

# Uncertainty in population projections: the state of the art

Raquel Rangel de Meireles Guimarães\*

In this paper I critically review the state of the art in population projections, focusing on how uncertainty is handled in three approaches: the classical cohort-component, the frequentist probabilistic model and the Bayesian paradigm. Next, I focus on recent developments on mortality, fertility and migration projections under the Bayesian setting, which have been clearly at the frontier of knowledge in demography. By evaluating the merits and limitations of each framework, I conclude that in the near future the Bayesian paradigm will offer the most promising approach to population projections, since it combines expert opinion, information that demographers have readily available from their empirical analyses and sophisticated statistical and computational methods to deal with uncertainty. Hence, the availability of population forecasts that take uncertainty carefully into account may enhance communication among demographers by allowing for greater flexibility in reflecting demographic beliefs.

**Keywords:** Population projections. Uncertainty. Cohort-component model. Frequentist approach. Bayesian approach.

---

\* Department of Economics at Federal University of Paraná (UFPR), Curitiba, Paraná, Brazil (raquel.guimaraes@ufpr.br).

## Introduction

The field of population projections is probably one of the richest in the area of demographic research. Keyfitz (1972) defines demographic forecasting as “the search for functions of population that are constant through time, or about which fluctuations are random and small.” According to the same author, uncertainty surrounding population estimates should be expressed in the form of probability distributions (KEYFITZ, 1972).

Notwithstanding this position of Keyfitz, demographers have traditionally conducted population forecasts using the cohort-component model from a deterministic perspective. In short, the cohort-component method involves a number of steps, each of which takes demographers’ expert opinions into consideration. Uncertainty is introduced into the projections at all phases of the process by means of the demographers’ judgment and experience. Based on their opinion and on past data analysis, a single-variant deterministic projection is built upon the most reasonable future behavior of the demographic components: fertility, mortality and migration. Hence, “projecting a population becomes an art influenced by scientific techniques” (DAPONTE et al., 1997, p. 1257).

In terms of the uncertainty of population estimates, the standard population projection approach does not allow demographers to explicitly state the probability that given demographic events will occur. In order to overcome this limitation, demographers began projecting the population using different sets of variants, reflecting the uncertainty in their estimates and leaving it to the users to choose the projections that best fit their needs. This is the so-called variant approach to population projections. Up until the 2010 Revision, this was the standard approach used by the UN Population Division to produce population estimates (UNITED NATIONS, 2009). Even though the variant approach represented an advance in demographic forecasts, some authors have argued that it has no probabilistic basis, resulting in inconsistencies in demographic estimates (GIROSI; KING, 2008; LEE; TULJAPURKAR, 1994).

Since the 1990s, demographers have attempted to incorporate uncertainty into population estimates by dealing with it in the form of probability distributions. The rationale behind probabilistic methods is that population forecasting is “unavoidably uncertain” (WILSON; REES, 2005, p. 340), especially when forecasts are for certain age groups, due to uncertainties in each of the components of the demographic dynamics – fertility, mortality and migration. Lee and Carter pioneered in incorporating uncertainty into demographic estimates in a probabilistic framework using the frequentist probabilistic projection approach. They projected mortality trends by extrapolating time series parameters (LEE; CARTER, 1992; LEE, 1992; LEE; TULJAPURKAR, 1994). Later, scholars from the IASA team set out to derive projections based on expert judgment (LUTZ et al., 1998; LUTZ; SCHERBOV, 1998). Frequentist probabilistic projections provide a useful approach for assessing change and deviations of mortality outcomes from the most likely scenario by quantifying the uncertainty in terms of probability. However, this approach has some problematic issues

that still remain to be solved. For instance, Girosi and King (2008) argue that the properties of the model developed by Lee and colleagues will not fit demographers' beliefs as to the future patterns of mortality. Also, it is argued that the frequentist probabilistic approach does not address the uncertainty of confidence intervals and of population estimates in an integrated approach (BIJAK, 2011).

Recently, Bayesian statistics have been stirring up interest in a variety of scientific fields as a result of the development of analytical tools and methods and the advancement of computational techniques. In demographic forecasting the Bayesian theory provides a clear framework to deal with both uncertainty and subjective assumptions. It also provides straightforward tools for making predictions (ALKEMA et al., 2011; BIJAK, 2011; DAPONTE et al., 1997; GIROSI; KING, 2008; PEDROZA, 2006). In this context, there is a growing opinion in the mainstream of population forecasting that the future belongs to Bayesian probabilistic predictions (BIJAK, 2011; GIROSI; KING, 2008).

While there have been clear advancements towards sophisticated population forecasts in academic research, equivalent advancements in demographic practice are still incipient. For instance, in 2010 the UN Population Division began developing probabilistic projections for the total fertility rate (TFR) in their 2010 Revision of the World Population Prospects in collaboration with researchers from the Center for Statistics and the Social Sciences (CSSS) at the University of Washington (ALKEMA et al., 2011; CHUNN et al., 2010; RAFTERY et al., 2012, 2014). Currently, the 2012 Revision of the World Population Prospects derives probabilistic projections of total fertility, and life expectancy at birth, for all countries and areas (UNITED NATIONS, 2013). In Brazil, population forecasts are developed by the Brazilian Institute of Geography and Statistics (IBGE) following the most likely scenario for total fertility rates, life expectancy and child mortality rates, and no uncertainty measure is provided with their estimates (IBGE, 2008). However, if decision makers were able to explicitly deal with forecast uncertainty, this would lead to better decision making. For instance, educational policy makers may benefit from probabilistic estimates of the number of children in order to draw up relevant measure (KEILMAN, 2008). Meaningful applications may also be speculated for public health services and pension systems.

Given the promising advances in Bayesian population projections and the fact that improvements in demographic forecasts can definitely contribute to better policy decisions, my goal with this paper is to critically review the state of the art in the field of population projections. My focus, however, is to review how uncertainty is incorporated into the derivation of population estimates, and also to evaluate the merits and limitations of each framework in demographic applications. I also focus on a detailed explanation of the research on Bayesian formulations for mortality, fertility and migration projections. I believe that, in the near future, the Bayesian paradigm will be the predominant approach to population projections, since it combines expert opinion, information that demographers have readily available from their empirical analyses and sophisticated statistical and computational methods to deal with

uncertainty.<sup>1</sup> Furthermore, the Bayesian framework offers elegant and appropriate solutions to missing data problems.

It is hoped that this paper will increase the awareness of the academic demographic community on the importance of the availability of forecasts with their respective uncertainties to improve policy planning and evaluation. Also, concern for the development of robust population forecasts will enhance communication among demographers by allowing better observation of the sources of uncertainty in the estimates.

## **Population projections: state of the art**

In this section I provide a brief overview of recent advances in the field of population projections. I describe the classical cohort-component model and then present the frequentist probabilistic projection model. Finally, I discuss the Bayesian projection model.

### *Classical cohort-component model*

The cohort-component method is widely used by demographers to make population projections (ARRIAGA et al., 1994). This method is based on the demographic balancing equation, which states that the future population is a function of the previous population plus the number of births and the number of immigrants, minus the number of deaths and the number of emigrants. In short, this method is applied for each sex and age group and relies on separate models for the demographic components – fertility, mortality and migration – which are later combined using the balancing equation to derive population totals (PRESTON et al., 2000).

The first step in the cohort-component method is to derive a base mid-year population. By evaluating the quality of this initial measure through experience and scientific knowledge, the demographer may adjust this population for age misreporting and under- and/or over-enumeration (PRESTON et al., 2000). Next, the demographer works on assumptions regarding fertility levels. In general, the total fertility rate is projected and then a set of age-specific fertility rates is assumed for the projection horizon. The same rationale is applied to the projection of mortality, where the life expectancy is projected and a set of age-specific death rates is assumed. Then, based on empirical regularities and judgments, the demographer assumes a level and pattern of net migration rates. These three sets of age-specific rates – fertility, mortality and migration – are applied to the base population to derive the projected population in a given year.

The deterministic approach to the cohort-component model is currently the dominant projection practice in demography. In its deterministic version, uncertainty in the cohort-component model is usually incorporated through the demographer's judgment of the most likely set of elements that will result in future changes in fertility, mortality and migration components. Generally a single-variant deterministic projection is derived, that is, "a single,

---

<sup>1</sup> Specifically in the population projections field, see, for instance, the *bayesTFR*, *bayesLife*, *bayesPop* and *bayesDem R* packages (RAFTERY et al., 2014).

best-guess population projection that assumes moderate levels of fertility, life expectancy, and migration in the future” (O’NEILL, 2005, p. 231).

To allow for some degree of uncertainty, the UN approach developed multi-variant projections until the 2008 Revision (UNITED NATIONS, 2009). In this approach, various scenarios were developed to reflect uncertainty around fertility levels, although not in probabilistic terms. A clear limitation of the scenario approach resides in the lack of integrated assessment. As Lee (1992) points out, UN medium variants are carefully treated by the specialists, while the high and low variants are the result of simplistic assumptions. Moreover, users may interpret that the interval between the high and low variants contain the actual future values for the population size, when in fact this interval has no probabilistic meaning (LEE; CARTER, 1992; LEE, 1992).

*Frequentist probabilistic projection model*

As a response to the limitations of deterministic models in providing consistent measures of uncertainty to demographic estimates, a probabilistic model for mortality projection was proposed by Lee and colleagues in the early 1990s. Known as the Lee-Carter model, it is now used by the U.S. Census Bureau as a benchmark for its population forecasts.

The Lee-Carter (LC) model allows the derivation of long-term forecasts of the level and age pattern of mortality and fertility and is based on matrix algebra. In this review I will discuss further the application of the LC model to project mortality rates. Consider  $m$  as a matrix of log-mortality rates,  $m = A \times T$ , where  $A$  refers to the number of age groups and  $T$  to time periods. The entry in the  $a$ -th row and  $t$ -th column of a matrix  $A$  is  $m_{a,t}$ , the central death rate for age group  $a$  in year  $t$ . The LC model postulates a linear relationship between the log-death rates and a parameter  $k$ :

$$\ln(m_{a,t}) = \alpha_a + \beta_a k_t + \varepsilon_{a,t} \quad (1)$$

where the average shape of the age profile is given by the  $\alpha_a$  coefficients and  $\beta_a$  gives the deviations from the average age profile when  $k$  varies. Since Eq. 1 is not estimable, a least-squares solution is found by using the first element of the singular value decomposition (LEE; CARTER, 1992). Since the solution for this problem may not be unique, the authors impose two restrictions to the model:

$$\sum_{a=1}^A \beta_a = 1 \quad (2)$$

$$\sum_{a=1}^A K_t = 0 \quad (3)$$

From Eq. 2, it implies that  $\alpha_a$  is the empirical average of the log-mortality rate in age group  $a$ . Since it is assumed that the disturbance term is normally distributed, it follows that:

$$m_{a,t} \sim N(\mu_{a,t}, \sigma^2) \quad (4)$$

in which the average estimator for  $m_{a,t}$  is given by:

$$\mu_{a,t} = \bar{m}_a + \beta_a k_t \quad (5)$$

Hence, the LC method assumes the absence of age and time interactions. That is,  $\beta_a$  is fixed over periods for all  $a$ , and  $k_t$  is fixed by age groups for all  $t$ . After estimating the parameters using a single-value decomposition, forecasts are conducted assuming that  $\beta_a$  is constant over time and employing an ARIMA model to the series of  $k$ . The authors found that a random walk with drift was appropriate for their data (LEE; CARTER, 1992), and their model allowed the computation of confidence intervals for the projected life expectancies.

Besides being a pioneer step in the application of a probabilistic method to demographic forecasts, the LC model has other very appealing characteristics. First, it is relatively simple and parsimonious, with a small number of parameters. Second, it relies on a relational model, preserving important features of the mortality pattern of the population of interest (LEE; MILLER, 2001). Also, the LC model may be applied to other applications besides the projection of mortality rates. For instance, Rodrigues (2010) implemented the LC approach to model hospitalization rates for Brazil.

Despite its manageability and applicability to several different demographic applications, the LC model nonetheless has its limitations. Girosi and King (2008) argued that the LC model tends to perform well on the short run, given slow changes in mortality, but it may lead to inconsistent long-run forecasts. Also, the LC model may produce implausible changes in projected age profiles given the independence of separate age-group forecasts. Furthermore, Thomas (2002) argued that the model tends to gradually forget the age pattern of death rates when it approaches low levels of mortality, and nothing prevents death rates from falling to zero in the LC model. Lee himself recognized that the confidence interval for the forecasts tends to be narrow as a result of the low entropy of the survival curve in contexts of high-levels of life expectancy (LEE, 2000).

Despite these restrictions, the methodological coherence of the LC model was an important breakthrough in demographic projections. Since it was introduced, several authors have worked to develop stochastic methods for this purpose (LEE; TULJAPURKAR, 1994).

### *Bayesian probabilistic projections*

Bayesian inference implies a subjectivist stances. By subjectivist stance, I mean a way of thinking whereby a probability is assigned to uncertainty. If an individual is coherent, then uncertainty measures follow the laws of probability (DE FINETTI, 1937). Second, Bayesian statisticians hold that empirical evidence is relevant to the understanding of a given phenomenon, but they also consider prior knowledge when making inferences.

Formally, the Bayesian paradigm postulates that, for an unknown quantity  $\theta$  and sample information provided in a vector  $x$ , the likelihood function  $L(x|\theta)$  provides empirical information on  $\theta$ : it is the probability of observing the sample given  $\theta$ . The prior distribution  $\pi(\theta)$  represents the initial uncertainty on  $\theta$ . Hence, the Bayesian inference on  $\theta$  is made in terms of the posterior distribution,  $\pi(\theta|x)$ , where:

$$\pi(\theta|x) \propto \pi(\theta)L(x|\theta) \quad (6)$$

In short, the Bayesian approach assumes the existence of a probability distribution that includes the knowledge, intuition or belief of a researcher with respect to the possible values of  $\theta$ , unconditional on the empirical evidence available from data. Hence, the essence of Bayesian inference is to transform prior beliefs and uncertainty about  $\theta$  and  $\pi(\theta)$  to the posterior knowledge,  $\pi(\theta|x)$ , by incorporating empirical evidence,  $L(x|\theta)$ .

The Bayesian paradigm offers appealing features for demographic forecasting. The issue of uncertainty in population forecasts requires the use of many subjective assumptions. To address this issue, the Bayesian framework is appropriate because knowledge and beliefs can be expressed in terms of the prior distribution. Also, forecasts are a natural analytical tool of Bayesian inference: the predictive distribution for  $\theta$  can be computed through the posterior distribution. In the past, Bayesian analysis was limited by the computing resources available at each moment, but, with recent developments in computing and analytical methods, complex tasks can now be carried out. There is an increasing interest in conducting Bayesian demographic analysis, and studies have been developed to model fertility schedules in small areas (ASSUNÇÃO et al., 2005, 2002; POTTER et al., 2010). In the field of projections, it is worth noting the work of Girosi and King (2008), which employed a Bayesian hierarchical model to predict mortality rates using information pooling from similar cross-sections (i.e., age groups, countries). Bijak (2011) applied the Bayesian paradigm to model and project the path of international migration in Europe. More details on the methodology of Bijak (2011) and Girosi and King (2008) are described in the next section, which is devoted solely to Bayesian formulations.

Recently, a team from the Center for Statistics and the Social Sciences (CSSS) at the University of Washington developed population estimates using Bayesian methods (ALKEMA et al., 2011; CHUNN et al., 2010; HEILIG et al., 2010; RAFTERY et al., 2012). Their approach followed the cohort-component model, but mortality and fertility were modeled using a Bayesian framework. Their Bayesian model employed to project total fertility is composed of two steps. The first model is aimed at projecting total fertility rates (TFR) in a first phase of the fertility transition, in which fertility decreases from high to low levels, reaching replacement-level. The team used a Bayesian approach to estimate the parameters of a bi-logistic function fitted for data for most countries worldwide, and the posterior distribution of the parameters for each country reflects its own fertility decline as well as the experience of all countries combined. A second model was used to project the second phase of fertility transition, when low fertility prevails: using a first-order autoregressive time-series frequentist model with a mean set at replacement-level, they generated random fluctuations of the TFR around the mean. For the mortality projection, a Bayesian model was employed to predict future paths of male life expectancy and female life expectancy assuming country-specific sex differentials observed in UN Population Division Projections. The CSSS team has not attempted to produce a probabilistic projection of international migration. From projected paths of future fertility and mortality, the scholars produced median population estimates with the paths corresponding to the 95 percent confidence intervals for all countries over the 2010-2050 period.

## Advancements in population projections using the Bayesian approach

### Mortality

The Bayesian model for mortality projections proposed here is inspired by the work of Girosi and King (2005). These authors presented a Bayesian hierarchical modeling approach for predicting mortality rates by pooling information from similar cross-sections (i.e., age groups, countries). By incorporating considerable information that demographers have on observed mortality data and future patterns, their model shows outstanding performance.

Girosi and King’s model is formalized as follows. Consider  $i, i = 1, \dots, N$  cross-sections and  $t, t = 1, \dots, T$  time periods. We observe  $d_{it}$ , which is the number of deaths in the cross-section  $i$  and time  $t$ . Define the log-mortality rate  $m_{it}$  as:

$$m_{it} = \frac{d_{it}}{p_{it}} \quad (7)$$

where  $p_{it}$  is the population at the cross-section  $i$  and time  $t$ . We assume a linear specification for the log-mortality:

$$m_{it} \sim N(\mu_{it}, \sigma_i^2) \quad (8)$$

$$\mu_{it} = Z_{it} \beta_i \quad (9)$$

where  $\mu_{it}$  is the expected log-mortality rate,  $m_{it}$  is assumed independent over  $t$  conditional on  $Z$ . Girosi and King noted that  $Z$  may include lagged terms of  $m_i$ .

The Bayesian model for the log-mortality rates assumes that the parameters  $\sigma_i^2$  and  $\beta_i$  are unknown and fixed quantities, and that their a priori uncertainty is expressed in terms of probability distributions. For  $\sigma$ , the prior distribution is  $P(\sigma)$ . For  $\beta$ , assume that its prior distribution is a function of hyperparameters  $\theta$ ,  $P(\beta|\theta)$ . Hyperparameters are defined as the parameters of a prior distribution, which are also random and have a hyperprior distribution,  $P(\theta)$ . The functional form of  $P(\sigma)$  and  $P(\theta)$  are chosen to be non-informative (flat). That means that  $\sigma$  and  $\theta$  are uniformly distributed between  $-\infty$  and  $+\infty$ . Put simply, the mean and variance of  $\sigma$  and  $\theta$  are set to be diffuse so they do not have an important effect on the results. On the other hand,  $P(\beta|\theta)$  is assumed to be highly “informative” in the sense that it is implemented by using the following kind of prior knowledge: “similar” cross sections (as defined by adjacent countries or age-groups) should have “similar” coefficients.

Assuming that a priori is independent of  $\beta$  and  $\theta$ , the posterior distribution is given by:

$$P(\beta, \sigma, \theta | m) \propto P(m | \beta, \theta) [P(\beta | \theta) P(\theta) P(\sigma)] \quad (10)$$

where the prior distribution is  $P(\beta, \sigma, \theta) \equiv P(\beta | \theta) P(\theta) P(\sigma)$ . We can summarize updated information using the posterior mean (Bayes estimator):

$$\beta^{Bayes} = \int \beta P(\beta, \sigma, \theta | m) d\beta d\sigma d\theta \quad (11)$$

Given this information, Girosi and King described three methods for forecasting: forecasting covariates; autoregressive models; and lagged covariates. Based on experimentation, the authors emphasized that the third strategy is preferable because it allows for a combination of statistical methods and expert judgment.



The great challenge in Girosi and King's approach is to define a prior distribution for  $\beta$ . Specifying a prior on the vector  $\beta$  requires nonsample knowledge about it, as I argued above. Because  $\beta$  refers to population measures, this vector does not always imply causal effects. The authors illustrated this issue by assuming tobacco consumption as covariate to explain changes in mortality. Even though this relationship is clear at the individual level, the relationship between tobacco consumption and lung mortality may be confused at the aggregate level by the country's development level. Also, demographers may feel uncomfortable regarding opinions about the similarity of coefficients. Rather, they may know fairly precisely how the expected value of mortality varies across cross sections.

To overcome limitations in the specification of a prior distribution of the coefficients, Girosi and King proposed a two-step strategy. First, they derive a prior density of the expected value of the mortality rates. Then, priors on the regression coefficients are specified. Priors on the expected value of the mortality rates are set based on genuine prior knowledge of age, time and country trends using a smooth approximation. As a result, posterior distributions are derived for the parameters using a Markov Chain Monte Carlo (MCMC) approach.

Finally, besides the inclusion of covariates, Girosi and King's model allows for five sources of prior knowledge: the age profile toward which the forecasts will tend, due to the smoothing imposed; the cross sections across which the researcher expects to see similar levels or trends in log-mortality; the degree of similarity and smoothing by setting the weight of the prior; the degree of smoothing imposed in different areas by choosing weights and relaxing the prior; and our ignorance, by setting the content of the null space, which is the portion of the forecasts that depend entirely on the data and are not influenced by the prior.

In summary, the appealing feature of Girosi and King's model is precisely the incorporation of prior knowledge for constructing projections within a sophisticated statistical framework. Thus, I believe that this method of projection of mortality seems to be quite adequate for demographic projections.

### *Fertility*

The best known Bayesian model for projecting fertility was developed by the CSSS team to simulate future values of the total fertility rate (TFR) for all countries (ALKEMA et al., 2011). Their approach accounts for three stages of the fertility transition: Phase I, Pre-Transition with Stable and High-Fertility; Phase II, Fertility Transition; and Phase III, Post-Transition and Low Fertility. However, models were developed only for Phases II and III as all countries have begun their demographic transition, and only the Phase II model is adopted in a Bayesian paradigm.

The model for the fertility transition can be briefly described as follows. Five-year decrements in the TFR are decomposed into a systematic decline plus random disturbances. The decline function is evaluated according to the level of the TFR at time  $t$  for country  $c$ ,  $f_{c,t}$ , and to some country-specific parameters  $\theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$  using a double logistic specification:

$$g(\theta_c, f_{c,t}) = \frac{-d_c}{1 + \exp\left(-\frac{2 \ln(9)}{\Delta_{c1}} (f_{c,t} - \sum_{i=1}^4 \Delta_{ci} + 0.5 \Delta_{c1})\right)} + \frac{d_c}{1 + \exp\left(-\frac{2 \ln(9)}{\Delta_{c3}} (f_{c,t} - \Delta_{c4} - 0.5 \Delta_{c3})\right)} \quad (12)$$

where  $U_c = \sum_{i=1}^4 \Delta_{ci}$  refers to different starting levels of the TFR decline,  $\Delta_{c4}$  refers to the level at the end of the transition,  $d_c$  refers to the overall pace of the fertility transition and the ratio  $\Delta_{ci}/(U_c - \Delta_{c4}), i = 1,2,3,4$  refers to differences in the timing of acceleration or deceleration in the pace of decline. Hence, each value of the decline parameters,  $\theta_c = (\Delta_{c1}, \Delta_{c2}, \Delta_{c3}, \Delta_{c4}, d_c)$ , refers to a special case of fertility decline.

As in Girosi and King’s model described previously, the CSSS team employs a Bayesian hierarchical model as a strategy to estimate the decline parameters  $\theta_c$  and their associated uncertainty. This model allows estimating the parameters for a given country taking into account information from other countries. This is an appealing feature of their model, as there are few observations for each country and values for each country gravitate toward a world mean. In other words, the Bayesian hierarchical model assumes that the prior distribution for the decline parameters is best described by the range of the decline parameters for all countries. Next, the prior distribution is updated using a country’s observed decline. Therefore, the posterior distribution summarizes information on a world pattern of fertility decline and the country’s experience. Computations are performed using MCMC algorithms.

In sum, the approach to project fertility levels proposed by the CSSS team is promising. However, it is not completely grounded in the Bayesian paradigm – only the demographic transition phase is modeled using a Bayesian hierarchical model. Therefore, further developments in this area are strongly encouraged.

### Migration

It is widely recognized by demographers that migration is the most complex component of demographic change. As Coleman (2008, p. 453) stated:

Of the three components of demographic change, data on migration are far below the quality of those on birth and death. Of the three, its theory is the least satisfactory, its trend by far the most volatile, and its future by far the most difficult to forecast. It is the only demographic component, at least potentially, under substantial and direct policy influence, which adds to the difficulty of prediction. Even its definition is unsatisfactory.

Although clearly overstated, Coleman’s argument summarizes some of the challenges involved in migration modeling and forecasting. Hence, future migration estimates are generally subject to considerable errors, which prevent derivation of accurate forecasts (KEILMAN, 2008).

Despite the difficulties described above, there is an increasing need for more accurate forecasts of migration flows. On one hand, policy makers are interested in numerical estimates of migrants to evaluate, for instance, their impact on labor markets. On the other hand, migration forecasts are inputs for population forecasts. Despite their importance and

probably due to the challenges in measurement and modeling, future migration flows are generally: projected using deterministic or scenario approaches; projected stochastically, but sometimes simplistically; or ignored (BIJAK, 2011).

However, uncertainty is definitely an important component in migration studies, perhaps more so than in fertility and mortality. To deal with uncertainty in migration projections, there are clear advantages of the Bayesian framework: as migration is a complex and multidimensional phenomenon and is therefore subject to a high level of uncertainty, the Bayesian paradigm offers a suitable approach.

As far as I know, the work of Bijak (2011) is one of the first attempts to summarize existing efforts to conduct migration forecasting and to compare different Bayesian models to derive migration estimates. Bijak claims that an averaging approach is desirable because there is no clear evidence towards a unique model that would provide better migration estimates. Hence, the Bayesian forecast averaging model allows for merging of the features of various predictive models and providing an interesting strategy to account for the uncertainty. The application of Bayesian models for migration modeling is a very novel approach in demographic forecasting, which will require a great deal of research. This method, nevertheless, will not be formalized in this article.

## Conclusion

In this paper I critically review the state of the art in population projections, focusing on how uncertainty is handled in each approach. I gave a special focus to novel Bayesian methods to produce population forecasts. I believe that the development of Bayesian demographic forecasts promises three important advantages for demographic research. First, such an approach represents a formal and probabilistic framework to deal with uncertainty; second, it will improve communication among demographers by allowing for greater flexibility in reflecting demographic beliefs; and third, it will provide a range of future estimates of the demographic components for policy makers. This approach is certainly promising and is at the frontier of knowledge in demography.

## References

ALKEMA, L.; RAFTERY, A. E.; GERLAND, P. et al. Probabilistic projections of the total fertility rate for all countries. **Demography**, v. 48, n. 3, p. 815-839, Aug. 2011.

ARRIAGA, E.; JOHNSON, P.; JAMISON, E. **Population analysis with microcomputers**. Washington D.C.: U.S. Census Bureau, 1994, v. II (Extract B).

ASSUNÇÃO, R. M.; POTTER, J. E.; CAVENAGHI, S. M. A Bayesian space varying parameter model applied to estimating fertility schedules. **Statistics in Medicine**, v. 21, n. 14, p. 2057-2075, 2002.

ASSUNÇÃO, R.; SCHMERTMANN, C.; POTTER, J.; CAVENAGHI, S. Empirical bayes estimation of demographic schedules for small areas. **Demography**, v. 42, n. 3, p. 537-558, 2005.

BIJAK, J. **Forecasting international migration in Europe: a Bayesian view**. New York: Springer, 2011.

CHUNN, J. L.; RAFTERY, A. E.; GERLAND, P. **Bayesian probabilistic projections of life expectancy for all countries**. [S.l.]: University of Washington, 2010.

COLEMAN, D. The demographic effects of international migration in Europe. **Oxford Review of Economic Policy**, v. 24, n. 3, p. 452-476, 21 Set. 2008.

DAPONTE, B. O.; KADANE, J. B.; WOLFSON, L. J. Bayesian demography: projecting the Iraqi Kurdish population, 1977-1990. **Journal of the American Statistical Association**, v. 92, n. 440, p. 1256-1267, 1 Dec. 1997.

DE FINETTI, B. La prevision: ses lois logiques, ses sources subjectives. **Reprinted in breakthroughs in statistics: foundations and basic theory (1992)**. [S.l.]: Ann. Inst. Henri Poincaré, 1937.

GIROSI, F.; KING, G. **Demographic forecasting**. Princeton, NJ: Princeton University Press, 2008.

HEILIG, G.; BUETTNER, T.; LI, N. et al. **A probabilistic version of the United Nations World population prospects: methodological improvements by using Bayesian fertility and mortality projections**. Lisbon, Portugal: [s.n.]. Available at: <[http://esa.un.org/peps/doc/S6\\_WP\\_14\\_Lisbon%20proceedings\\_Rev7.pdf](http://esa.un.org/peps/doc/S6_WP_14_Lisbon%20proceedings_Rev7.pdf)>. 28 Apr. 2010

IBGE – Instituto Brasileiro de Geografia e Estatística. **Projeção da população do Brasil por sexo e idade, 1980-2050**: revisão 2008. Rio de Janeiro: IBGE, 2008.

KEILMAN, N. European Demographic forecasts have not become more accurate over the past 25 years. **Population and Development Review**, v. 34, n. 1, p. 137-153, 2008.

KEYFITZ, N. On future population. **Journal of the American Statistical Association**, v. 67, n. 338, p. 347-363, 1 Jun. 1972.

LEE, R. The Lee-Carter method for forecasting mortality, with various extensions and applications. **North American Actuarial Journal**, v. 4, n. 1, p. 80-93, 2000.

LEE, R.; CARTER, L. Modeling and forecasting U. S. mortality. **Journal of the American Statistical Association**, v. 87, n. 419, p. 659-671, 1992.

LEE, R. D. Stochastic demographic forecasting. **International Journal of Forecasting**, v. 8, n. 3, p. 315-327, Nov. 1992.

LEE, R. D.; TULJAPURKAR, S. Stochastic population forecasts for the United States: beyond high, medium, and low. **Journal of the American Statistical Association**, v. 89, n. 428, p. 1175-1189, 1 Dec. 1994.

LEE, R.; MILLER, T. Evaluating the performance of the Lee-Carter method for forecasting mortality. **Demography**, v. 38, n. 4, p. 537-549, 2001.

LUTZ, W.; SANDERSON, W. C.; SCHERBOV, S. Expert-based probabilistic population projections. **Population and Development Review**, v. 24, p. 139-155, Jan. 1998.

LUTZ, W.; SCHERBOV, S. An expert-based framework for probabilistic national population projections: the example of Austria. **European Journal of Population/Revue Européenne de Démographie**, v. 14, n. 1, p. 1-17, 1998.

O'NEILL, B. C. Population scenarios based on probabilistic projections: an application for the millennium ecosystem assessment. **Population & Environment**, v. 26, n. 3, p. 229-254, 2005.

PEDROZA, C. A Bayesian forecasting model: predicting U.S. male mortality. **Biostatistics**, v. 7, n. 4, p. 530-550, 1 Oct. 2006.

POTTER, J. E.; SCHMERTMANN, C. P.; ASSUNÇÃO, R. M.; CAVENAGHI, S. M. Mapping the timing, pace, and scale of the fertility transition in Brazil. **Population and Development Review**, v. 36, n. 2, p. 283-307, 2010.

PRESTON, S.; HEUVELINE, P.; GUILLOT, M. **Demography**: measuring and modeling population processes. [S.l.]: John Wiley & Sons, 2000.

RAFTERY, A. E.; ALKEMA, L.; GERLAND, P. Bayesian population projections for the United Nations. **Statistical Science**, v. 29, n. 1, p. 58-68, Feb. 2014.

RAFTERY, A. E.; LI, N.; SEVCIKOVA, H.; GERLAND, P.; HEILIG, G. K. Bayesian probabilistic population projections for all countries. **Proceedings of the National Academy of Sciences**, v. 109, n. 35, p. 13915-13921, Aug. 2012.

RODRIGUES, C. G. **Dinâmica demográfica e internações hospitalares**: uma visão prospectiva para o Sistema Único de Saúde (SUS) em Minas Gerais, 2007 a 2050. Belo Horizonte: Centro de Desenvolvimento e Planejamento Regional – Cedeplar/Universidade Federal de Minas Gerais – UFMG, 2010.

THOMAS, B. Approaches and experiences in projecting mortality patterns for the oldest-old. **North American Actuarial Journal**, v. 6, n. 3, p. 14, 2002.

UNITED NATIONS. **World population prospects**: the 2008 revision, highlights and advance tables. New York: Department of Economic and Social Affairs, Population Division, 2009.

\_\_\_\_\_. **World population prospects**: the 2012 revision, highlights and advance tables. New York: Department of Economic and Social Affairs, Population Division, 2013.

WILSON, T.; REES, P. Recent developments in population projection methodology: a review. **Population, Space and Place**, v. 11, n. 5, p. 337-360, 2005.

## Author

*Raquel Rangel de Meireles Guimarães* is Associate Professor at the Department of Economics at Federal University of Paraná (UFPR). She has PhD degree in Demography from the *Centro de Desenvolvimento e Planejamento Regional – Cedeplar* of the *Universidade Federal de Minas Gerais – UFMG*, has master degree in International Comparative Education, from the Stanford University, and master degree in Demography from Cedeplar/UFMG.

## Address

Av. Prefeito Lothário Meissner, 632 - Térreo  
Jardim Botânico  
80210-170 - Curitiba/PR - Brasil

## Resumo

### *Incertezas nas projeções populacionais: o estado da arte*

O artigo apresenta, inicialmente, uma revisão crítica do estado da arte em projeções de população, focando em como a incerteza é tratada em três abordagens: no modelo clássico de coorte-componente; no modelo probabilístico frequentista; e no paradigma bayesiano. Em seguida, a análise se concentra sobre desenvolvimentos recentes nos modelos de projeções bayesianos de fecundidade, mortalidade e migração, os quais têm claramente se destacado na fronteira do conhecimento em demografia. Ao avaliar os méritos e limitações de cada abordagem, conclui-se que o paradigma bayesiano irá se destacar no futuro próximo como a abordagem mais promissora para as projeções de população, uma vez que combina a opinião de especialistas, as informações que os demógrafos têm disponíveis a partir de suas análises empíricas, assim como métodos estatísticos e computacionais sofisticados para lidar

com a incerteza. Assim, a disponibilidade de previsões demográficas acuradas que levem em conta a incerteza pode melhorar a comunicação entre os demógrafos, permitindo uma maior flexibilidade na determinação das previsões demográficas.

**Palavras-chave:** Projeções de população. Incerteza. Modelo de coorte componente. Abordagem frequentista. Abordagem bayesiana.

## Resumen

### *La incertidumbre en las proyecciones de población: el estado del arte*

El artículo presenta en primer lugar una revisión crítica del estado del arte sobre las proyecciones de población, centrándose en la forma en que se aborda la incertidumbre en tres enfoques diferentes: el modelo clásico de cohorte-componente, el probabilístico frequentista y el paradigma bayesiano. A continuación, el análisis se centra en la evolución reciente de los modelos de proyecciones bayesianos de la fecundidad, la mortalidad y la migración, que claramente se ubican en la frontera del conocimiento en demografía. A partir de la evaluación de las ventajas y limitaciones de cada enfoque, se concluye que el paradigma bayesiano se destacará en el futuro cercano como el abordaje más prometedor para las proyecciones de población, ya que combina la opinión de los expertos, la información de la que disponen los demógrafos a través de sus análisis empíricos y métodos estadísticos y computacionales sofisticados para lidiar con la incertidumbre. De este modo, la disponibilidad de pronósticos demográficos precisos que tengan en cuenta la incertidumbre puede mejorar la comunicación entre los demógrafos, permitiendo una mayor flexibilidad para elaborar las proyecciones demográficas.

**Palabras clave:** Proyecciones de población. Incertidumbre. Modelo de cohorte-componente. Enfoque frequentista. Enfoque bayesiano.

Recebido para publicação em 13/01/2013

Aceito para publicação em 19/11/2014