DIFFERENCES IN ESTIMATING FUTURE FERTILITY OF THE CZECH REPUBLIC USING VARIOUS STATISTICAL MODELS

Filip Hon – Jitka Langhamrová

Abstract

The fertility of women in the Czech Republic underwent significant changes after the change of political regime in the 1990s. First, there was a decline in fertility in the 1990s and in the turn of the millennium. Subsequently, in recent years fertility has started to increase again a little. Women, however, acquire children at a higher age than in the past. The aim of the paper is to carry out projections of age-specific fertility rates using various statistical models to indicate possible future development of these rates. Thanks to this projection it is possible to observe whether current trends would manifest themselves, for example, in further postponement of fertility to older age or in further improvement of the overall fertility level. Both parametric and non-parametric models can be used for these statistical projections. The results show that projections vary greatly according to by which model age - specific fertility rates are estimated, depending on how flexible each model is and which recent changes it takes most into account. Thus, in the case of the Czech Republic and its irregular development of age-specific fertility rates, these models may rather serve only as a basis for demographers in compiling population forecasts.

Key words: fertility, statistical modeling, projection, age-specific fertility rates

JEL Code: J11, J13

Introduction

As a result of the second demographic transition, which began in the 1990s after the change of political regime, fertility began to decline. At the same time, women postpone reproduction to older age (Van de Kaa, 2002).

Of course, bigger attention should be paid to its analysis in response to reduced fertility. The aim of the article is to analyse fertility in the Czech Republic from the perspective of modern statistical approaches using Rstudio, which are currently increasingly used abroad and there is a paucity of literature about using similar models on datasets of Czech Republic. One such approach is the Quadratic Spline model (Schmertmann, 2003). Next is the Functional demographic model (Hyndman, Ullah, 2007). Both models introduced in this paper has been calculated on the latest available data. Based on this analysis, both models can then be used to estimate possible future development.

Functional demographic model is based on the Lee-Carter model, which has been already slightly more publicly appreciated in the Czech Republic. For example, there can be found a literature about using Lee-Carter model for a projection purposes and comparison of results with Eurostat's expert estimates (Šimpach, 2015).

The paper first describes briefly the methodology. Subsequently, it describes the development of specific fertility rates after the Velvet Revolution, which is followed by projection in the last part of this paper.

1 Methodology

In this chapter the basic terms used in the analytical part will be described in three subchapters. The first subchapter gives a brief explanation of what specific fertility rates are and why they are used. The second subchapter deals with the Quadratic Spline model and the last subchapter is focused on the calculation of the Functional demographic model.

1.1 Calculation of age specific fertility rates and total fertility rate

Fertility of women varies based on age. For a country at the level of demographic development such as the Czech Republic, for example, a woman of 30 years of age will logically have a higher fertility rate than a woman of 15 years of age and so on.

$$f_{t,x} = \frac{N_{t,x}^{live}}{\overline{F}_{t,x}} \qquad (1)$$

where

$N_{t,x}^{live}$ – number of live births to women in age group *x* in a given year *t*,

$\overline{F}_{t,x}$ – number of women in age group *x* in a given year *t*.

The calculation of age specific fertility rates is done only for women in the reproductive period, since for others it would logically be zero. The fact that fertility varies according to age is important for example for population projections and in other demographic analyses (Roubíček, 1997).

Age specific fertility rates can be used to calculate an important and frequently used indicator of total fertility rate,

$$TFR = \sum_{x=15}^{49} f_x, \qquad (2)$$

which gives a hypothetical total number of children, that would be born to a woman during her reproductive period assuming no change in fertility levels and zero mortality during that period (Langramrová, Šimpach, 2013).

1.2 Quadratic Spline model

This subchapter deals with a brief (see Schmertmann, 2003) methodology of the Quadratic Spline model, which is based on the fact that the curve of age specific fertility rates for all ages of a woman's reproductive period is possible to approximate only on the basis of several values. These values are called parameters and, after approximation by means of regression splines, determine the remaining values (Schmertmann, 2003).

These model parameters are defined as

α, is the lowest age at which the specific fertility rate is not zero,

P, is the age at which the specific fertility rate is highest,

R, is the value at which the specific fertility rate at age P is found,

H, is the youngest age after P, when the age specific fertility rate drops to half the highest value found at the age of P.

It is necessary to add to the parameters that after their initial estimation it is appropriate to further refine them by using iterations minimizing the square error on the whole curve (Schmertmann, 2003).

If there is need to use only parameters mentioned above, you also need to determine knot positions from the index ages to draw a curve using quadratic splines.

1) Knot 0 is found at the earliest age at which the specific fertility rate is not zero,

$$t_0 = \alpha, \qquad (3)$$

where t is the knot designation,

2) knot 1 is between ages α and P,

$$t_1 = (1 - W)\alpha + WP, \qquad (4)$$

where,

$$W = \min\left[0.75; 0.25 + 0.025(P - \alpha)\right] \qquad (5)$$

3) knot 2 is at the age of P when the specific fertility rate is highest,

$$t_2 = P, \qquad (6)$$

4) knot 3 is exactly in the middle between ages *P* and *H*,

$$t_3 = (P + H)/2, \qquad (7)$$

5) knot 4 is exactly in the middle between age *H* and age β when reproduction is stopped and the specific fertility rate is for the first time since the age α minus one zero,

$$t_4 = (H + \beta)/2, \qquad (8)$$

6) reproduction ends at the age of β, ie most often at the age of fifty.

In the case of modeling age specific fertility rates, these knots should, in most cases, provide a meaningful way of carrying out the required analysis using the above mentioned parameters for countries at different stages of demographic development (Schmertmann, 2003).

The model is often used to distribute age specific fertility rates across age when five-year age intervals are used as source data (Schmertmann, 2003). It can also be used for projection of future specific fertility rates, using estimation of time series of development of individual parameters (Lipps, Betz, 2004).

Of course, because of addiction to the position of knots the Quadratic Spline model may not be flexible enough in some cases. For example, in the USA, the shape of the specific fertility rates curve is changing which is the Quadratic Spline model unable to capture. This is because higher specific fertility rates occur at an early age (Peristera, Kostaki, 2007).

On the other hand, at the cost of less flexibility, the model is much better usable for projections than some of similar non-parametric models (De Beer, 2011).

1.3 Functional demographic model

In this section is described, a method for calculating Functional demographic model, for capacity reasons again only in brief (see Hyndman, Ullah, 2007). This model treats data as functional. Estimation and forecast are robust to outliers. Unlike other models it use more than one principal component and use nonparametric smoothing.

It is based on functional time series $\{x_i, y_t(x_i)\}$, $t = 1, \ldots, n$, $i = 1, \ldots, p$,

where

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \qquad (9)$$

where

$y_t(x)$ is the log of the observed fertility rate,

$\{x_1, \ldots, x_p\}$ are single years of age,

$y_t(x)$ are forecasted for $x \in [x_1, x_p]$ and $t = n + 1, \ldots, n + h$.

$\varepsilon_{t,i}$ is an iid standard normal random variable (Hyndman, Ullah, 2007).

Variance for fertility can be calculated using the formula,

where

$$\sigma_t^2 (x) \approx [1000 - p_t(x)] M_t^{-1}(x) p_t^{-1}(x), \tag{10}$$

where

$p_t(x)$ is the observed fertility rate for females of age x in year t,

$M_t(x)$ is the female population of age x at 30 June in the year t (Hyndman, Ullah, 2007).

Curve of $f_t(x)$ is for each t and $x \in [x_1, x_p]$ from $\{x_i, y_t(x_i)\}$, $i = 1, 2, \ldots, p$ smoothed by nonparametric smoothing. This curve is decomposed via a basis function,

where

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\, \varphi_k(x) + e_t(x) \tag{11}$$

where

$\mu(x)$ is a measure of location of $f_t(x)$,

$\{\phi_k(x)\}$ is a set of orthonormal basis functions,

$e_t(x) \sim N(0, v(x))$.

For forecast of future age specific fertility rates there is need to forecast at first $\beta_{t,k}$ for $k = 1, \ldots, K$ and $t = n + 1, \ldots, n + h$ (see Hyndman, Ullah, 2007).

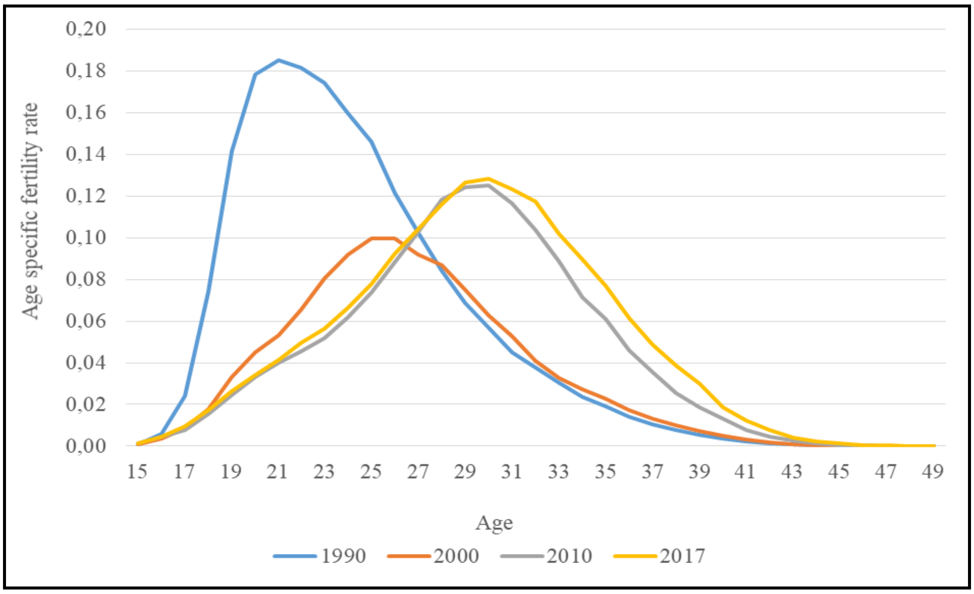
2 Age specific fertility rates and total fertility rate from 1990 and their forecast to 2025

This chapter deals with the application of the statistical models described above for the purpose of projection of future fertility rates. First, for the continuity of projection into the overall context, it is necessary to show the development of age specific fertility rates in the recent past. This development has been shown since 1989 when the Velvet Revolution and the change of political regime was followed by a second demographic transition (Van de Kaa, 2002).

The Quadratic Spline model is modeled (see Schmertmann, 2003, 2005, 2017), forecast with assistance of predictions of the parameters (Lipps, Betz, 2004) are made using the Box-Jenkins methodology and assuming a constant parameter α. The functional demographic model is modeled in RStudio software based on the recommendations defined in the 'demography' package (Hyndman et al., 2019).

For the purposes of the projection, both models have the same length of time series from 1989 to 2017, which is the year of the Velvet Revolution. Choosing the length of time series for fertility is one of the topics that needs to be further published. While, on the one hand, in the Box-Jenkison methodology, a time series of thirty observations can be considered too short for a stable statistical modeling (O'Donovan, 1983) on the other, it is also possible to achieve more realistic demographic results in the Czech Republic's fertility modeling for example by shortening time series to the period from 1998 (Šimpach, 2015).

Fig. 1: Age specific fertility rates in selected years, Czech Republic



Source: Human Fertility Database, 2020

As shown in Figure 1, age specific fertility rates in the Czech Republic have changed significantly since 1990.

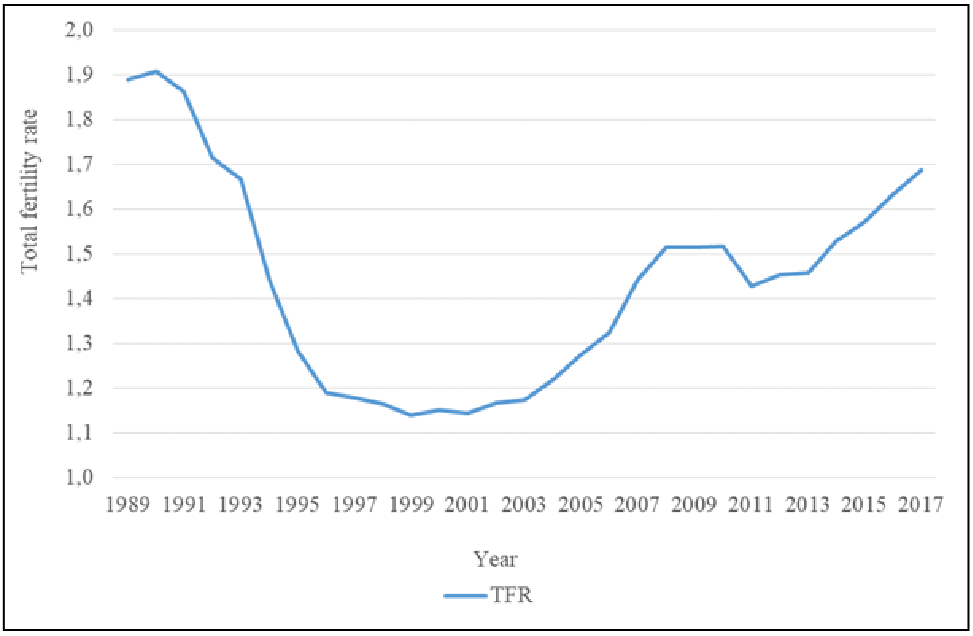
In the early 1990s, the highest age specific fertility rates were just after the age of twenty. Currently, women reach the highest age specific fertility rates after thirty years of age. This shift in age when reproduction is most common took place mainly between years 1990 and 2010. The difference between the values in years 2010 and 2017 is no longer so pronounced. From year 2010 to the present a slight increase in overall fertility can be observed, especially due to higher fertility among women in higher age.

Postponement of reproduction to a higher age compared to 1990 is due to factors accompanying the second demographic transition, such as greater opportunities for travel, career growth or a higher proportion of people studying at universities and many others (Van de Kaa, 2002).

In the case of the Czech Republic, it is also worth noting the very low fertility at the turn of the millennium and its subsequent recovery, among other things, due to the recuperation of postponed births (Fiala et al., 2018). Nevertheless, recuperation is not sufficient enough and fertility rates are far from that being reached at 1990 levels.

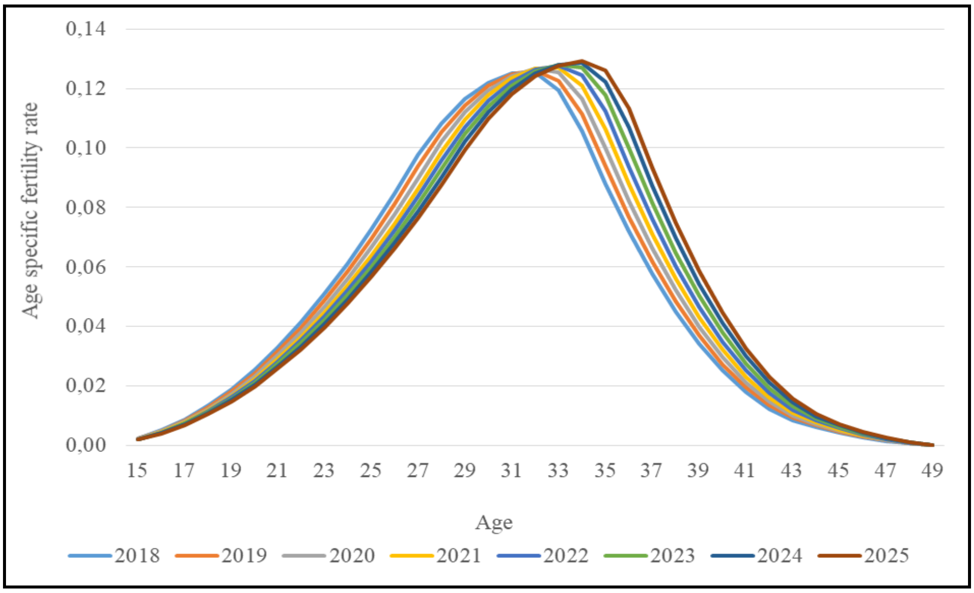
The development of total fertility rate in Figure 2 reflects the facts already mentioned above, such as very low fertility at the turn of the millennium. Furthermore, there is a slight improvement in recent years, but the values are still far below the two children per woman.

Fig. 2: Total fertility rate from 1989 to 2017, Czech Republic



Source: Human Fertility Database, 2020

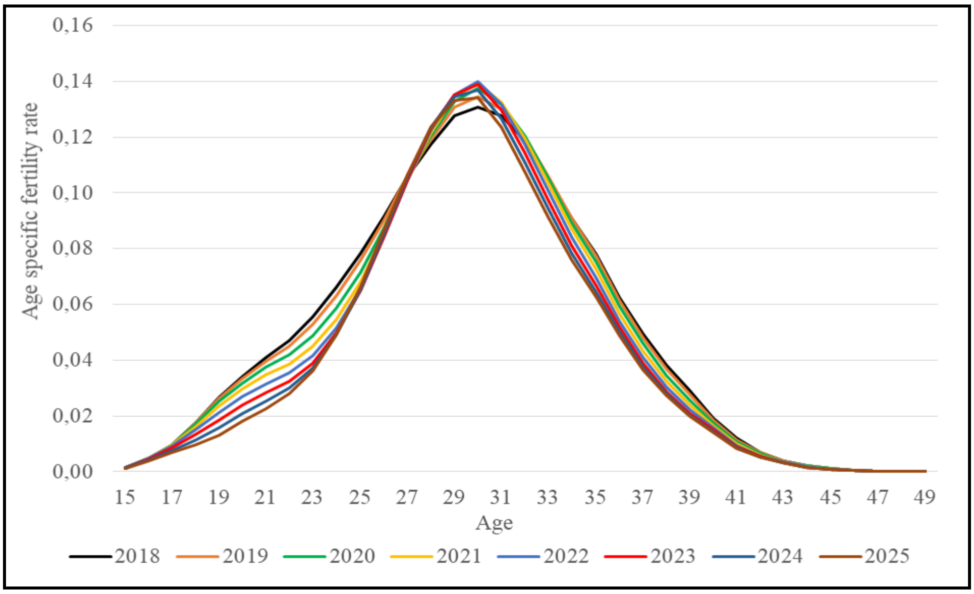
Fig. 3: Age specific fertility rates forecasted by Quadratic Spline model to 2025, Czech Republic



Source: Human Fertility Database, 2020

In the case of the Quadratic Spline model projection shown in Figure 3, there is a further shift in age when specific fertility is highest to an older age. Higher age specific fertility rates are achieved by women of older age and, on the contrary, fertility of women before reaching the age of 30 continues to decrease. As a result, it appears that overall slight improvement in fertility should continue based on model projection

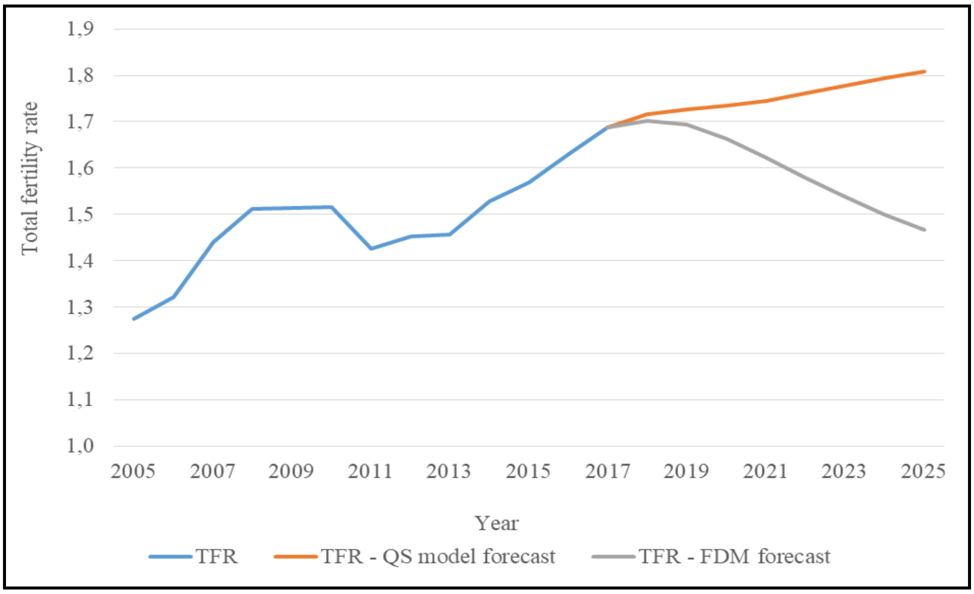
Fig. 4: Age specific fertility rates forecasted by Functional demographic model to 2025, Czech Republic



Source: Human Fertility Database, 2020

Figure 4 shows the results obtained using the Functional demographic model. Like the previous Quadratic Spline model, there is expected a further fertility rates reduction in women around the age of twenty. Otherwise, the results differ significantly from the previous model. Increasing fertility at the age when specific fertility is the highest model expected only in the coming years, then the model expects its decline almost to the current levels. At the same time, the model does not assume an increase in fertility at an older age.

Fig. 5: Total fertility rate and its forecast to 2025, Czech Republic



Source: Human Fertility Database, 2020

Of course, the differences between the models described above must also be reflected in the total fertility rate projection. In Figure 5, in summary characteristics such as total fertility rate, it is clearly seen that, although the same time series since 1989 has been used in both models, each model assumes a completely different future development. The Quadratic Spline model puts more weight on further overall fertility improvement, while the Functional demographic model expects another fluctuation.

Conclusion

This paper contributed to a reduction of a paucity of literature about using modern statistical models for the purpose of modeling age-specific fertility rates on datasets of Czech Republic. Completely unique in terms of new interesting facts, was also a comparison of differences in projections according to individual models. Thanks to the new findings that were gained during the analysis presented in this paper, it was shown that although the same time series had been analysed in both models since 1989, the resulting projection was very different. The Quadratic Spline model, by assuming the modeling of this time series, would assume a further shift in age when specific fertility is highest to an older age. Higher age specific fertility rates should be achieved by women over 30 years of age, while increasingly lower age specific fertility rates should be achieved by younger women. Overall, fertility should increase slightly. On the contrary, the Functional demographic model does not assume an increase in fertility in older age or overall fertility. This model also expects fertility reduction in women around the age of 20. Of course, both models are only projections, which may not be the same as the actual values to be achieved in the future.

Naturally, the issue in the Czech Republic must continue to be addressed. In particular, it is necessary to give a detailed look at the length of the analysed time series and how the choice of this length can influence the projection results in case of uneven fertility development in the Czech Republic. Further research should aim to deepen the methodological knowledge of these methods and to test in the Czech Republic other publically less appreciated possibilities of statistical modeling of fertility.

Acknowledgment

This article was supported by the Grant Agency of the Czech Republic No. GA ČR 19-03984S under the title Economy of Successful Ageing together with the Internal Grant Agency, project No. F4/7/2020 under the title Natalist policy in the Czech Republic.

References

Beer, J. D. (2011). A new relational method for smoothing and projecting age specific fertility rates: TOPALS. *Demographic Research*, 24.

Fiala, T., Langhamrová, J., Pechholdová, M., Ďurček, P., & Šprocha, B. (2018). Population development of Czechia and Slovakia after 1989. *Demografie, revue pro výzkum populačního vývoje*, 3.

Human Fertility Database. (2019). Retrieved from: http://humanfertility.org/

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

Hyndman, R. J., Booth, H., Tickle, L., & Maindonald, J. (2019). Forecasting Mortality, Fertility, Migration and Population Data. Retrieved from: https://cran.r-project.org/web/packages/demography/demography.pdf

Langhamrová, J., & Šimpach, O. (2013). Základy demografie – Materiály ke cvičením. Prague: Oeconomica.

Lipps, O., & Betz, F. (2004). Stochastic Population Projection for Germany- based on the Quadratic Spline approach to modelling age specific fertility rates. *MEA Discussion paper*.

O'Donovan, T. M. (1983). Short term forecasting: an introduction to the Box-Jenkins approach. Chichester: Wiley.

Kostaki, A., & Paraskevi, P. (2007). Modeling fertility in modern populations. *Demographic Research*, 16

Roubíček, V. (1997). Úvod do demografie. Prague: Codex Bohemia.

Schmertmann, C. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic Research*, 9.

Schmertmann, C. (2005). Quadratic spline fits by nonlinear least squares. *Demographic Research*, 12.

Schmertmann, C. (2017). Quadratic Spline Interpolation of Fertility Schedules. Retrieved from: http://fsu.schmert.net/ecp5117/QS_examples.html

Šimpach, O. (2015). Application of the Modern Approach to age specific fertility rates stochastic modelling in the Czech Republic. *Mathematical Methods in Economics*.

Kaa, D. J. van de. (2002). The Idea of a Second Demographic Transition in Industrialized Countries. *Welfare Policy Seminar of the National Institute of Population and Social Security*.

Contact

Filip Hon
University of Economics, Prague
Winston Churchill sq. 1938/4, 130 67 Praha 3-Žižkov
xhonf01@vse.cz

Jitka Langhamrová
University of Economics, Prague
Winston Churchill sq. 1938/4, 130 67 Praha 3-Žižkov
langhamj@vse.cz