ISSN 0038-982X (2011): 2 p.73-89

UDK 303.094.6:314.38(497.11)
DOI:10.2298/STNV1102073N
Originalni naučni rad
Primljeno: 09.06.2011.

FUNCTIONAL DATA ANALYSIS IN FORECASTING SERBIAN FERTILITY

*Vladimir NIKITOVIĆ**

A new approach, combining functional data analysis and principal components decomposition in order to forecasting demographic rates, introduced recently by Hyndman and his associates, is tested on official data series of Serbian age-specific fertility rates available for period 1950-2009. The original concept of the method with its extensions and improvements is applied to region-specific data for the country (Central Serbia and Vojvodina). One of the most important benefits of the method reflected in confirmation that is essentially to model and forecast more than one principal component in order to adequately address sources of variation in fertility. Similarly, modelling and forecasting fertility rates with regards to age and not total fertility rates shows how important it is for the recognized tendency of postponing childbearing in Serbia to be included in coefficients of functional time series. Besides, the method is based completely on evaluation of historical data, without subjective views of forecasters having to be taken into account. Coherent functional product-ratio forecasts of two regions proved to be highly convergent on the long-term not allowing for outliers to contaminate the forecast.

Key words: *population forecasting, functional approach, fertility, Serbia*

Introduction

The volume of national population forecastes produced by stochastic methodology has grown significantly in recent years. Even if the main advantages of this approach, such as probabilistic consistency among all forecast variables and resulted indecies, were well recognized in the literature almost two decades ago (Lee, Tuljapurkar, 1994), it was widely accepted along with the recent fast development of fast computers and improvements in statistical methods. Furthermore, there is an increase in adopting these methods by official national statistic agencies. The project lead by the Dutch official agency was among the first to result in long-term internationally

* Demographic Research Centre of Institute of Social Sciences, Belgrade.

consistent stochastic population forecasts for 18 relatively data-abundant European countries (Statistics Netherlands, 2005).

Since the future size and structure of population are the core of any social and economic plan, the use of stochastic population forecast is the only way to correctly address demographic issues regarding rapid ageing of population. Thus, any improvement in methodology used to forecast fertility and mortality has its clear reflection in improving decisions relating policy making. Given inherent consistency of stochastically based forecast assumptions only such an approach can provide meaningful and widely useful indicators of future uncertainty connected to demographic processes. It is of special interest to improve the forecasting methods of fertility since fluctuations of its level during the last several decades (the baby-boom after World War II and the subsequent decline) were the most important initiator of the process of actual population ageing.

Forecasting fertility is the key issue of the paper. Even though different approaches on stochastic fertility forecasting are proposed in literature, the approaches which are generally based on extrapolation techniques are widely accepted. The ultimate goal of the stochastic forecast is to explicitly address uncertainty about future level of forecasted parameters. Nevertheless, stochactic methods for forecasting fertility are remarkably less developed than methods for forecasting mortality. Some of the main reasons are related to difficulties in forecasting structural changes in fertility, as well as to high degree of impact that selected model has on the estimation of uncertainty (Lee, 1974; Lee, Tuljapurkar, 1994). In that sense, the main advantages in further development of methodology used to forecast fertility should decrease the level of subjective impact of the forecaster.

The most convinient way in terms of data and calculation requirements is to directly forecast total fertility rate (TFR), as presented in Wilson and Bell (2004). Trying to apply procedure similar to Lee-Carter method used for mortality forecasting, Lee (1993) concluded that there is still a need for subjectively imposing limits of prediction interval in order to reduce its width and avoid demographically unrealistic forecasts, as well as to pre-specife the long-term median value of total fertility. Some authors, as Keilman and his associates (2000; 2002), tried to model age-specific rates, but those attepmts, apart from the need for subjectively imposed limits on future TFR, did not prove better than those just employing total fertility.

Some recent developments in this field generally represent further improvements of principal component approach introduced by Lee-Carter method for forecasting mortality. Most recently, Hyndman and Ullah (2007) proposed combination of functional data approach and principal components decomposition to forecast fertility and mortality rates. Several refinments of the method were published thereafter, mostly in the way of improving forecast accuracy, and of extending the application scope (migration included, coherency between sexes and regions) and reducing computational time (Hyndman, Booth, 2008; Hyndman, Shang, 2009; Hyndman et al, 2011).

This method, in short, assumes underlying smooth function of observed data, which should be decomposed via principal components decomposition in order to get coeficients that can be forecasted by means of some univariate time series method. So far, the method (approach) has been tested on approximately 20 populations all around the world, regarding mortality rates, but just to Australian and Sweden regarding fertility rates. In this paper, the approach has been tested on Serbian time series of fertility rates for the past 60 years (1950-2009). The advantages and limitations of the method are considered through its versions published in the period 2007-2011, while the general principles from Hyndman and Ullah (2007) present the basic point of the analysis.

Data and Methods

Demographic indicators forecasted for the Serbian population, made by official statistical agencies, represent results of so called „decomposing approach" to producing forecast in every projection set from the first publication in 1951 until the most recent in 2005 (Nikitović, 2010). Namely, deterministic procedure implied that region-specific hypothesis should be specified separately by every single region for every demographic component. Secondly, forecast indicators for the total population of the country were simply the results of aggregation of projected region populations. Apart from other widely recognized sources of bias induced by deterministic approach, such a perfect correlation accross age and sex as well as perfect serial correlation in components of demographic change, the official procedure, described above, assumed perfect correlation between regions of the country.

In recent papers of Nikitović (2007; 2009; 2010), main advantages of stochastic approach over common deterministic were shown on the example of Serbian population. Unlike the practice of national Statistical Office, where hypothesis on fertility usually determines the variant of population projections due to no alternatives on future mortality and migration[1], full stochastic distribution of parameters of demographic components were produced in order to get internally consistent forecasts of demographic indicators. Yet, stochastic element was provided by applying time series methods to summary indicators, such as total fertility, life expectancy at

[1] Exception to this rule was the projection 1981-2011 regarding migration (Nikitović, 2004).

birth and net migration, rather than to age-specific rates. This is mainly due to computational simplicity. Forecasting future paths of fertility is difficult, due to typical structural changes over time, as well as long-term influence of misspecified fertility assumption on population structure (Lee, 1974; 1998).

The original concept of functional data paradigm introduced in Hyndman and Ullah (2007) assumes combination of functional time series analysis and principal components decomposition. In order to analyse fertility rates and produce their point forecasts, along with prediction intervals, several steps should be taken:

a) Box and Cox transformation[2] of observed data, $y^*_t(x)$, in order to reduce out-of-sample variance, results in transformed data set, $y_t(x)$;

b) Estimate of smooth functions, $s_t(x)$, using a nonparametric smoothing technique applied to $y_t(x)$ for each year t

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

where $\varepsilon_{t,x}$ are iid standard normal variates and $\sigma_t(x)$ allows the variance to change with age and time;

c) Estimate $\mu(x)$ as the mean of $s_t(x)$ across years; estimate coefficients $\beta_{t,k}$ and principal component functions $\phi_k(x)$, k=1,..., K, using a principal components decomposition[3] of $[s_t(x)-\mu(x)]$

$$s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\, \phi_k(x) + e_t(x)$$

where $e_t(x)$ is the model error, assumed to be serially uncorrelated;

d) Estimate time series models[4] for $\beta_{t,k}$;

e) Forecast principal component scores $\beta_{t,k}$ for $t = n + 1, \ldots, n + h$ using the fitted time series models;

f) Multiply the forecasted principal component scores by fixed principal components to obtain forecasts of $s_t(x)$;

g) The estimated variances of the error terms (the sum of: smoothing error in estimating $s_t(x)$, mean function error in estimating $\mu(x)$, error in forecasting the scores $\beta_{t,k}$, and error in the model residuals $e_t(x)$) are used to compute prediction intervals for the forecasts.

After several extensions of the approach, Hyndman, Booth and Yasmeen (2011) recently proposed a method for coherent mortality forecasting. It involves forecasting of an interpretable product and ratio functions of rates using the functional approach presented in Hyndman and Ullah (2007). The

[2] Hyndman and Ullah (2007) use logarithms as a special case of Box and Cox transformation when λ=0.

[3] Hyndman and Shang (2009) proposed functional partial least squares regression, instead.

[4] Hyndman and Ullah (2007) used univariate robust ARIMA models, while Hyndman and Booth (2008) employed exponential smoothing state space models.

method was applied to sex-specific (two subpopulations) mortality rates of Sweden, as well as to state-specific (six subpopulations) mortality rates of Australia. It could even be possible to incorporate dimensions of sex and region in defining the product, in order to achieve coherency between the dimesions as well as within their boundaries (Hyndman et al, 2011).

The functional approach to data analysis was first tested on fertility rates for two subpopulations of Serbia, as two main constitutional parts of the country (Central Serbia and Vojvodina)[5]. Application of such an approach for forecasting mortality rate in Serbia would surely be the sequel to this analysis, which is a topic for some other paper. Forecasting fertility was challenging in Serbian example for at least two reasons:

a) using several principal components proved to be more important for fertility than for mortality forecasts, since the first component explains smaller proportion of the variation in the former than in the latter vital rate (Hyndman, Booth, 2008). Thus, the advantage of the functional approach relating to other stochastic methods used nowadays in projecting Serbian fertility should be more clearly noted than in the case of mortality;

b) modelling age-specific rates instead of total fertility rate provides more accurate forecasts given clear differences between age groups in contribution to overall variability.

Given the tradition in official projections of Serbia, two subpopulations of the country served to testing coherent region-specific forecasting of fertility, too. By construction of the method, coherency in fertility forecasts between two regions should overcome shortage of perfect regional correlation in official projections. Besides, the aim was to test if it is provide more consistent point forecasts and width of prediction intervals between regions than individually produced functional forecasts would for each region due to nature of coherency constraint.

Annual age and region specific fertility rates of Serbian population by single years of age (15-49) and by two regions of the country for the period 1950-2009 were obtained from the Statistical Office of the Republic of Serbia. The rates are defined as the number of live births during each calendar year, according to the age of the mother, per 1000 female resident population of the same age at 30 June.

[5] Third constitutional part of the country, Kosovo and Metohija, is omitted from the analysis due to lack of data for the period 1989-2009 (data for the 1989-1997 period are considered as unreliable).

Results

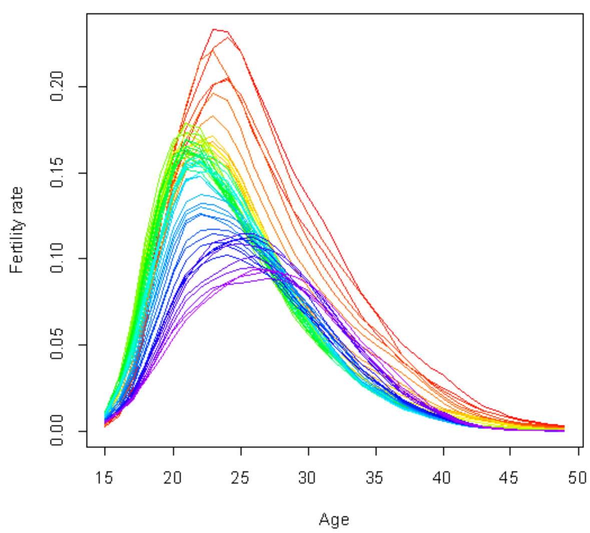
Functional analysis of time series

As noted in Hyndman and Booth (2008), there is no obvious consensus on best transformation to use for fertility data. The value of λ=0.4 produces relatively small out-of-sample forecast errors and narrowest prediction intervals when applied to the analysed data sets in this paper for the period 1950-2009. Transformed data, $y_t(x)$, were then smoothed using a weighted median smoothing B-spline, constrained to be concave (Figure 1) as suggested in Hyndman and Booth (2007).

Unfortunately, available data sets disaggregated by age of mother are restricted only to the period after the World War II. Longer data sets would surely easily depict early 1950s as the baby-boom period in both regions of Serbia.

Figure 1a

Smoothed age-specific fertility rates of Central Serbia viewed as functional time series for ages 15-49, observed from 1950 to 2009

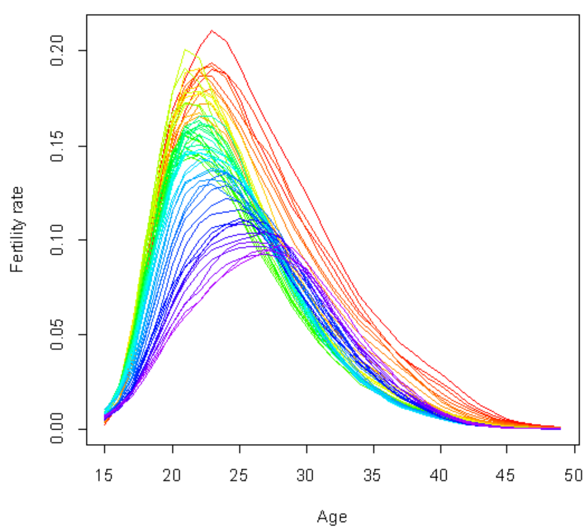


Note: In the web version of the paper, curves are ordered chronologically according to the colors of the rainbow – the oldest are shown in red, with the most recent in violet.

However, Figure 1 demonstrates very high rates at all ages in that period, followed by rapid decline during late 1950s and early 1960s, and then an increase at higher ages in more recent years. Strong decrease of fertility after baby-boom period coincided with highly intensive industrialization that transformed Serbia profoundly, from a pre-war predominantly rural and agricultural country to a fast-developing industrial society entering late phases of demographic transition (Nikitović, 2004). Some of general determinants of fertility decline, like the adoption of new norms and values regarding childbearing and partnership, growing levels of female labor force, and birth control, are recognized in the case of Serbia, too. But, there is a particularity concerning birth control in Serbia that had great impact on early break of after-war high fertility. While in the rest of Europe introduction of modern contraceptive methods influenced fertility decline, in Serbia, it was high increase of abortion rate. Unfortunatelly, it is still highly common contraceptive methods in low feritlity Serbia nowadays since the total abortion rate is double the total fertility rate (Rašević, Sedlecki, 2011).

Figure 1b

Smoothed age-specific fertility rates of Vojvodina viewed as functional time series for ages 15-49, observed from 1950 to 2009.



Increase of fertility at higher ages is initiated by increased interest in tertiary education and the trend to postpone raising children in exchange for career pursuit. The very high psychological and economic value of parenthood in Serbia should also be taken into account (Rašević, 2009).

Robust approach of Hyndman and Ullah (2007) offered robust estimation of principal components in order to handle outliers. Namely, it uses simple statistical tests for outliers in order to assign weights zero and one to outlying and non-outlying observations respectively. Later extensions of the method implied two different approaches. The more simple one just shortened the time scope of data series (excluding periods with outliers) used for fitting the model, as in Hyndman and Booth (2008), while the more sophisticated approach employed some of the new graphical methods in this field, as functional HDR box plot (Hyndman and Shang, 2008). This proved to be sensitive enough to discover extreme data points that most of the earlier popular techniques would omit. In this paper, data sets are tested for outliers by several outlier detection methods for functional data (Table 1).

Table 1

Detected Outliers by Various Methods Aplied to the Serbian Fertility Data

Method	Detected outliers	
	Central Serbia	Vojvodina
Robust Mahalanobis distance	1950-1954	1950
Likelihood ratio test	1950-1954	1950-1952
Integrated square forecast errors	None	None
Functional depth	None	None
Functional bagplot	1950-1954	1950
Functional HDR boxplot*	1950, 1952, 1955, 1956	1950, 1952, 1955, 1956
*Note**: Coverage probability of the outer region was set to 93%.		

Three methods suggested 1950-1954 was the period for exclusion from further analysis for Central Serbia, while there is no obvious concensus on that for Vojvodina. Clearly, the baby-boom period is the issue. If longer data sets were available, (at least including the World War II and first several after-war years), baby-boom years would probably not be pointed out as outliers. In Serbia as a whole, baby-boom period was shorter compared to the most of European countries (sharp decline of total fertility rate occurred 10-15 years earlier). One of the most influential reasons could lie in early liberalization of women's right to abortion. Since the first acts in 1952, set of laws largely liberalized this right until 1969 and enabled abortion to be the dominant contraceptive technique in the period when the modern contraceptive means were not yet available in Serbia (Rašević, Sedlecki, 2011).

A difference between Central Serbia and Vojvodina in TFR level for the 1950-2009 period cannot be considered significant if figures representing this summary indicator are analyzed as noted at first sight (Figure 2). Yet, given the start level in 1950, the fall of TFR was not so steep in Vojvodina comparing to Central Serbia.

Years reflecting short baby-boom period in Central Serbia are identified as the outliers by functional high density region (HDR) boxplot method. This is easy to explain since the relatively short period of the time series (1950-2009) covers mostly the below-replacement fertility period without structural breaks in it. For that reason, this method suggested shorter series (1957 as the start year if coverage probability of outer HDR region is set to 93%) as the basis for fitting the functional model. The same suggests the analysis of TFR series if one wants to get more plausible prediction intervals (Nikitović, 2010). Besides that, analysis of historic forecast errors in official projections of the population of Serbia from 1971 to present time also advocates more narrow intervals of forecast than would have been gotten if the 1950 had been the starting year of the series (Nikitović, 2004; 2007).

Figure 2

Total Fertility Rate in Central Serbia (solid line) and Vojvodina (dashed line) in 1950-2009

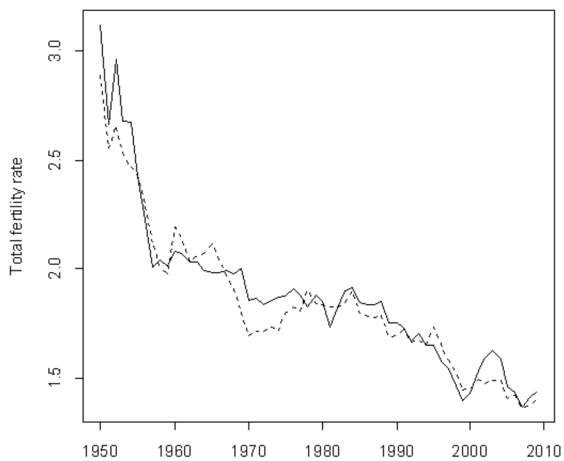
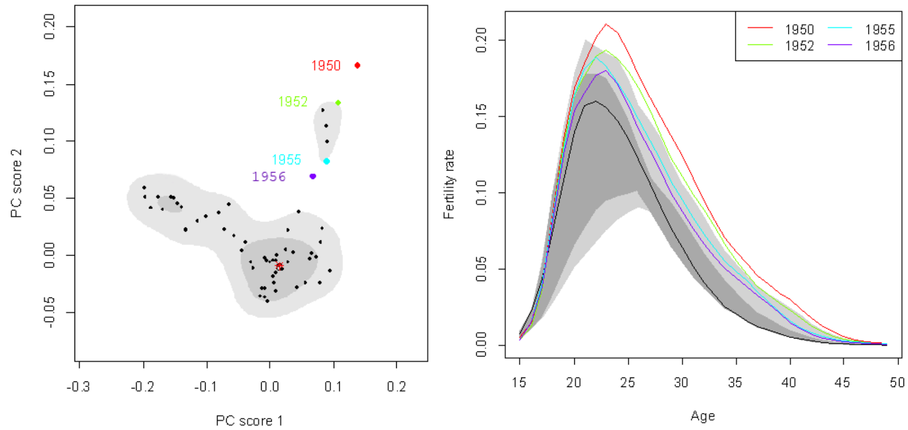


Figure 3 presents fertility in Vojvodina through functional HDR boxplot, according to Hyndman and Shang (2008). The figure shows bimodality which indicates that samples may come from two populations with different locations. The closer analysis of figure 1 suggests that two regions correspond to two periods with different level (it is obvious even if TFR at Figure 2 is analyzed). Namely, the decrease from baby-boom period to actual level happened through two stages, i.e. not so suddenly and sharply as in the case of Central Serbia, where baby-boom period was identified as outliers. Decrease of total fertility after baby-boom period in Central Serbia was stronger than in Vojvodina on account of steeper decrease at ages of highest reproductive potential (22-35). Similarly, the presence of outer HDR region (even with 93% coverage probability) around specific baby-boom years confirms assumption that baby-boom years probably should not be considered as outliers, at least in the case of Vojvodina.

Figure 3

Bivariate and functional HDR boxplot for the Vojvodina fertility rates



Note: The dark and light grey regions show 50% (inner) and 93% (outer) HDR respectively. The black line is the modal curve. The curves outside the outer region are outliers.

Forecasting Functional Time Series Coefficients

In this analysis, the number of basic functions, $\phi_k(x)$, k=1,..., K, to be calculated using principal components decomposition, was set to K=6. Hyndman and Booth (2008) found that the method is insensitive to the choice of K, provided K is large enough. The risk of choosing a small K reflects in poor forecast accuracy. Even if it seems that K=3 is quite appropriate for this analysis, the only cost in choosing larger K was additional computational time, which could be neglected given the speed of modern computers.

Due to relatively short time series of fertility rates and stated considerations on outliers' detection, the whole available period (1950-2009) was tested for the fitting and forecasting procedures. The use of ARIMA methods for forecast of functional basis coefficients, derived from principal decomposition in the functional data approach, resulted in extremely wide prediction intervals of TFR for Central Serbia[6] due to influence of outliers

[6] In 2050 the width of 95% prediction interval of TFR is 0.36-6.43 with point forecast of 1.73.

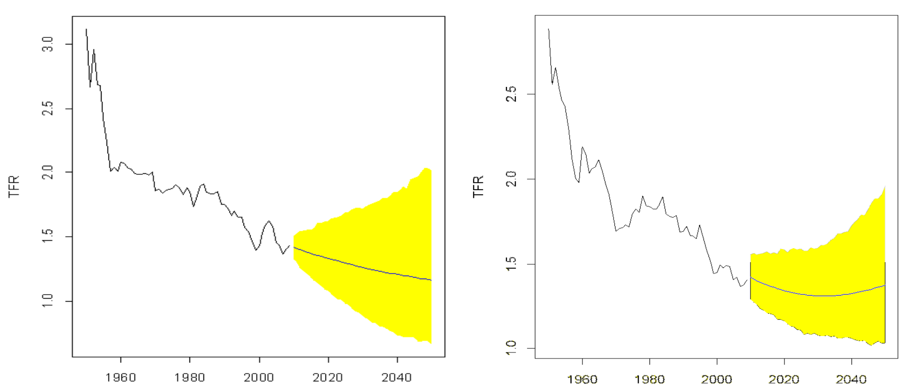
from the short baby-boom period, even if adjustment of forecast variance suggested in Hyndman and Booth (2008) is applied. This is the confirmation of findings which came from outlier detection methods that were applied. Point forecast in 2050 reflects quite optimistic increase of TFR which was experienced last time in the beginning of 1990s. Such a distribution of outcomes is in collision with forecasts produced by diverse stochastic methodology, from nationally correlated forecasts for EEA countries (Statistics Netherlands, 2005) to several individual national forecasts (Alho, 2001; Keilman et al, 2002; Wilson, Bell, 2004; Torri, Vignoli, 2007).

ARIMA methods used to forecast functional basis coefficients resulted in quite realistic prediction intervals of TFR for Vojvodina[7]. Unlike Central Serbia, baby-boom years are not recognized as outliers rather as the first stage of after-war decreasing trend of TFR, which is in accordance with previously reported results of outlier detection.

On the other side, exponential state-space models with additive trend and error term provide more realistic point forecast of TFR for both analyzed regions (Figure 4) when used to forecast functional basis coefficients, as suggested in Hyndman and Booth (2008). Even further decrease of this

Figure 4

Forecasts of total fertility rate in Central Serbia and Vojvodina for 2010-2050



summary indicator is in accordance with forecasts for countries with similar history of TFR trend in past 60 years (Statistics Netherlands, 2005). But, it could be argued that these models produce too confident forecasts of TFR in Vojvodina (width of 95% prediction interval of around 0.9) for the 41-year period, while the lower 95% limit of the interval is just about one, which is a

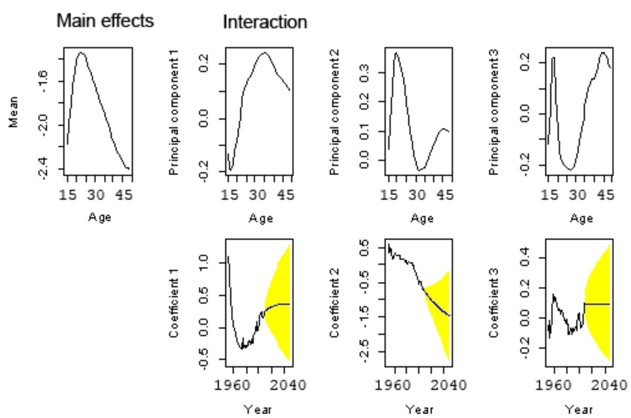
[7] The width of 95% prediction interval of TFR is 0.68-2.16 with point forecast of 1.28.

bit higher than ever recorded (Statistics Netherlands, 2005). At the same time, upper limit of the 95% prediction interval does not allow for replacement level to be achieved, even being very close to it, until the end of projection period, in both regions (Figure 4). It seems realistic, given the length of the projection horizon and the past 50 years of below-replacement fertility in the country. Yet, as Alho (2001) pointed out in the forecast for Lithuanian population (similar demographic characteristic and recent historic background comparing to Serbia), there are no convincing arguments that the future level of TFR will be higher or lower than the present one, implying the random walk as the best guess.

Only the first three functional principal components and their associated scores for Serbian fertility are presented for the sake of simplicity in interpretation (Figure 5). The remaining three components that were used in the process of modelling and forecasting account for less than 2.0% of

Figure 5a

The first three functional principal components and associated scores (forecasts with 95% prediction intervals in 2010-2050) for fertility data of Central Serbia



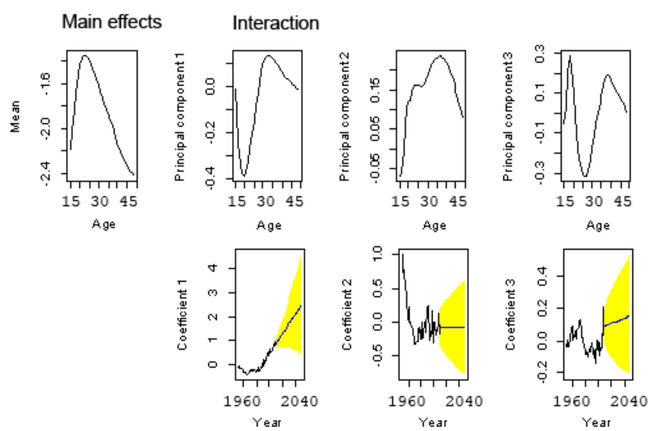
variation and do not have simple and direct demographic interpretation. Forecasts of functional principal scores shown in Figure 5 are based on exponential smoothing state space models with additive trend and error term as described in Hyndman and Booth (2008).

The first three basis functions explain 51.0%, 45.1% and 2.0% of the variation in fertility of Central Serbia. By analogy, the values for Vojvodina are: 56.9%, 38.5% and 2,3%[8]. Clearly, principal components are modelling the fertility rates of mothers in different age groups.

In Central Serbia $\phi_1(x)$ models mothers at ages higher than 25, $\phi_2(x)$ models young mothers in their teens and first half of 20s, $\phi_3(x)$ models the difference between very young mothers up to age of 20 and late mothers aged 35 and above (Figure 5a). Some of general determinants of fertility change, commented earlier, can be recognized through coefficients associated with each functional principal component. The decrease in $\beta_{t,1}$ (young mothers) during the whole period, accelerated in the last 20 years. The time series model predicts this tendency as the most probable one until the middle of the century. The fastest decrease after baby-boom period shows $\beta_{t,2}$ (older mothers), while its recent noticeable recovery, due to tendency to postpone childbearing, is expected to be quite slow in the forthcoming decades. The fast increase in $\beta_{t,3}$ (the difference between very young and very late mothers) just after the baby-boom, due to decline in fertility of older women, diminished over time. Recent increase of fertility in

Figure 5b

The first three functional principal components and associated scores (forecasts with the 95% prediction intervals in 2010-2050) for fertility data of Vojvodina



closing end of reproductive ages resulted again in the rise of $\beta_{t,3}$, but the level of difference between age groups captured by time series model is not predicted to change in the next decades. Interpretation of functional principal components for fertility in Vojvodina (Figure 5b) would be similar to

[8] The remaining principal components explain 0.7%, 0.4% and 0.2% of the variation in fertility of Central Serbia, and 1.0%, 0.5% and 0.2% of the variation in fertility of Vojvodina. In both regions, only 0.6% of the variation in fertility is not explained by the model.

Central Serbia. The main difference is that the first two basis functions, that account for 95.4% of the variation, model older mothers. The recent noticeable tendency in fertility increase of mothers in early 30s is predicted to continue over the next decades. At the same time, the time series model does not forecast the situation that the coefficient representing fertility changes in mothers in their late 30s rises to the baby-boom level again, but instead it presupposes that their current level will not be changing in the decades to come.

* * *

Even if an assumption of perfect correlation between regions of a country does not usually lead to a higher degree of error cancellation when forecasts for regions are aggregated into one single country forecast (Lee, 1998), additional information on co-variance structure between regions would undoubtely improve forecast accuracy. Yet, stochastic co-variance matrices imply rather complex calculations. In that way, the most recent extension to the method of Hyndman and his associates does not require vector models for coherent or non-divergent forecasts betwen subpopulations. The core of the method assumes functional product-ratio method that models and forecasts the geometric mean of subpopulation rates and the ratio of subpopulation rates to product rates. Coherence is imposed by constraining the forecast ratio function through stationary time series models. Features of the functional principal decomposition, described earlier, imply that time series coefficents of the product functions are all uncorrelated. Similarly, the ratio coefficients are approximately uncorrelated with each other due to use of products and ratios (Hyndman et al, 2011).

Apart from being easily interpretable and computationally efficient, the method ensures coherence in terms of long-term non-divergence of forecast rates of subpopulations. Application of the method to fertility rates of Central Serbia and Vojodina proved its ability to maintain certain structural relationships based on extensive historic observation. Since outliers can be handled by the default of the method, the whole available period (1950-2009) was modelled by ARIMA (product functions) and ARFIMA (ratio functions) methods, which enabled the long-term forecasts of the ratio functions to converge to the mean ratios. The resulted point forecasts of TFR for two subpopulations of the country were much closer to each other (1.37 and 1.34 in 2050 for Central Serbia and Vojvodina, respectively) than in independent forecasts by ARIMA methods (1.73 and 1.28 in 2050) described in detail earlier. Above all, that means that point frecast (along with unrealistically high upper limit of prediction interval) of TFR for Central Serbia is levelled down in order to converge the TFR for Vojvodina (not so heavily influenced by baby-boom outliers) on the long run. The 80%

prediction interval in 2050 ranges from 0.86 to 2.04 in Central Serbia, and from 0.84 to 2.03 in Vojvodina. The concept of balanced margin of error in all regions is often more desirable for the planning purposes than to risk a serious error for one region (Hyndman et al, 2011).

Conclusions

As regards to fertility, the greatest benefit from the method employed in this paper probably arises from using more than one principal component in modelling and forecasting. Since the methodology of fertility forecasting is much less developed than the one of mortality, the reason for this being widely known difficulties in fertility forecasting, any method that is able to capture and model diverse sources of variation could be considered an improvement. Besides, simplicity of calculation and model interpretation play an important role. This is particularly important if we are aware that one of the biggest obstacles for wide acceptance of stochastic approach to population forecasting is its rather complex calculation basis. Thus, a lot of stochastic forecasts are simplified by taking into account variation in total rates rather than in age-specific, as was the case in the first stochastic projection of Serbian population (Nikitović, 2010).

In that way, the new method employed in this paper, provided an improvement in forecasting fertility rates of Serbia. It showed the importance of modelling different age groups for the purpose of better forecast accuracy. Furthermore, the functional principal decomposition of fertility rates confirmed presence of tendency to postpone childbearing in Serbia. The amount of variation due to this source shows how important is to model its relative contribution to overall fertility level in both analyzed regions of the country. However, it turns out to be almost equally important to model fertility separately, both of young and older mothers in Central Serbia. The tendency to postpone childbearing in this region is forecasted by the model that assumes further fertility decrease in younger ages and slow increase in closing ages of reproductive span.

Simple tests for outlier detection, which were incorporated in the original robust approach to forecasting time series coefficients derived from functional principal decomposition (Hyndman, Ullah, 2007), did not identify baby-boom years as outliers, which in case of Central Serbia resulted in quite optimistic point forecast, along with unrealistic upper limit of prediction interval. Yet, the analysis of more sensitive tests for detecting outliers points out the problem of inadequate length of the time series available rather than it shows the need for that relatively short baby-boom span (short in comparison with the majority of European countries) to be excluded from the same series. This is confirmed by overly confident forecasts of models that excluded baby-boom years. Finally, coherent extension of the method was proven to produce highly convergent forecasts for the two regions on the long-term, while at the same time not allowing outliers to contaminate the forecast. Having relatively high percentage of correlation in total fertility rate observed between Central Serbia and Vojvodina, the use of independent forecasts for the two regions, as described in this paper, would not result in great cancellation of errors when forecasts for regions are aggregated into the forecast for the whole country. However, further improvements in forecasting Serbian population should certainly benefit from the new methodology tested in this paper. This is especially because of the fact that implementation of the approach is somewhat facilitated by the use of demography package (made by R. J. Hyndman; the last version is from 2011) for free statistical software R.

Literature

ALHO, J. M. (2001). "Stochastic Forecast of the Lithuanian Population 2001-2050", Research Report P98-1023-R, European Union's Phare ACE Research Project.

HYNDMAN, R. J., H. BOOTH (2008). "Stochastic Population Forecasts Using Functional Data Models for Mortality, Fertility and Migration", *International Journal of Forecasting*, No. 24(3), 323–342.

HYNDMAN, R. J., H. BOOTH, F. YASMEEN (2011). "Coherent Mortality Forecasting: The Product-Ratio Method with Functional Time Series Models", Working Paper 01/11 (Clayton: Monash University, Department of Econometrics and Business Statistics)

HYNDMAN, R. J., H. L. SHANG (2009). "Forecasting Functional Time Series", *Journal of the Korean Statistical Society*, No. 38(3), 199–221.

HYNDMAN, R. J., H. L. SHANG (2010). "Rainbow Plots, Bagplots and Boxplots for Functional Data", *Journal of Computational & Graphical Statistics*, No. 19(1), 29–45.

HYNDMAN, R. J., M. S. ULLAH (2007). "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach", *Computational Statistics & Data Analysis*, No. 51, 4942–4956.

KEILMAN, N., D. Q. PHAM (2000), "Predictive Intervals for Age-Specific Fertility", *European Journal of Population*, No. 16, 41-66.

KEILMAN, N., D. Q. PHAM, A. HETLAND (2002). "Why Population Forecasts Should Be Probabilistic – Illustrated by the Case of Norway", *Demographic Research*, No. 6, 408-454.

LEE, R. D. (1974). "Forecasting Births in Post-Transition Populations: Stochastic Renewal with Serially Correlated Fertility", *Journal of the American Statistical Association*, No. 69(347), 607-617.

LEE, R. D. (1993). "Modeling and Forecasting the Time Series of U.S. Fertility: Age Distribution, Range, and Ultimate Level", *International Journal of Forecasting*, No. 9, 187−202.

LEE, R. D. (1998). "Probabilistic Approaches to Population Forecasting", *Population and Development Review* 24, Supplement: Frontiers of Population Forecasting.

LEE, R. D., S. TULJAPURKAR (1994). "Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low", *Journal of the American Statistical Association*, No. 89(428), 1175-1189.

NIKITOVIĆ, V. (2004). *Tačnost projekcija stanovništva Srbije* (Beograd: Geografski institut "Jovan Cvijić" SANU).

NIKITOVIĆ, V. (2007). "Stohastička projekcija stanovništva Centralne Srbije na osnovu empirijskih projekcionih grešaka", *Stanovništvo,* god. XLV, br.1, 7-31.

NIKITOVIĆ, V. (2010). *Demografska budućnost Srbije: Imigracija kao izvesnost?* (Beograd: Službeni glasnik, Institut društvenih nauka).

RAŠEVIĆ, M. (2009). "Populaciona politika: stanje i očekivanja", *Stanovništvo,* god. XLVII, br.1, 53-66.

RAŠEVIĆ, M., K. SEDLECKI (2011). "Pitanje postojanja abortusne kulture u Srbiji", *Stanovništvo,* god. XLIX, br.1, 7-21.

STATISTICS NETHERLANDS (2005). *Changing Population of Europe: Uncertain Future* (Hague: Statistics Netherlands).

TORRI, T., D. VIGNOLI (2007). "Forecasting the Italian Population, 2005-2055: A Stochastic Approach", *Genus*, No. 1-2, 93-118.

WILSON, T., M. BELL (2004). "Australia’s Uncertain Demographic Future", *Demographic Research*, No. 11, 194-234.

Vladimir Nikitović

Funkcionalna analiza vremenskih serija u prognoziranju fertiliteta Srbije

R e z i m e

Novi pristup prognoziranju demografskih stopa, koji kombinuje funkcionalnu analizu vremenskih serija i metod glavnih komponenata, nedavno predstavljen od strane Hyndman-a i njegovih saradnika, testiran je na zvaničnim vremenskim serijama stopa fertiliteta specifičnih prema starosti za Srbiju u periodu 1950-2009. Originalni koncept metoda, sa poboljšanjima i ekstenzijama objavljivanim naknadno, primenjen je na podatke za Centralnu Srbiju odnosno Vojvodinu. Jedna od najvažnijih prednosti metoda ogleda se u potvrdi činjenice da je bitno modelirati i prognozirati više od jedne glavne komponente da bi se adekvatno obuhvatili izvori varijacije u fertilitetu. Slično, modeliranje i prognoziranje stopa fertiliteta specifičnih prema starosti umesto ukupne stope fertiliteta ukazuje koliko je važno da prepoznata tendencija odlaganja rađanja u Srbiji bude obuhvaćena u koeficijentima funckionalnih vremenskih serija. Pored toga, metod se u potpunosti zasniva na oceni istorijskih podataka, uopšte ne zahtevajući subjektivni upliv prognostičara. Konačno, koherentne prognoze fertiliteta dva regiona, zasnovane na metodu geometrijske sredine između funkcionalnih serija, potvrdile su očekivanu izrazitu konvergenciju na duži rok ne dozvoljavajući uticaj nestandardnih opservacija.

Ključne reči: *prognoze stanovništva, funkcionalna analiza, glavne komponente, fertilitet, Srbija*