156–190.

Borges, G.M. (2015). Subnational fertility projections in Brazil: A Bayesian probabilistic approach application. Paper presented at the Annual Meeting of the Population Association of America, San Diego, USA, April 30–May 2, 2015.

Méndez Borges, G. (2017). Theories and measures of demographic convergence: an pplication to subnational levels in Latin America in ¿Convergencia demográfica? Análisis comparativo de las tendencias demográficas subnacionales en América Latina y el Caribe. Río de Janeiro: Asociación Latinoamericana de Población.

Chackiel, J. (2004). La transición de la fecundidad en América Latina 1950-2000. *Papeles de población*, *10*(41), 9-58.

Esteva, G. (2009). Más allá del desarrollo: la buena vida in La agonía de un mito ¿cómo reformular el desarrollo? Junio, Año XXIII, II época, pp. 7-17. http://www.alainet.org/publica/445.phtml

Instituto Nacional de Estadística y Censos INDEC. (2012). *Dinámica y estructura de la población. Tasa bruta de natalidad por provincia. Años 1980-2009*. http://www.indec.gov.ar/principal.asp?id_tema=7924, 6.7.22

Lutz, W., Goldstein, J. R. (2004). Introduction: How to deal with uncertainty in population forecasting? *International Statistical Review*. 72, 1–4.

Pantelides, E. A. (2006). La Transición de la fecundidad en la Argentina 1869-1947. CENEP, Centro de Estudios de Población, Cuaderno del CENEP No. 54.

Pantelides, E. A. (1989). La Fecundidad Argentina desde Mediados del Siglo XX. CENEP, Centro de Estudios de Población, Cuaderno del CENEP No. 41

Pantelides, E. A. (1983). La transición demográfica argentina: un modelo no ortodoxo. CENEP

Raftery, A. E. et. al. (2009). White paper: Probabilistic projections of the total fertility rate for all countries for the 2010 World Population Prospects.

Ramírez, H. E. (2008). Desarrollo, subdesarrollo y teorías del desarrollo en la perspectiva de la geografía crítica. *Revista Escuela de Historia*, *1*(7), 1-23

Ševčíková, H, Raftery A. E., Gerland, P. (2018). Probabilistic projection of subnational total fertility rates. *Demographic Research*, *38*(60), 1843-1884.

Ševcıková, H., Alkema, L., Raftery, A. (2011). bayesTFR: An R package for probabilistic projections of the total fertility rate. *Journal of Statistical Software* 43, 1–29.

Tilly, C. (1981). *As sociology meets history*. Academic Press.

Watkins, S. C. (1990). From local to national communities: The transformation of demographic regimes in Western Europe, 1870–1960. *Population and Development Review*, *16*(2), 241–272.

Watkins, Susan C. (1991). *From provinces into nations: Demographic integration in Western Europe, 1870–1960.* Princeton University Press.