ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Point and interval forecasts of age-specific fertility rates: a Comparison of functional principal component methods

Han Lin Shang

April 2012

Working Paper 10/12

Point and interval forecasts of age-specific fertility rates: a comparison of functional principal component methods

Han Lin Shang

Department of Econometrics & Business Statistics, Monash University, Melbourne, Vic 3145, Australia

Abstract

Accurate forecasts of age-specific fertility rates are critical for government policy, planning and decision making. With the availability of Human Fertility Database (2011), we compare the empirical accuracy of the point and interval forecasts, obtained by the approach of Hyndman and Ullah (2007) and its variants for forecasting age-specific fertility rates. The analyses are carried out using the age-specific fertility data of 15 mostly developed countries. Based on the one-step-ahead to 20-step-ahead forecast error measures, the weighted Hyndman-Ullah method provides the most accurate point and interval forecasts for forecasting age-specific fertility rates, among all the methods we investigated.

Keywords: functional data analysis, functional principal component analysis, forecast accuracy comparison, random walk with drift, random walk, ARIMA model

JEL classification code: J11, J13, C14

Email address: `HanLin.Shang@monash.edu` (Han Lin Shang)

URL: `http://monashforecasting.com/index.php?title=User:Han` (Han Lin Shang)

1. Introduction

Accurate forecasts of age-specific fertility rates are critical for allocation of current and future resources to fertility control programmes and child care services. There has been a surge of interest in forecasting fertility rate in recent years, mainly driven by the need for accurate forecasts to inform government policy and planning. Such an interest is reflected by many contributions in the 1$^{\text{st}}$ Human Fertility Database Symposium in 2011. Some of the contributions include the work by Bongaarts and Sobotka (2011) and Goldstein and Cassidy (2010), to name a few.

This paper is motivated by the lack of forecasting method comparison for age-specific fertility rates. When a demographer is asked to forecast age-specific fertility rates, an obvious question is that which method gives the best accuracy of the point and interval forecasts. Thanks to the availability of Human Fertility Database (2011), this paper aims to contribute to the demographic literature by providing a forecast accuracy comparison of some functional principal component methods. We reveal the methods that produce the most accurate point and interval forecasts for the age-specific fertility rates.

In broad terms, our analyses are extensions of the early work by Booth et al. (2006) and Shang et al. (2011). However, the comparison presented here differs from those two previous work by the following three factors:

1) We use a rolling origin approach to examine one-step-ahead to 20-step-ahead forecast errors. By contrast, Booth et al. (2006) utilized a fixed origin approach and examined the average of one-step-ahead to fifteen-step-ahead forecasts. Shang et al. (2011) utilized a rolling origin approach and examined only the one-step-ahead and ten-step-ahead forecasts.
2) We compare the forecast accuracy of age-specific fertility rates for the ages from 15 to 49, whereas Booth et al. (2006) and Shang et al. (2011) compared the forecast accuracy of age-specific log mortality rates for the ages from 0-94 and 0-89, respectively.
3) In the fertility rate forecasting, the nonparametric smoothing technique is the penalized regression spline with concavity constraint (see Hyndman and Ullah, 2007, for detail). In the mortality rate forecasting, the nonparametric smoothing technique is the penalized regression spline with monotonic constraint (see Ramsay, 1988, for detail).

When modeling and forecasting age-specific fertility rates, we often encounter the problem of large dimensionality, as the number of age variable increases. In the demographic literature, there exist at least a parametric and a nonparametric solution to overcome this problem, namely fitting Gamma curves or using principal component analysis. In the parametric school of thought, Rogers (1986), Thompson, Bell, Long and Miller (1989), Bell (1992) and Keilman and Pham (2000) suggested fitting curves to annual age-specific fertility rates. They proposed to approximate the age-specific fertility rates by Gamma curves; reduce the dimension (number of ages) to the number of Gamma curve parameters; forecast the Gamma curve parameters using time series techniques to obtain the forecasted Gamma curves; then using the forecasted Gamma curves to produce the forecasts of age-specific fertility rates (with fixed Gamma density). In the nonparametric school of thought, Bozik and Bell (1987) and Bell (1988, 1992, 1997) utilized the principal component analysis to produce the point and interval forecasts of age-specific fertility rates with only a small number of orthogonal principal components. In the sense of mean squared error, principal component approximation minimizes the loss of information, while reducing dimensionality. In this paper, we consider the nonparametric means of forecasting age-specific fertility rates. Our reasonings are: since we do not know the underlying data generating process in practice, the parametric assumption of Gamma density may not be satisfied; secondly, the nonparametric method is a data-driven approach, and follows the paradigm of "let data speak for themselves".

As emphasized by Bell (1997), a characteristic of age-specific fertility rates is the smooth shape over age of the data each year, it seems desirable to use modeling and forecasting methods that capture this smooth shape over age, in hope to improve the accuracy of short-term forecasts. This motivated the work by Hyndman and Ullah (2007), which has received increasing attention in demographic forecasting (see Peristeva and Kostaki, 2005; Booth, Hyndman, Tickle and De Jong, 2006; Rueda-Sabater and Alvarez-Esteban, 2008; Iqbal and Hoque, 2008; Hyndman and Shang, 2009; Lazar and Denuit, 2009; Dowd, Cairns, Blake, Coughlan, Epstein and Khalaf-Allah, 2010; Yang, Yue and Huang, 2010; Shang, Booth and Hyndman, 2011, for example). As a generalization of Lee and Carter's(1992) method, this functional data analysis approach views age-specific fertility rates from a functional perspective. Instead of modeling and forecasting different ages in a given year, we consider age as a smooth and continuous variable that can be intrinsically of infinite dimension. Borrowing the language of functional analysis (Conway, 1994), each year of age-specific fertility rate can be considered as a realization of a stochastic process f in a function space. This approach combines ideas from functional data analysis (Ramsay and Silverman, 2005) and nonparametric smoothing (Eubank, 1999), and it has the following features:

1. Data are pre-smoothed before fitting functional principal component analysis, in order to smooth out missing observations.
2. Fertility rates are modeled as continuous functions of age, so that patterns of variation among years are captured by functional principal components.
3. This method can be robust against possible outlier.
4. This method produces point and interval forecasts for a whole function, which is a future realization of the stochastic process.

The rest of this paper is given as follows. In Sections 3 and 5, we revisit the Hyndman-Ullah (HU) method, robust Hyndman-Ullah (HUrob) method, geometrically decaying weighted Hyndman-Ullah (HUw) method, for calculating the point and interval forecasts. Illustrated by the 15 countries' data sets described in Section 2, we compare the accuracy of the point and interval forecasts in Sections 4 and 6, respectively. A discussion is given in Section 7, along with some thoughts on how the methods and comparison presented here might be further extended.

2. Data sets

The data sets used in this study were taken from the Human Fertility Database (2011). The fertility rates are defined as the number of live births during the calendar year, according to the age of the mother among the female resident population of the same age at 30th June.

Fifteen mostly developed countries were selected in our analyses, solely based on the length of data series. These fifteen countries selected all have reliable data series commencing at/before 1951 and end in 2006. The forecasting period is set to be from 1987 to 2006, while the remaining data are the training set. Note that it was desirable to have the number of available data beyond 50 years, in order to obtain consistent sample estimators (Box et al., 2008, p.17). In general, the more data we have (the longer the time series), the more likely it is that we will select and estimate a model that approximate the underlying structure of the data well. The selected countries are shown in Table 1. Age is given in single years and we restrict the age range from 15 to 49, in order to avoid excessive fluctuation at younger and older ages.

[Insert Table 1 about here.]

As pointed out by Bell (1992), how many of the available data should be used in modeling? There may be known events thought to affect the data, (such as, Spanish flu pandemic in 1918-1919) or reasons to suspect that the structure of the data may have changed over time (before and after the second World War). There are two ways of dealing with this issue: drop the affected data (such as the method of HUrob) or attempt to model the effects (such as the method of HU).

Apart from the modeling consideration, the relevant of the data from distant past may not be quite useful for forecasting. Forecasts from time-series models depend most heavily on recent data with diminishing weight given to those data from distant past. This motivates the development of the HUw method.

3. Review of point forecast methods

In this section, we present a unified framework for the three HU methods. The accuracy of point and interval forecasts is compared with some univariate methods, namely the random walk (RW), random walk with drift (RWD), and autoregressive integrated moving average (ARIMA) model.

3.1. Hyndman-Ullah (HU) method and its variants

The features of the HU methods can be summarized as follows:

1) The age-specific fertility rates are first smoothed using penalized regression spline with a concavity constraint (see He and Ng, 1999, for detail). Assuming there is an underlying continuous and smooth function $\{f_t(x); x \in [x_1, x_p]\}$ that is observed with errors at

discrete ages in year t, we can express it as

$$m_t(x_i) = f_t(x_i) + \sigma_t(x_i)\epsilon_{t,i}, \quad i = 1, 2, \dots, p, \quad t = 1, 2, \dots, n, \tag{1}$$

where $m_t(x_i)$ denotes the observed fertility rates for age x_i in year t; $\sigma_t(x_i)$ allows the amount of noise to vary with x_i in year t, thus allowing the possible presence of heteroscedastic error; and $\epsilon_{t,i}$ is an independent and identically distributed (iid) standard random variable. The nonparametric smoothing method is extremely useful for transforming data from five-year age groups into single-year. This also facilities a range of models, where annual data by single age are more convenient to work with (see also Liu, Gerland, Spoorenberg, Vladimira and Andreev, 2011).

2) With a set of smooth functions $\{f_1(x), f_2(x), \dots, f_n(x)\}$, the mean function $a(x)$ can be estimated by

$$\hat{a}(x) = \sum_{t=1}^{n} w_t f_t(x). \tag{2}$$

$w_t = \frac{1}{n}$ represents equal weighting used in the HU method, whereas $\{w_t = \lambda(1-\lambda)^{n-t}, t = 1, 2, \dots, n\}$ represents a set of geometrically decaying weights used in the HUw method. The geometrically decaying weight parameter λ must be estimated in a data-driven manner (see Section 3.4 for more details).

3) Using functional principal component analysis (FPCA), a set of curves $\{w_t[f_t(x) - \hat{a}(x)]; t = 1, 2, \dots, n\}$ is decomposed into orthogonal functional principal components and their uncorrelated scores. That is,

$$f_t(x) = \hat{a}(x) + \sum_{j=1}^{J} \hat{b}_j(x)\hat{k}_{t,j} + e_t(x), \tag{3}$$

where $\{\hat{b}_1(x), \hat{b}_2(x), \dots, \hat{b}_J(x)\}$ represents a set of equally weighted or geometrically decaying weighted functional principal components; $\{\hat{k}_{t,1}, \hat{k}_{t,2}, \dots, \hat{k}_{t,J}\}$ is a set of uncorrelated principal component scores; $e_t(x)$ is the residual function with mean zero; $J < n$ is the number of retained principal components. Following Hyndman and Booth (2008) and Shang et al. (2011), we chose $J = 6$ which should be larger than any of the components required.

Denote $v_t = \int_x e_t^2(x)dx$ as the integrated squared error for year t. If $v_t \geq s + \theta\sqrt{s}$ and s is the median of $\{v_1, v_2, \dots, v_n\}$ and θ is a tuning parameter with the default value of 3, then year t is considered as an outlier and should be removed from analyses. By assigning zero weights to outliers before applying FPCA, this forms the idea of the HUrob method.

4) By conditioning on the observed data $\boldsymbol{\mathcal{I}} = \{m_1(x), m_2(x), \dots, m_n(x)\}$ and the set of functional principal components $\boldsymbol{B} = \{\hat{b}_1(x), \hat{b}_2(x), \dots, \hat{b}_J(x)\}$, the h-step-ahead point forecast of $m_{n+h}(x)$ can be expressed as

$$\hat{m}_{n+h|n} = \text{E}[m_{n+h}(x)|\boldsymbol{\mathcal{I}}, \boldsymbol{B}] = \hat{a}(x) + \sum_{j=1}^{J} \hat{b}_j(x)\hat{k}_{n+h|n,j}, \tag{4}$$

where $\hat{k}_{n+h|n,j}$ denotes the h-step-ahead forecast of $k_{n+h,j}$ using a univariate time series, such as an exponential smoothing state space model of Hyndman et al. (2008).

3.2. Random walk (RW) methods

It is pertinent to compare the point forecast accuracy of the HU methods with naïve linear extrapolation of age-specific fertility rates, such as the RW and RWD methods (see Alho and Spencer, 2005, pp.274-276 for an introduction). The linear extrapolation of age-specific fertility rates was achieved by applying the RW and RWD models to each age,

$$m_{x,t+1} = c + m_{x,t} + e_{x,t+1}, \quad t = 1, 2, \dots, n-1, \quad x = x_1, x_2, \dots, x_p,$$

where $m_{x,t}$ represents the age-specific fertility rate at age x in year t, c represents the drift term, and $e_{x,t+1}$ represents the normal iid error with mean zero.

The h-step-ahead point and interval forecasts are given by

$$\hat{m}_{x,n+h|n} = \text{E}[m_{x,n+h}|m_{x,1}, m_{x,2}, \dots, m_{x,n}] = ch + m_{x,n},$$
$$\text{var}(\hat{m}_{x,n+h|n}) = \text{var}[m_{x,n+h}|m_{x,1}, m_{x,2}, \dots, m_{x,n}] = \text{var}(m_{x,n}) + \text{var}(e_{x,n+h}),$$

where h represents the forecast horizon. The RW model is the same as the RWD model with

$c = 0$. Assuming the various sources of error are all normally distributed, a $100(1-\alpha)\%$ prediction interval for $m_{x,n+h}$ is constructed as

$$\hat{m}_{x,n+h|n} \pm z_{\alpha}\sqrt{\text{var}(\hat{m}_{x,n+h|n})}, \qquad (5)$$

where z_{α} is the $(1-\alpha/2)$ standard normal quantile. Computationally, the point and intervals forecasts of the RW and RWD can be obtained via the `rwd` function in the *forecast* package (Hyndman, 2011*b*) in R (R Development Core Team, 2011).

3.3. Autoregressive integrated moving average (ARIMA) model

One popular approach of forecasting age-specific fertility rates involves the use of time-series models. In this approach, we select a particular time-series model for the series to be forecasted, and use the fitted model to produce point and interval forecasts. The ARIMA model discussed by Box et al. (2008) comprises one popular class of models (see also, Abraham and Ledolter, 1983). The ARIMA model has been applied to forecast fertility rate and its related problems by Lee (1974, 1975), Saboia (1977), McDonald (1979, 1981), Miller and Hickman (1981), Land and Cantor (1983), Miller and McKenzie (1984), de Beer (1985), Carter and Lee (1986), Bozik and Bell (1987, 1989) and Bell (1992). As a demonstration, we consider a stationary time series. If the series is not stationary, one should take differences in order to achieve stationarity. The optimal model is assumed to be ARMA(p,q), where p and q are the orders associated with the AR and MA components. The linear extrapolation of age-specific fertility rates can also be achieved by applying the ARMA(p,q) model to each age,

$$(m_{x,t} - \mu) = \sum_{i=1}^{p}\beta_i(m_{x,t-i} - \mu) + w_{x,t} + \sum_{j=1}^{q}\psi_j w_{x,t-j}, \quad t = \max(p,q)+1,\dots,n, \qquad (6)$$

where μ represents the time-varying mean of a time series, $(\beta_1, \beta_2, \dots, \beta_p)$ represent the coefficients of the autoregressive (AR) component, $(\psi_1, \psi_2, \dots, \psi_q)$ represent the coefficients of the moving average (MA) component, $m_{x,t}$ represents the age-specific fertility rate at age x in year t, the error term $\{w_{x,t}, w_{x,t-1}, \dots, w_{x,t-q}\}$ has mean zero and follows a normal distribution with the estimated variance $\hat{\sigma}^2_w$.

The one-step-ahead point and interval forecasts are given by

$$\hat{m}_{x,n+1|n} = \mathrm{E}[m_{x,n+1}|m_{x,1}, m_{x,2}, \dots, m_{x,n}] = \hat{\mu} + \sum_{i=1}^{p} \hat{\beta}_i(m_{x,n+1-i} - \hat{\mu}),$$
$$\mathrm{var}[m_{x,n+1}|m_{x,1}, m_{x,2}, \dots, m_{x,n}] = \hat{\sigma}_w^2 \left(1 + \hat{\psi}_1^2 + \hat{\psi}_2^2 + \dots + \hat{\psi}_q^2\right). \quad (7)$$

The parameter values are commonly estimated by the maximum likelihood criterion. A $100(1-\alpha)\%$ prediction interval for the $y_{n+h}(x)$ can be constructed similarly as (5). For h-step-ahead forecasts, (6) and (7) can be applied iteratively.

It is noteworthy that Hyndman and Khandakar (2008) developed an algorithm, named `auto.arima` in the *forecast* package in R, for automatically selecting the optimal ARIMA model based on a statistical criterion (such as Akaike's (1974) Information Criterion).

3.4. Evaluating point forecast accuracy

We divide each data set into a fitting period and a forecasting period. We applied a rolling origin approach as follows: The forecasting period is initially set to be the last 20 years ending in 2006. Using the data in the fitting period, we compute the one-step-ahead to 20-step-ahead forecasts, and determine the forecast errors by comparing the forecasts with the actual out-of-sample data. Then, we increase the fitting period by one year, and compute the one-step-ahead to 19-step-ahead forecasts, and calculate the forecast errors. This process is repeated until the fitting period extends to 2005.

In the HUw method, we also divide the data in the fitting period into a training set and a validation set. The validation set is set to be the last 20 years ending in 1986. Using the data in the training period, we compute the one-step-ahead to 20-step-ahead forecasts, and determine the forecast errors. Then, we increase the training period by one year, and compute the one-step-ahead to 19-step-ahead forecasts, and calculate the forecast error. This process is repeated until the fitting period extends to 1985. The optimal value of $\lambda \in (0,1)$ is chosen by minimizing an overall forecast error measure. We utilize a one-dimensional optimization algorithm of Nelder and Mead (1965), to minimize the objective function and to find its corresponding optimal value of λ. Other minimization algorithms are also possible in the `optim` function in R.

There are many ways to measure the point forecast accuracy, such as mean absolute forecast error (MAFE), mean absolute percentage forecast error, and root mean square forecast error. There are also many ways to measure the bias, such as the mean forecast error (MFE) and mean algebraic percentage forecast error. Following the early work by Booth et al. (2006) and Shang et al. (2011), we use the MAFE and MFE in this paper. They are defined by

$$\text{MAFE}_h = \frac{1}{15 \times (21-h) \times 35} \sum_{i=1}^{15} \sum_{k=1}^{21-h} \sum_{j=1}^{35} \left| f_k^{(i)}(x_j) - \hat{f}_k^{(i)}(x_j) \right|,$$

$$\text{MFE}_h = \frac{1}{15 \times (21-h) \times 35} \sum_{i=1}^{15} \sum_{k=1}^{21-h} \sum_{j=1}^{35} \left[f_k^{(i)}(x_j) - \hat{f}_k^{(i)}(x_j) \right],$$

where $i = 1, 2, \dots, 15$ represents the country index, j represents 35 different ages from 15 to 49, and $h = 1, 2, \dots, 20$ represents the forecast horizon. The MAFE is the average of absolute errors, |actual – forecast|, and measures forecast precision, regardless of sign. The MFE is the average of error, (actual – forecast), and is a measure of bias. These measures are used to evaluate the point forecast accuracy of the age-specific fertility rates.

4. Comparison of point forecast accuracy

Following the work by Shang et al. (2011), results are presented by country and for two averages: the simple average and a weighted average using weights based on birth count in the most recent year, that is 2006 in our data. For each country, the weight is calculated as the country's birth count in 2006 divided by the sum of each country's birth count in 2006 and scaled to sum to 15. The weights are 2.96, 0.13, 0.58, 0.17, 1.09, 0.10, 1.30, 0.16, 0.30, 0.09, 0.09, 0.79, 0.17, 0.12, and 6.95 in the country order of Table 1.

Table 2 provides summaries of the point forecast accuracy based on the MAFEs for the one-step-ahead to 20-step-ahead forecasts of fertility rates, averaged over the ages, years in the forecasting period and countries. As measured by the simple and weighted averages of the MAFEs over countries, the HU methods tend to perform slightly better than the naïve univariate methods on average. The HUw method performs the best in all the methods investigated for every forecast horizon.

[Insert Table 2 about here.]

Table 3 shows the corresponding MFEs for the one-step-ahead to 20-step-ahead forecasts. With the exception of the RWD method, other methods tend to overestimate fertility rates, that is, forecasts are greater than the holdout samples. As the forecast horizon increases, the bias increases from $h = 1$ to $h = 15$, then the bias reduces and approaches to zero from $h = 16$ to $h = 20$. The reduction in bias from $h = 16$ to $h = 20$ may due to the cancelation of positive and negative bias and less number of observations at longer horizons. The RWD method tends to underestimate fertility rates giving smaller forecasts in comparison to the holdout samples. As the forecast horizon increases, the bias increases from $h = 1$ to $h = 20$.

[Insert Table 3 about here.]

Figure 1a shows the MAFEs of the one-step-ahead to 20-step-ahead forecasts for different methods and forecast horizons, averaged over the ages, years in the forecasting period, and 15 countries. As the forecast horizon increases, the values of MAFEs increase linearly (see also Smith and Sincich, 1991). Among all the methods investigated, the HUw method provides uniformly smaller MAFEs for all the forecast horizons.

Figure 1b displays the corresponding MFEs. With the exceptions of the RWD, other methods generally overestimate the fertility rates. In the case of the RWD, the bias intensifies as the forecast horizon increases. For all other methods, the bias increases from $h = 1$ to $h = 15$. Because there are a small number of observations at longer horizons, this may reduce the bias from $h = 16$ to $h = 20$.

[Insert Figure 1 about here.]

As suggested by a referee, we also include the point forecast comparison of the total fertility rates. The total fertility rate is the sum of the age-specific fertility rates for a given year. With the last 20 years as the test samples, the forecast accuracy and bias are compared using the forecasted age-specific fertility rates. Figures 2a and 2b show the MAFEs and MFEs of the one-step-ahead to 20-step-ahead forecasts for different methods and forecast horizons, averaged over the ages, years in the forecasting period, and 15 countries. As the forecast horizon increases, the values of MAFEs increases quadratically. The RW method performs the best in all the methods investigated. With the exception of the RWD, other methods generally overestimate the fertility rates. The MFEs of the total fertility rates are similar in shape to the MFEs of the age-specific fertility rates, because the latter one is averaged over age.

[Insert Figure 2 about here.]

5. Evaluating interval forecast accuracy

In recent years, some statistical agencies have attempted to compute stochastic population forecasts with a prediction interval. See for example Lutz and Scherbov (1998) for Austria, Alho (1998) for Finland, Keilman and Pham (2000) for Norway. These methods have been motivated by earlier work on stochastic forecasts by for instance Alho (1990), Lee and Tuljapurkar (1994) and Tayman et al. (2007). The current paper is a contribution to the literature in this area. Its aim is to compare the accuracy of interval forecasts among the functional principal component methods and univariate methods for forecasting age-specific fertility rates.

To avoid the repetitiveness, we refer the readers to the work by Shang et al. (2011) for the constructions of prediction interval for the HU methods. The prediction intervals of the RW and ARIMA models are constructed parametrically with the variances described in (5) and (7).

The method of evaluating interval forecast accuracy for age-specific fertility rates is given as follows. For each year in the forecasting period, the one-step-ahead to 20-step-ahead prediction intervals were calculated at the 80% and 95% nominal coverage probabilities, and were then tested against the actual proportion of out-of-sample data that fell within the calculated prediction intervals (Swanson and Beck, 1994; Tayman, Smith and Lin, 2007). The empirical coverage probabilities are defined as the proportion of observations that fall into the calculated 80% and 95% prediction intervals, where the denominator is the total number of observations in the forecasting period (for example, 35 ages $\times$ 20 years = 700 observations). We calculated the coverage probability deviance, which is the absolute difference between a nominal coverage probability and the empirical coverage probability, and use this measure to evaluate the interval forecast accuracy of each method. For instance, with a nominal coverage probability of 80%, the maximum coverage probability deviance is 80% (i.e., when the empirical coverage probability is 0), while the minimum coverage probability deviance is 0 (i.e., when the empirical coverage probability is 80%). Results are given in terms of the coverage probability deviance.

Apart from the coverage probability deviance as a means of finding the optimal method, we also calculate the size of prediction interval. Following the early work by Lee and Tuljapurkar (1994) and Tayman et al. (2007), the size of the prediction interval is defined as a *half-width* by dividing one-half of the difference between the upper and lower bounds of the interval forecasts, and multiplying the result by 100. For symmetrical intervals, the half-width reflects the percentage distance between the point forecast (center) and the lower and upper bounds of the prediction interval.

6. Comparison of interval forecast accuracy

The simple and weighted average coverage probability deviances of the age-specific fertility rates are shown in Tables 4 and 5, with the nominal coverage probabilities of 80% and 95%. Based on both averages, the HUw method performs the best in all the methods investigated. The results show that all HU methods tend to deviate from the nominal coverage probability, and such a deviation increases as the forecast horizon increases. This reflects that the accuracy of the interval forecasts in capturing the extent of uncertainty diminishes with the increases in forecast horizon. As measured by the weighted averages, the coverage probability deviances of the univariate methods are larger than the HU methods. Among the univariate methods, the RW method produces a smaller coverage probability deviance.

With the exception of the HUw method, other methods tend to underestimate the model uncertainty, resulting in a smaller empirical coverage probability than the nominal coverage probability. The extent of underestimation of the model uncertainty intensifies, as the forecast horizon increases. On the contrary, the HUw method overestimates the model uncertainty, especially for the shorter forecast horizons.

Apart from the coverage probability deviance, we also calculate the half-width of the prediction interval, shown in Table 6. The HUw and ARIMA methods show larger half-width than the others. The better coverage probability of the HUw method may due to a larger half-width of the prediction interval. The larger half-width of the ARIMA model is rather not surprising. As the forecast horizon increases, the selected ARIMA model has less and less predictiveness about the future, resulting in a much larger half-width. On the contrary, the RW and RWD assume that the future forecasts do not differ significantly from the present values, thus restricting the half-width of the prediction interval to some extent.

7. Discussion

The above comparative analysis of forecasting methods for fertility rates is the most comprehensive to date. We shall summarize the main points:

1) Model uncertainty should be considered as a criterion for finding the best model.
2) The RW model performs better than the RWD model for forecasting age-specific fertility rates.
3) The HUw model performs the best in all the methods investigated for producing the point and interval forecasts of the age-specific fertility rates.
4) The RW model performs the best in all the methods investigated for producing the point forecasts of the total fertility rates.

7.1. Point forecasts

Our overall findings regarding point forecasts of fertility rates are that the HUw method is more accurate than any other method investigated for producing the one-step-ahead to 20-step-ahead forecasts (see Tables 2, 4, 5). The success of the HUw method suggests that assigning greater weights to the recent years lead to generally smaller age-specific errors, given that such errors are cumulated over forecast horizon.

In contrast, the univariate methods perform worse than the HU methods, in terms of point forecast accuracy. This is mainly due to two factors. First, the smoothing of fertility rates means that the observational error is treated separately from dynamic changes over years. Second, the HU methods forecast the changes in fertility rates through different sets of principal components and principal component scores, whereas the univariate methods extrapolate data without modeling any possible non-linear pattern.

7.2. *Interval forecasts*

The coverage probability deviances of the HU methods are better than the univariate methods, as measured by the weighted averages. This may because the RW and RWD methods model only the parameter and model uncertainty, whereas the HU methods model one additional source of uncertainty (see also Alho and Spencer, 2005, p255). That additional source of uncertainty stems from the use of a nonparametric smoothing technique, namely penalized regression spline with the concavity constraint, in order to smooth the noisy fertility rates. As the forecast horizon increases, the coverage probability deviances of the HU methods become larger, as reflected by diminishing accuracy of the interval forecasts in capturing the extent of uncertainty associated with the increases in forecast horizon.

It should be noted that the coverage probability deviance does not distinguish between the upper and lower bounds. Concretely, the coverage probability deviance would be the same, when the empirical coverage probability is 0.9 or 0.7 at the nominal coverage probability of 0.8. Further research is needed to develop other accuracy measures of interval forecasts.

7.3. *Limitation*

With the focus on the principal component approaches, this comparative analysis is concluded on the basis of the weighted averages across the 15 countries, with the forecasting period from 1987 to 2006. A different result may be obtained for a different forecasting period, as reflected by the gradual pattern change in the age-specific fertility rates.

To examine if the differences among methods are statistically significant, a hypothesis testing is often performed. However, statistical tests of the results are problematic, because the countries and years included cannot be treated as random independent samples of a well-defined population. Therefore, we did not test the statistical significance among the methods investigated.

7.4. *Future research*

A future research is to consider the idea of model averaging to improve point and interval forecast accuracy. The idea of model averaging is not new, and has been studied in statistics since the work by Bates and Granger (1969). In the demographic literature, there has been little interest among demographers in combining forecasts from different methods. Nonetheless, some notable exceptions include Smith and Shahodullah (1993), Ahlburg (1998, 2001) and Sanderson (1998), whose pioneering work, particularly in the context of census tract forecast, have done much to awaken others, including the present author. However, a difficulty in using the model averaging approach is the question of estimating the optimal weights associated with different methods in a data-driven manner. Although Bayesian modeling averaging provides a promising solution, such an investigation awaits in the field of demographic forecasting.

7.5. Implementation

Implementation of the methods used in this article is straightforward using the readily-available R package *demography* (Hyndman, 2011*a*). The data requirements are historical fertility rates and birth counts in a complete matrix format by age and year. Such data are readily available for many European countries from the Human Fertility Database (2011).

Acknowledgements

The author would like to thank the editor-in-chief, the associate editor, and two referees for their insightful comments, which led to a much improved version of the manuscript. The author thanks Professors Heather Booth and Rob Hyndman for introducing him to the field of demographic forecasting.

References

Abraham, B. and Ledolter, J. (1983), *Statistical Methods for Forecasting*, John Wiley and Sons, New York.

Ahlburg, D. A. (1998), Using economic information and combining to improve forecast accuracy in demography, Working paper, Industrial Relations Center, University of Minnesota, Minneapolis.

Ahlburg, D. A. (2001), Population forecasting, *in* J. S. Armstrong, ed., 'Principles of Forecasting', Kluwer Academic Publishers, New York, pp. 557–575.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Alho, J. M. (1990), 'Stochastic methods in population forecasting', *International Journal of Forecasting* **6**(4), 521–530.

Alho, J. M. (1998), A stochastic forecast of the population of Finland, Reviews 1998/4, Statistics Finland, Helsinki.

Alho, J. M. and Spencer, B. D. (2005), *Statistical Demography and Forecasting*, Springer, New York.

Bates, J. M. and Granger, C. W. J. (1969), 'The combination of forecasts', *Operations Research Society* **20**(4), 451–468.

Bell, W. (1988), Applying time series models in forecasting age-specific fertility rates, Working paper, Bureau of the Census.

Bell, W. (1992), ARIMA and principal components models in forecasting age-specific fertility, *in* N. Keilman and H. Cruijsen, eds, 'National Population Forecasting in Industrialized Countries', Swets & Zeitlinger, Amsterdam, pp. 177–200.

Bell, W. (1997), 'Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates', *Journal of Official Statistics* **13**(3), 279–303.

Bongaarts, J. and Sobotka, T. (2011), Demographic explanation for the recent rise in European fertility: Analysis based on the tempo and parity-adjusted total fertility, Working paper, Population Council.

Booth, H., Hyndman, R. J., Tickle, L. and De Jong, P. (2006), 'Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions', *Demographic Research* **15**, 289–310.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008), *Time Series Analysis: Forecasting and Control*, 4th edn, John Wiley, Hoboken, New Jersey.

Bozik, J. and Bell, W. (1987), Forecasting age specific fertility using principal components, *in* 'Proceedings of the American Statistical Association. Social Statistics Section', San Francisco, CA, pp. 396–401.

Bozik, J. E. and Bell, W. R. (1989), Time series modelling for the principal components approach to forecasting age specific fertility, SRD Research Report, Statistical Research Division, Bureau of the Census, Washington, D.C.

Carter, L. R. and Lee, R. D. (1986), 'Joint forecasts of U.S. marital fertility, nuptiality, births and marriages using time series models', *Journal of the American Statistical Association* **81**(396), 902–911.

Conway, J. B. (1994), *A Course in Functional Analysis*, 2nd edn, Springer-Verlag, New York.

de Beer, J. (1985), 'A time series model for cohort data', *Journal of the American Statistical Association* **80**(391), 525–530.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2010), 'Evaluating the goodness of fit of stochastic mortality model', *Insurance Mathematics and Economics* **47**(3), 255–265.

Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing*, 2nd edn, Marcel Dekker, New York.

Goldstein, J. R. and Cassidy, T. (2010), Cohort postponement and period measures, Working paper number 15,, Max Planck Institute for Demographic Research.

He, X. and Ng, P. (1999), 'COBS: Qualitatively constrained smoothing via linear programming', *Computational Statistics* **14**(3), 315–337.

Human Fertility Database (2011), *Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria)*. Available at www.humanfertility.org (data downloaded on 16/11/2011).

Hyndman, R. J. (2011*a*), *demography: Forecasting mortality, fertility, migration and population data*. R package version 1.11. http://CRAN.R-project.org/package=demography.

Hyndman, R. J. (2011*b*), *forecast: Forecasting functions for time series*. R package version 3.18. http://CRAN.R-project.org/package=forecast.

Hyndman, R. J. and Booth, H. (2008), 'Stochastic population forecasts using functional data models for mortality, fertility and migration', *International Journal of Forecasting* **24**(3), 323–342.

Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**(3).

Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2008), *Forecasting with Exponential Smoothing: the State Space Approach*, Springer, Berlin.

Hyndman, R. J. and Shang, H. L. (2009), 'Forecasting functional time series (with discussion)', *Journal of the Korean Statistical Society* **38**(3), 199–221.

Hyndman, R. J. and Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Iqbal, M. A. and Hoque, M. A. (2008), 'Robust forecasting of fertility trends in Bangladesh', *Middle East Journal of Age and Ageing* **5**(1), 13–22.

Keilman, N. and Pham, D. Q. (2000), 'Predictive intervals for age-specific fertility', *European Journal of Population* **16**(1), 41–66.

Land, K. C. and Cantor, D. (1983), 'ARIMA models of seasonal variation in U.S. birth and death rates', *Demography* **20**(4), 541–568.

Lazar, D. and Denuit, M. M. (2009), 'A multivariate time series approach to projected life tables', *Applied Stochastic Models in Business and Industry* **25**(6), 806–823.

Lee, R. D. (1974), 'Forecasting births in post-transitional populations: Stochastic renewal with serial correlated fertility', *Journal of the American Statistical Association* **69**(347), 607–617.

Lee, R. D. (1975), 'Natural fertility, population cycles and the spectral analysis of births and marriages', *Journal of the American Statistical Association* **70**(350), 295–304.

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

Lee, R. and Tuljapurkar, S. (1994), 'Stochastic population forecasts for the United States: Beyond high, medium and low', *Journal of the American Statistical Association* **89**(428), 1175–1189.

Liu, Y., Gerland, P., Spoorenberg, T., Vladimira, K. and Andreev, K. (2011), Graduation methods to derive age-specific fertility rates from abridged data: A comparison of 10 methods using HFD data, Working paper, Columbia University.

Lutz, W. and Scherbov, S. (1998), 'An expert-based framework for probabilistic national population projections: The example of Austria', *European Journal of Population* **14**(1), 1–17.

McDonald, J. (1979), 'A time series approach to forecasting Australian total live births', *Demography* **16**(4), 575–602.

McDonald, J. (1981), 'Modeling demographic relationships: An analysis of forecast functions for Australian births', *Journal of the American Statistical Association* **76**(376), 782–792.

Miller, R. B. and Hickman, J. C. (1981), Time series modelling of births and birth rates, Working paper 8-81-21, Graduate School of Business, University of Wisconsin-Madison.

Miller, R. B. and McKenzie, S. K. (1984), Time series modelling of monthly general fertility rates, SRD Research Report 84/16, Statistical Research Division, Bureau of the Census,, Washington, D.C.

Nelder, J. A. and Mead, R. (1965), 'A simplex method for function minimization', *The Computer Journal* **7**(4), 308–313.

Peristeva, P. and Kostaki, A. (2005), 'An evaluation of the performance of kernel estimator for graduating mortality data', *Journal of Population Research* **22**(2), 185–197.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Ramsay, J. O. (1988), 'Monotone regression splines in action', *Statistical Science* **3**(4), 425–441.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, 2nd edn, Springer, New York.

Rogers, A. (1986), 'Parameterized multistate population dynamics and projections', *Journal of the American Statistical Association* **81**(393), 48–61.

Rueda-Sabater, C. and Alvarez-Esteban, P. C. (2008), 'The analysis of age-specific fertility patterns via logistic models', *Journal of Applied Statistics* **35**(9), 1053–1070.

Saboia, J. L. M. (1977), 'Autoregressive integrated moving average (ARIMA) models for birth forecasting', *Journal of the American Statistical Association* **72**(358), 264–270.

Sanderson, W. C. (1998), 'Knowledge can improve forecasts! a review of selected socio-economic population projection models', *Population and Development Review* **24**(supplement), 88–117.

Shang, H. L., Booth, H. and Hyndman, R. J. (2011), 'Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods', *Demographic Research* **25**(5), 173–214.

Smith, S. K. and Shahodullah, M. (1993), Evaluating population projection errors for census tracts, paper presented at the Annual Meeting of the American Statistical Association.

Smith, S. K. and Sincich, T. (1991), 'An empirical analysis of the effect of length of forecast horizon on population forecast errors', *Demography* **28**(2), 261–274.

Swanson, D. A. and Beck, D. M. (1994), 'A new short-term county population projection method', *Journal of Economic and Social Measurement* **20**(1), 25–50.

Tayman, J., Smith, S. K. and Lin, J. (2007), 'Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models', *Population Research and Policy Review* **26**(3), 347–369.

Thompson, P. A., Bell, W. R., Long, J. F. and Miller, R. B. (1989), 'Multivariate time series projections of parameterized age-specific fertility rates', *Journal of the American Statistical Association* **84**(407), 689–699.

Yang, S. S., Yue, J. C. and Huang, H.-C. (2010), 'Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models', *Insurance Mathematics and Economics* **46**(1), 254–270.

Country	HU methods			Univariate methods		
	HU	HUrob	HUw	RW	RWD	ARIMA
Australia	1921	1921	1921	1921	1921	1921
Austria	1951	1951	1951	1951	1951	1951
Canada	1921	1921	1921	1921	1921	1921
Czech Republic	1950	1950	1950	1950	1950	1950
England & Wales	1938	1938	1938	1938	1938	1938
Finland	1939	1939	1939	1939	1939	1939
France	1946	1946	1946	1946	1946	1946
Hungary	1950	1950	1950	1950	1950	1950
Netherlands	1950	1950	1950	1950	1950	1950
Scotland	1945	1945	1945	1945	1945	1945
Slovakia	1950	1950	1950	1950	1950	1950
Spain	1922	1922	1922	1922	1922	1922
Sweden	1891	1891	1891	1891	1891	1891
Switzerland	1932	1932	1932	1932	1932	1932
USA	1933	1933	1933	1933	1933	1933

Table 1: Commencing year of the initial fitting period for each country and method. The abbreviations of different methods are elaborated in Section 3.

h	HU methods			Univariate methods		
	HU	HUrob	HUw	RW	RWD	ARIMA
1	0.0017	0.0022	0.0013	0.0016	0.0017	0.0015
2	0.0026	0.0031	0.0022	0.0028	0.0031	0.0025
3	0.0036	0.0040	0.0032	0.0040	0.0046	0.0036
4	0.0046	0.0051	0.0043	0.0052	0.0060	0.0047
5	0.0056	0.0061	0.0053	0.0064	0.0074	0.0059
6	0.0067	0.0072	0.0064	0.0075	0.0087	0.0071
7	0.0077	0.0082	0.0075	0.0086	0.0101	0.0082
8	0.0088	0.0093	0.0086	0.0098	0.0115	0.0094
9	0.0099	0.0104	0.0097	0.0108	0.0128	0.0105
10	0.0109	0.0115	0.0107	0.0119	0.0141	0.0116
11	0.0119	0.0126	0.0116	0.0129	0.0155	0.0127
12	0.0128	0.0137	0.0125	0.0139	0.0169	0.0137
13	0.0140	0.0149	0.0138	0.0149	0.0182	0.0149
14	0.0150	0.0160	0.0145	0.0159	0.0195	0.0159
15	0.0163	0.0171	0.0160	0.0168	0.0208	0.0167
16	0.0168	0.0182	0.0161	0.0175	0.0221	0.0174
17	0.0175	0.0192	0.0168	0.0183	0.0234	0.0177
18	0.0179	0.0202	0.0176	0.0192	0.0248	0.0183
19	0.0190	0.0201	0.0188	0.0200	0.0261	0.0192
20	0.0202	0.0225	0.0196	0.0209	0.0276	0.0202
Mean	0.0111	0.0121	0.0108	0.0119	0.0148	0.0116
Median	0.0114	0.0121	0.0111	0.0124	0.0148	0.0122
Mean(w)	0.0082	0.0087	0.0076	0.0093	0.0112	0.0086
Median(w)	0.0085	0.0090	0.0081	0.0096	0.0113	0.0089

Table 2: MAFEs of the age-specific fertility rates for the one-step-ahead to 20-step-ahead point forecasts by method and forecast horizon, averaged over the ages, years in the forecasting period, and 15 countries.

h	HU methods			Univariate methods		
	HU	HUrob	HUw	RW	RWD	ARIMA
1	-0.0002	-0.0002	0.0000	-0.0001	0.0004	0.0000
2	-0.0002	-0.0002	-0.0001	-0.0003	0.0009	-0.0001
3	-0.0003	-0.0003	-0.0002	-0.0006	0.0012	-0.0003
4	-0.0005	-0.0005	-0.0004	-0.0010	0.0014	-0.0005
5	-0.0008	-0.0008	-0.0007	-0.0014	0.0015	-0.0008
6	-0.0013	-0.0013	-0.0011	-0.0020	0.0016	-0.0013
7	-0.0017	-0.0016	-0.0015	-0.0026	0.0016	-0.0016
8	-0.0022	-0.0021	-0.0018	-0.0031	0.0017	-0.0021
9	-0.0026	-0.0026	-0.0023	-0.0036	0.0017	-0.0025
10	-0.0031	-0.0031	-0.0030	-0.0041	0.0019	-0.0029
11	-0.0038	-0.0038	-0.0036	-0.0045	0.0020	-0.0034
12	-0.0046	-0.0045	-0.0045	-0.0049	0.0023	-0.0039
13	-0.0053	-0.0052	-0.0055	-0.0051	0.0026	-0.0043
14	-0.0057	-0.0056	-0.0063	-0.0051	0.0032	-0.0045
15	-0.0057	-0.0056	-0.0057	-0.0050	0.0040	-0.0044
16	-0.0053	-0.0052	-0.0057	-0.0046	0.0051	-0.0036
17	-0.0042	-0.0041	-0.0048	-0.0041	0.0064	-0.0018
18	-0.0028	-0.0027	-0.0043	-0.0037	0.0076	-0.0006
19	-0.0015	-0.0013	-0.0037	-0.0031	0.0089	0.0007
20	-0.0015	-0.0010	-0.0037	-0.0027	0.0100	0.0015
Mean	-0.0027	-0.0026	-0.0029	-0.0031	0.0033	-0.0018
Median	-0.0024	-0.0023	-0.0033	-0.0034	0.0020	-0.0017
Mean(w)	-0.0004	-0.0004	-0.0010	0.0005	0.0042	0.0016
Median(w)	-0.0004	-0.0004	-0.0006	0.0002	0.0032	0.0004

Table 3: MFEs of the age-specific fertility rates for the one-step-ahead to 20-step-ahead point forecasts by method and forecast horizon, averaged over the ages, years in the forecasting period, and 15 countries.

h	HU methods			Univariate methods		
	HU	HUrob	HUw	RW	RWD	ARIMA
1	0.0918(+)	0.0915(-)	0.0654(+)	0.1602(+)	0.1418(+)	0.0776(+)
2	0.0986(+)	0.0993(-)	0.0665(+)	0.1322(+)	0.0995(+)	0.0792(+)
3	0.1095(-)	0.1075(-)	0.0674(+)	0.1044(+)	0.0943(+)	0.0991(-)
4	0.1210(-)	0.1187(-)	0.0787(+)	0.0949(+)	0.1042(-)	0.1099(-)
5	0.1283(-)	0.1273(-)	0.0844(+)	0.0880(+)	0.1138(-)	0.1243(-)
6	0.1316(-)	0.1284(-)	0.0937(+)	0.0850(+)	0.1449(-)	0.1358(-)
7	0.1367(-)	0.1343(-)	0.1002(+)	0.0876(+)	0.1778(-)	0.1447(-)
8	0.1386(-)	0.1374(-)	0.1070(+)	0.0944(-)	0.2053(-)	0.1513(-)
9	0.1435(-)	0.1425(-)	0.1060(+)	0.0967(-)	0.2319(-)	0.1549(-)
10	0.1541(-)	0.1546(-)	0.1125(+)	0.1020(-)	0.2514(-)	0.1536(-)
11	0.1621(-)	0.1585(-)	0.1167(+)	0.1103(-)	0.2688(-)	0.1621(-)
12	0.1619(-)	0.1589(-)	0.1118(+)	0.1149(-)	0.2851(-)	0.1663(-)
13	0.1695(-)	0.1652(-)	0.1262(+)	0.1198(-)	0.2967(-)	0.1762(-)
14	0.1752(-)	0.1736(-)	0.1409(+)	0.1219(-)	0.3042(-)	0.1843(-)
15	0.1762(-)	0.1740(-)	0.1471(+)	0.1295(-)	0.3117(-)	0.1906(-)
16	0.1848(-)	0.1810(-)	0.1464(+)	0.1417(-)	0.3185(-)	0.1987(-)
17	0.1890(-)	0.1829(-)	0.1495(+)	0.1567(-)	0.3190(-)	0.2100(-)
18	0.1860(-)	0.1803(-)	0.1579(+)	0.1657(-)	0.3156(-)	0.2159(-)
19	0.2038(-)	0.1981(-)	0.1452(+)	0.1762(-)	0.3229(-)	0.2267(-)
20	0.1867(-)	0.1771(-)	0.1338(-)	0.1943(-)	0.3295(-)	0.2495(-)
Mean	0.1525(-)	0.1496(-)	0.1129(+)	0.1238(-)	0.2318(-)	0.1605(-)
Median	0.1580(-)	0.1566(-)	0.1122(+)	0.1173(-)	0.2601(-)	0.1585(-)
Mean(w)	0.1101(-)	0.1016(-)	0.0846(+)	0.1486(-)	0.2428(-)	0.1854(-)
Median(w)	0.1087(-)	0.1035(-)	0.0714(+)	0.1486(-)	0.2760(-)	0.1780(-)

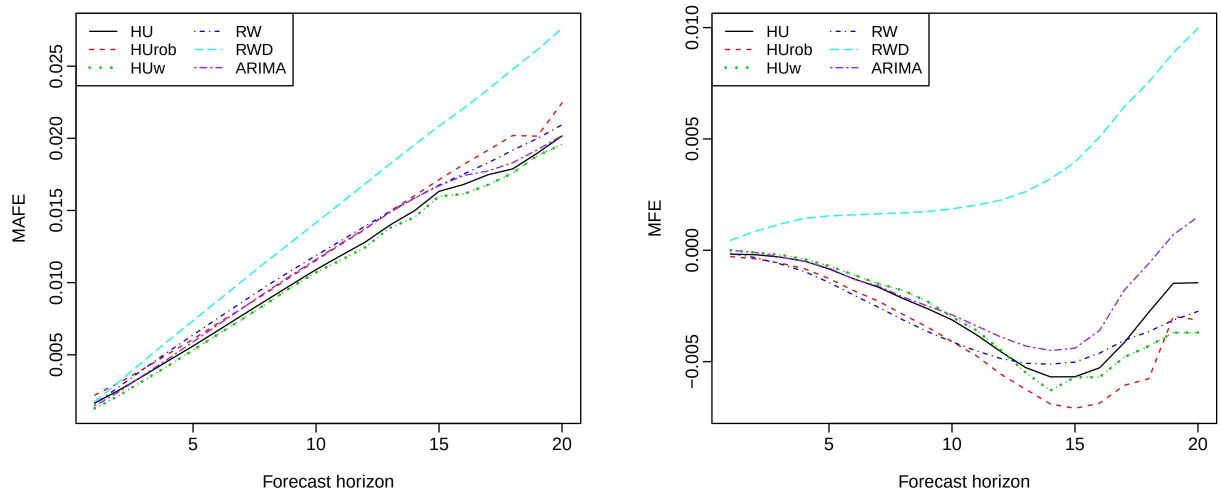
Table 4: Coverage probability deviances of the forecasted fertility rates by method and forecast horizon, averaged over different countries at the nominal coverage probability of 80%. Plus sign indicates that the empirical coverage probability averaged over countries is greater than the nominal coverage probability. For instance, 0.0918(+) means that the coverage probability deviance, which is the mean(|empirical coverage probability - nominal coverage probability|) is 0.0918, and the averaged empirical coverage probability is greater than the nominal coverage probability.

h	HU methods			Univariate methods		
	HU	HUrob	HUw	RW	RWD	ARIMA
1	0.0435(-)	0.0657(-)	0.0461(+)	0.0401(+)	0.0389(+)	0.0308(-)
2	0.0540(-)	0.0663(-)	0.0457(+)	0.0403(+)	0.0362(+)	0.0352(-)
3	0.0675(-)	0.0742(-)	0.0428(+)	0.0432(+)	0.0461(+)	0.0548(-)
4	0.0836(-)	0.0938(-)	0.0420(+)	0.0471(+)	0.0581(-)	0.0766(-)
5	0.0935(-)	0.0991(-)	0.0439(+)	0.0483(-)	0.0759(-)	0.0968(-)
6	0.0969(-)	0.1046(-)	0.0436(+)	0.0540(-)	0.0926(-)	0.1175(-)
7	0.1053(-)	0.1067(-)	0.0524(+)	0.0609(-)	0.1135(-)	0.1377(-)
8	0.1097(-)	0.1121(-)	0.0562(+)	0.0681(-)	0.1323(-)	0.1544(-)
9	0.1132(-)	0.1138(-)	0.0573(+)	0.0703(-)	0.1467(-)	0.1714(-)
10	0.1135(-)	0.1225(-)	0.0740(-)	0.0743(-)	0.1618(-)	0.1847(-)
11	0.1167(-)	0.1262(-)	0.0797(-)	0.0805(-)	0.1871(-)	0.2000(-)
12	0.1177(-)	0.1260(-)	0.0808(-)	0.0860(-)	0.2118(-)	0.2117(-)
13	0.1238(-)	0.1339(-)	0.0921(-)	0.0938(-)	0.2392(-)	0.2215(-)
14	0.1357(-)	0.1437(-)	0.0959(-)	0.1050(-)	0.2640(-)	0.2295(-)
15	0.1382(-)	0.1472(-)	0.0948(-)	0.1066(-)	0.2809(-)	0.2378(-)
16	0.1457(-)	0.1494(-)	0.1010(-)	0.1124(-)	0.2995(-)	0.2469(-)
17	0.1505(-)	0.1552(-)	0.1110(-)	0.1168(-)	0.3174(-)	0.2567(-)
18	0.1482(-)	0.1613(-)	0.1167(-)	0.1219(-)	0.3282(-)	0.2624(-)
19	0.1520(-)	0.1790(-)	0.1186(-)	0.1280(-)	0.3379(-)	0.2719(-)
20	0.1589(-)	0.1855(-)	0.1395(-)	0.1334(-)	0.3419(-)	0.2757(-)
Mean	0.1134(-)	0.1233(-)	0.0767(-)	0.0815(-)	0.1855(-)	0.1737(-)
Median	0.1151(-)	0.1242(-)	0.0768(-)	0.0774(-)	0.1745(-)	0.1924(-)
Mean(w)	0.0525(-)	0.0524(-)	0.0516(-)	0.0756(-)	0.2163(-)	0.1146(-)
Median(w)	0.0499(-)	0.0494(-)	0.0513(-)	0.0452(-)	0.2488(-)	0.1316(-)

Table 5: Coverage probability deviances of the forecasted fertility rates by method and forecast horizon, averaged over different countries at the nominal coverage probability of 95%. Plus sign indicates that the empirical coverage probability averaged over countries is greater than the nominal coverage probability. For instance, 0.0562(+) means that the coverage probability deviance, which is the mean(|empirical coverage probability - nominal coverage probability|) is 0.0562, and the averaged empirical coverage probability is greater than the nominal coverage probability.

	Coverage probability at 80%						Coverage probability at 95%					
	HU methods			Univariate methods			HU methods			Univariate methods		
h	HU	HUrob	HUw	RW	RWD	ARIMA	HU	HUrob	HUw	RW	RWD	ARIMA
1	12	12	23	17	17	10	18	19	58	26	26	15
2	17	18	31	24	24	16	26	27	63	37	37	25
3	21	23	39	30	30	22	33	34	72	45	46	34
4	26	27	46	34	35	28	39	42	80	52	54	43
5	30	32	51	38	40	33	46	50	87	59	61	51
6	34	37	55	42	44	38	52	57	91	64	68	60
7	38	42	63	46	48	44	59	64	99	70	74	68
8	42	47	67	49	52	49	65	71	106	75	80	77
9	46	51	72	52	56	54	71	79	105	79	86	86
10	50	56	76	55	60	60	77	86	118	84	92	96
11	54	61	81	58	64	66	83	94	123	88	97	108
12	58	66	85	60	67	72	89	101	130	92	103	122
13	61	72	91	63	70	80	94	111	143	96	108	140
14	65	77	96	66	74	88	99	119	144	100	113	161
15	68	82	100	68	78	97	105	127	160	104	119	188
16	71	88	106	71	81	105	110	136	160	108	124	218
17	74	94	109	73	84	103	115	146	168	112	130	218
18	77	104	109	75	88	99	119	161	182	115	135	202
19	79	97	116	78	91	94	123	151	180	119	140	175
20	84	100	116	80	95	101	130	156	202	122	146	197

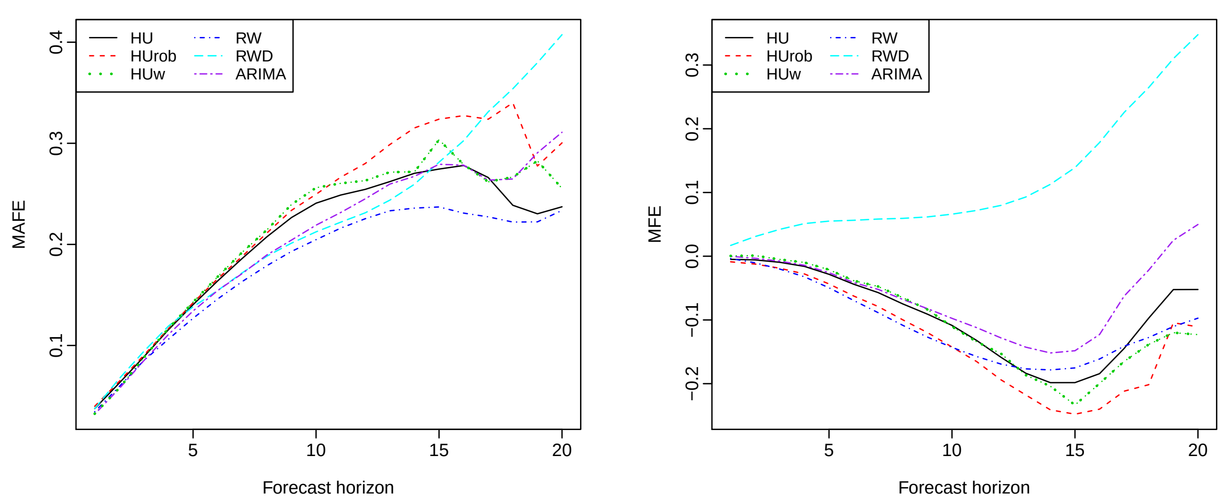
Table 6: Half-width of the forecasted fertility rates by method and forecast horizon, averaged over different countries at the nominal coverage probabilities of 80% and 95%.



(a) MAFEs averaged over years in the forecasting period, ages and countries.

(b) MFEs averaged over years in the forecasting period, ages and countries.

Figure 1: MAFEs and MFEs of the age-specific fertility rates for the one-step-ahead to 20-step-ahead point forecasts by method and forecast horizon.



(a) MAFEs averaged over years in the forecasting period, ages and countries.

(b) MFEs averaged over years in the forecasting period, ages and countries.

Figure 2: MAFEs and MFEs of the total fertility rates for the one-step-ahead to 20-step-ahead point forecasts by method and forecast horizon.