Modelling Women's Fertility Intentions and Their Fertility Decisions in the Bayesian Approach, the Case of Poland

Wioletta Grzenda

Institute of Statistics and Demography
Collegium of Economic Analysis
SGH Warsaw School of Economics

Introduction

The research topic discussed in this paper, which is focused on fertility intentions and their realisation is often considered in many demographic studies. The purpose of this work was to draw attention to the research methods used, because their selection can significantly extend the results of these studies. Childbearing intention is defined as a detailed plan for having a child over a certain period of time, which is shaped by the desires, preferences or attitudes of the individual, taking into account the resources and skills of the individual (Brzozowska and Mynarska, 2017). The authors of this work considered the fertility intentions and their realisation by comparing the percentage shares for declarations of parenthood and their realisation for the groups of childless women, with one child and two children or more, as well as in cross-sections by sex. Therefore, the question arises: what impact do other characteristics of the respondents have on these intentions? In searching for the answer to this question, analysing intentions and their realisation using only percentage shares in different cross-sections may be insufficient. In (Sobotka and Testa, 2008), a classic logistic regression model was used to study parental intentions among childless women and men aged 18–39. This approach made it possible to include only two conditions in the study: the "negative" intention to become a parent and the "positive" intention. In this paper, we propose to expand this approach by proposing the use of discrete choice models to model fertility intentions measured on a wider than dichotomous scale.

Examination of the realisation of declared intentions requires having data from a sufficiently long period. Therefore, the realisation of childbearing intentions in short term is most often examined. However, under such conditions, the construction of econometric models is problematic because of the insufficient number of observations for individual cross-sections due to the number of children possessed so far and the declared intentions. Therefore, in this study we decided to analyse the procreative decisions of all women aged 18 to 49 years.

It is important to consider their age in women's fertility studies. According to Bratti and Tatsiramos (2012), the negative relationship between the age of a woman at the time of the birth of the first child and total fertility indicates that delaying motherhood can be an important factor determining fertility decline. This means that time should be a key aspect in examining fertility decisions. Therefore, in this study to analyse this phenomenon, the Cox model was proposed (Cox 1972; 1975; Cox, Oakes, 1984). The basic version of this model in the study of fertility decisions is not sufficient if we want to include in the study both constant features and features that change over time. To model such features an appropriate analysis tool can be the extended Cox model with time-independent and time-dependent variables.

In this work, all models were considered in Bayesian approach. Bayesian methods often are applied when the estimation of the models by the classical methods based on maximizing likelihood function can be problematic due to the difficulties associated with finding the global maximum. Moreover, in Bayesian approach, model parameters are random variables. Therefore, a frequent way to compare parameters of models and evaluate the differences between them is to compare the posterior expected values as well as to use graphic methods. In this paper, we have used a method that allows the evaluation of whether there are statistically significant differences between the impact of investigated factors on the transition from the state of being childless to the state of having a first child and the impact of these factors on the transition from the state of having one child to the state of having two children. The use of such analytical methods allowed identification and comparison of the impact of the investigated factors on decisions regarding the birth of a child in the case of childless women and women with one child. Selected results of this and other analyses are described in this work.

Research methods

In our study, we used two methods: the multinomial logit model (MNL) and the extension of the Cox model - the hazard model with time-independent and time-dependent variables. For analysing fertility intentions multinomial models were selected, because we considered the dependent variable, which takes more than two values. Let us suppose that the *i*-th unit $(i = 1, \ldots, n)$ has to select one of J unordered categories. In the multinomial logit model, random components ε_{ij} $(j = 1, \ldots, J)$ are independent and have the same Gumbel distribution (the type I extreme-value distribution). Moreover, the unobserved stochastic parts of the utility are uncorrelated for all alternatives and have the same variance (McFadden, 1974).

Then, the probability of observing the choice by the i-th unit $(i = 1, \dots, n)$ of j-th category $(j = 1, \dots, J)$ is given by the formula:

$$p_{ij} = \frac{exp(\mathbf{x'}_{ij}\boldsymbol{\beta})}{\sum_{k=1}^{J} exp(\mathbf{x'}_{ik}\boldsymbol{\beta})}, i = 1, \dots, n, j = 1, \dots, J,$$

where $\mathbf{x}$ denotes the vector of explanatory variables and $\boldsymbol{\beta}$ is the vector of parameters. Assuming that the options from which individuals make their choices are mutually exclusive, the formula for the likelihood function for this model has the form:

$$L(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x}) = \prod_{j=1}^{J} (p_{ij})^{d_{ij}},$$

where d_{ij} is an indicator function that assumes a value of 1, when the i-th unit $(i = 1, \dots, n)$ selected j-th category $(j = 1, \dots, J)$, and 0 otherwise. The likelihood function is the basis for estimating the models in both classical approach and Bayesian approach considered in this paper. The model can be used *inter alia* to analyse the fertility intentions, because variables considered in the model only describe the characteristics of the respondents and not the categories selected.

The fertility decisions of women can be conditioned by numerous factors and constantly change over time. Therefore, the proper research tools for their analysis are survival time models. To assess the impact of potential factors on the transition from the state of being childless to the state of having a first child and on the transition from one child to the second child, the Cox model (Cox 1972; 1975) has been used. Let $\mathbf{x} = [x_1, \dots, x_k]^T$ denote a vector of independent variables, and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_k]$ represent a vector of regression coefficients. Then the hazard rate for the Cox model of proportional hazards is usually written as:

$$h(t|x) = h_0(t)\exp(\mathbf{x}\boldsymbol{\beta}),$$

where $h_0(t)$ denotes the baseline hazard rate. The generalization of the Cox model of proportional hazards is a model containing both time-independent and time-dependent variables. Then, the vector of independent variables can be presented in the form $\mathbf{x}(t) = [x_1, \dots, x_{k_1}, x_1(t), \dots, x_{k_2}(t)]^T$, where x_i for $i = 1,2, \dots, k_1$ denotes the time-independent variable, and x_j for $j = 1,2, \dots, k_2$ denotes the time-dependent variable. The hazard rate for a model with time-independent and time-dependent variables is defined as:

$$h(t|\mathbf{x}(t)) = h_0(t)\exp\left(\sum_{i=1}^{k_1} \beta_i x_i + \sum_{j=1}^{k_2} \delta_j x_j(t)\right).$$

To estimate unknown parameters in the Cox model, the partial likelihood function is used, which for the model considered in this work has the form:

$$L_P(\boldsymbol{\beta},\boldsymbol{\delta};\mathbf{t},\boldsymbol{\nu})=\prod_{i=1}^{n}\left(\frac{exp(\mathbf{x}_i\boldsymbol{\beta}+\mathbf{x}_i(t)\boldsymbol{\delta})}{\sum_{j\in R(t_i)}exp(\mathbf{x}_j\boldsymbol{\beta}+\mathbf{x}_i(t)\boldsymbol{\delta})}\right)^{\nu_i},$$

where $\mathbf{t}=[t_1,\dots,t_n]^T$ is the survival time vector, $\boldsymbol{\nu}=[\nu_1,\dots,\nu_n]^T$ is the vector of censoring variables and $R(t_i)$, $i=1,2,\dots,n$ denote risk set at the moment t_i.

In this paper, Bayesian methods have been used to perform the estimation of these models (Ibrahim et al., 2001). In this approach, the parameters of a model are random variables, so these methods offer a much richer foundation for statistical inference and interpretation of the outcome. In the Bayesian approach, statistical inference about the parameter vector $\boldsymbol{\beta}=[\beta_1,\dots,\beta_k]$ is based on the posterior distribution $p(\boldsymbol{\beta}|D)$, which is given by the formula:

$$p(\boldsymbol{\beta}|D)\propto p(D|\boldsymbol{\beta})p(\boldsymbol{\beta})$$

where D denotes observed data. For the multinomial logit model, this formula may be written in an equivalent form:

$$p(\boldsymbol{\beta}|\mathbf{x},\mathbf{y})\propto L(\boldsymbol{\beta};\mathbf{y},\mathbf{x})p(\boldsymbol{\beta}),$$

while for the Cox model, it can be presented in the form:

$$p(\boldsymbol{\beta},\boldsymbol{\delta}|\mathbf{t},\boldsymbol{\nu})\propto L_P(\boldsymbol{\beta},\boldsymbol{\delta};\mathbf{t},\boldsymbol{\nu})p(\boldsymbol{\beta})p(\boldsymbol{\delta}).$$

In this paper, Markov Chain Monte Carlo (MCMC) methods were used to determine a posterior distributions. In particular, the Metropolis algorithm (Gelman et al.,2000) and the Gamerman algorithm (Gamerman, 1997) were used. The MCMC methods are general simulation methods used to estimate the vector of parameters in the Bayesian model (Congdon, 2006).

Reference Data

The data set for this analysis comes from the Generations and Gender Survey (GGS) for Poland. The data used in this paper come from the first and second rounds, conducted in 2010-2011 and 2014-2015, respectively. To model intentions and fertility decisions, women were selected at the time of the second survey at the age of 18-49. 2,879 observations were obtained in this way. For women who participated in the first and second surveys, information from both surveys was combined. There were 2,422 women selected in this way.

At the first stage of our study, the fertility preferences of women surveyed in the first survey were analysed on the basis of the following question present in the survey "Do you intend to have a child (or another child) in the next 3 years?". To this question, 11.19% of women answered definitely yes, 14.82% probably yes, 23.53% probably no, 50.45% definitely no. Then the same question was considered, but separately in a group of women without children (656 observations) and already having at least one child (1766 observations). For the former group of women, the following distribution of answers was obtained: 21.34% of women answered definitely yes, 25.15% probably yes, 25.61% probably no, 27.90% definitely no. However, for women who already had a child, 7.42% of women answered definitely yes, 10.99% probably yes, 22.76% probably no, 58.83% definitely no. In the next stage, explanatory variables were constructed, which were then included in modelling. Table 1 presents selected features. For the features changing over time, their values are given at the time of the first survey. In addition, for the needs of modelling, the last two levels of the education variable have been combined into one.

Table 1 Sample characteristics of women used to model fertility intentions, data from GGS, Poland

Variable	Categories	Labels of levels	Percent
Age group	From 18 to 29 years old	1	18.99
	From 30 to 39 years old	2	40.67
	40 years and more	3	40.34
Child	No	0	27.09
	Yes	1	72.91
Marital status	Married	1	32.99
	Unmarried, widowed, separated or divorced (single)	0	67.01
Education	Higher	1	29.52
	Post-secondary and secondary professional	2	35.05
	Secondary general	3	10.12
	Basic vocational	4	21.59
	Primary school	5	3.72
Employment	Employed in the public sector	1	22.91
	Employed in the private sector	2	32.62
	Performing a different type of work	3	12.01
	Not working	4	32.45
Place of residence	City of 100 thousand residents and more	1	39.88
	City under 100 thousand residents	2	20.60
	Rural areas	3	39.51
Region of Poland	Central (łódzkie, mazowieckie)	1	15.61
	Southwest (dolnośląskie, opolskie)	2	10.40
	South (małopolskie, śląskie)	3	17.13
	Northwest (wielkopolskie, zachodniopomorskie, lubuskie)	4	18.62
	North (kujawsko-pomorskie, warmińsko-mazurskie, pomorskie)	5	14.49
	East (lubelskie, podkarpackie, świętokrzyskie, podlaskie)	6	23.74
Place of residence in childhood	City of 100 thousand residents and more	1	22.25
	City under 100 thousand residents	2	34.31
	Rural areas	3	43.44
Mother's education	Primary school	0	32.20
	More than primary education	1	67.80
Father's education	Primary school	0	30.84
	More than primary education	1	69.16
Parent relationship	The parents lived together	0	79.11
	No	1	20.89

Current material situation of a household in the respondent's assessment	Good	1	13.91
	Average	2	62.59
	Poor or lack of answer	3	23.49

Modelling of fertility intention

Bayesian multinomial logit model was used to model fertility intentions. The dependent variable was defined on the basis of a question from the survey: "Do you intend to have a child (or another child) in the next 3 years?". Taking into account the set of possible answers to this question, this variable had four categories. As the reference category, the category most often chosen by the respondents was chosen, namely "definitely not". Taking into account large sample size, all considered models were estimated using non-informative a prior distributions. Namely, for the parameter vector $\boldsymbol{\beta}$, all models assumed a prior normal distributions with the mean equal to 0 and variance equal to 100. Gamerman's algorithm was used for sampling (Gamerman, 1997). The convergence of the generated chains was assessed using the Geweke test.

As a result of the analysis, it was found that the following characteristics, apart from having a child, have a major impact on fertility intentions: age group, marital status, education, employment and the relationship of parents. For all interpretations considered, the reference level for the intention was "definitely not". Women without children were more than 12 times more likely to say "definitely yes", almost four times more likely to say "probably yes" and twice as likely to say "probably not" in each of the cases considered compared to other women. It was received that women aged 18 to 29 had more than 9 times greater chances to indicate the answer "definitely yes" compared to women from the oldest age group of 40 years and older, while women aged 30 to 39 years were more than 13 times more likely to indicate this answer. On the other hand, the chances of indicating the answer "probably yes" and "probably not" for both considered groups of women were similar and were about 8 and 3 times higher, respectively, compared to women from the oldest age group. As one would expect, single women were less likely to answer "definitely yes" by about 80% compared to married women.

Interesting results were obtained for the variable describing the level of education. Only the impact of higher education turned out to be significant, with women having such education having more than twice as high chances to indicate "definitely yes" and "probably yes" in comparison to the least educated women. In addition, the employment description variable had a significant impact on fertility intentions. The best chances to indicate "definitely yes" were obtained for working women compared to unemployed ones, with no major differences between the public and private sectors. Interestingly, in the case of the answer "probably yes" the highest chances were obtained for women employed in the private sector, each time compared to unemployed women. An important determinant also turned out to be the parents' relationship, the women whose parents lived in a relationship had about 50% greater chance of indicating the answer "definitely yes" and "probably yes". In the case of other variables, only selected levels had a significant impact on the intentions tested, except for variables describing: education of parents, place of residence and material situation of the household of the respondent, for whom no statistically significant impact on the declared intentions regarding fertility was obtained.

Verification of fertility intentions

In the next stage of the study, the fertility intentions of women with actual birth events until the second survey were verified. Among women with at least one child a high percentage of "definitely not" answers to the question "Do you intend to have a child (or another child) within the next 3 years?" was observed. 58.83% of women gave such an answer, while among childless women this percentage was almost by half lower. Analysing birth events, it was found that 22.56% of childless women gave birth to their first child, while among women with one child only 8.55% gave birth to a second child. These results roughly match the declarations of "definitely yes", the question arises as to the women who indicated the variant "probably yes" and other variants. Therefore, a more detailed analysis of intentions and their realisation was made. It was received that among all women who expressed a strong intention to give birth to a child, i.e. they chose the "definitely yes" option, only 24.35% gave birth to their first child, and only 18.82% gave birth to a second child between the surveys analysed. Among women who chose the "probably yes" option, 11.42% gave birth to their first child and 12.53% gave birth to their second child. Among women who chose the "probably not" option, 4.39% gave birth to their first child and 3.51% gave birth to their second child. Among women who chose the "definitely not" option, 1.31% gave birth to their first child and 2.86% gave birth to their second child.

Based on the analyses carried out, one can notice large discrepancies between reproduction plans and their realisations. Therefore, the research question arises as to what factors and to what extent influence the decision to give birth to the first child and how their influence changes in the case of the decision to give birth to a second child.

Modelling of the waiting times to first and second births

In the analysis of fertility decisions, the moment of the birth of the first and subsequent children is of great importance, so in these studies survival models are an appropriate research tool. The Cox model was applied to two groups of women: those who were exposed to the risk of the birth of a first child (2,879 observations) and those who were exposed to the risk of the birth of a second child (1,015 observations). Therefore, we estimated two Cox models in the Bayesian approach, namely the first model for the transition from the state of being childless to the state of having a first child, and the second model for the transition from having one child to having a second child. Based on the preliminary selection of variables, the following variables were selected: age group, education, employment, place of residence, parent relationship, current material situation of a household in the respondent's assessment and father's education.

The values of most variables were determined either at the time of the second survey, in the case of childless women, or at the time of the birth of first child, but the variables: education and employment were included in the model as variables that change over time. Next, for all variables the assumption of their proportional hazards was verified. For those variables for which this assumption was not met, time-dependent index variables were created on the basis of graphs of hazard functions and survival functions obtained using parametric methods. The 96th month (from 18 years of age) for the variable describing age group and the 120th month for the variable describing employment were chosen as partition points. Therefore, it was assumed that number 1 added to the name of the variable level denotes the time period up until this partition point, and number 2 after this point.

Due to large sample size, non-informative independent normal prior distributions were used with a mean equal to 0 and a variance of 10^6 for all regression coefficients of these models. Moreover, in both models, the number of burn-in samples was assumed to be 2,000 and posterior samples equal to 10,000. These settings were applied in order to minimize the effect of initial values on posterior inference. The highest posterior density (HPD) intervals for all parameters were determined for α=0.05. In this paper, the adaptive rejection Metropolis Sampling Algorithm (ARMS) was used for estimations.

On the basis of the posterior values of the hazard ratio determined for women who were exposed to the risk of the birth of their first child, it can be concluded, among others, that the impact of the variable describing the age of the respondents varies depending on the time. In the first investigated 96 months women aged 18 to 29 years old had a 38.08% lower hazard of childbirth, compared to women from the age group 40 years and more, while in the following months the hazard was lower by 86.02%. Women 30 to 39 years old in the first 96 months had a 59.49% greater hazard of childbirth, while in the following months the hazard was lower by 37.57% compared to the women from oldest age group. The level of education of women did not have much influence on their decision to give birth to their first child, women without higher education had only a slightly lower hazard of childbirth than women with higher education. Interesting results were obtained for the variable describing employment. In the first 120 months, women employed in the public sector had a 89.08% greater hazard of childbirth than unemployed women, while in the following months the hazard was lower by 30.29%. Women employed in the private sector in the first 120 months had a 42.88% higher hazard of childbirth compared to unemployed women, while in the following months the hazard was lower by 28.37%. Women performing other type of work than employment in the public or private sector, such as self-employment had in the first 120 months a 135.66% greater hazard of childbirth than unemployed women; in the later period the hazard was lower by 20.24% compared to unemployed women. In addition, it was found that women living in cities had a lower risk of having a child than women living in rural areas. Having parents in relationship and father's basic education also had a positive effect on the birth of a child.

A similar analysis was carried out for women who were exposed to the risk of the birth of a second child. As an example, the results for variable employment have been presented. In the first 120 months, women employed in the public sector had a 48.94% greater hazard of having a second child than unemployed women, while in the following months the hazard was lower by 79.68% compared to unemployed women. Women employed in the private sector in the first 120 months had a 24.61% greater hazard of having a second child than unemployed women, while in the following months the hazard was lower by 71.25% compared to unemployed women. Women performing other type of work than employment in the public or private sector in the first 120 months had a 114.6% greater hazard of having a second child than unemployed women, while in the following months the hazard was lower by 80.27% compared to unemployed women.

Therefore, the question arises to what extent the impact of the examined traits differentiates the decision to give birth to the first child compared to the decision to give birth to a second child. Before proceeding to a more detailed analysis, the hypothesis about the equality of the posterior expected values vectors was verified. It was received that at any significance level it should be rejected. Therefore, there are statistically significant differences in the impact of the examined factors on the decision to give birth to the first and second child.

Another approach may be to compare the posterior density function for investigated parameters. Sample posterior density functions for the corresponding coefficients are presented in Fig. 1 and Fig. 2.

Fig. 1 The posterior densities for the characteristic *employed in the public sector 1* for first births and second births; data from GGS, Poland

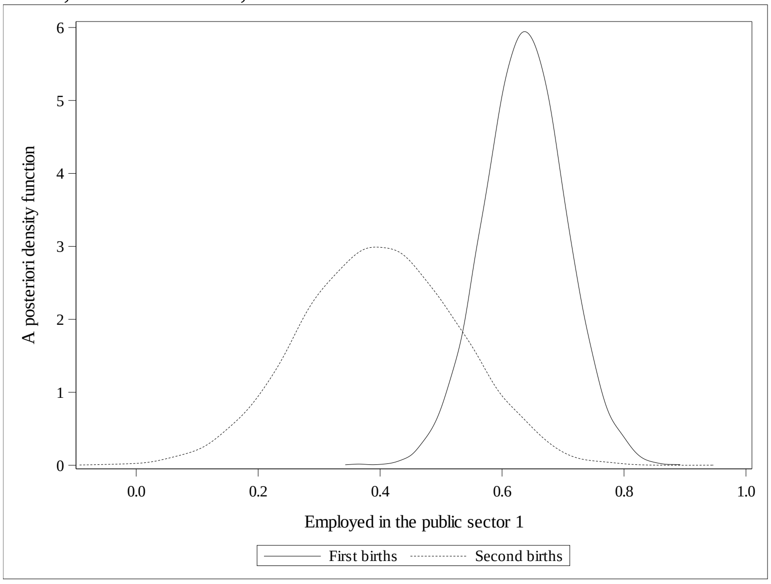
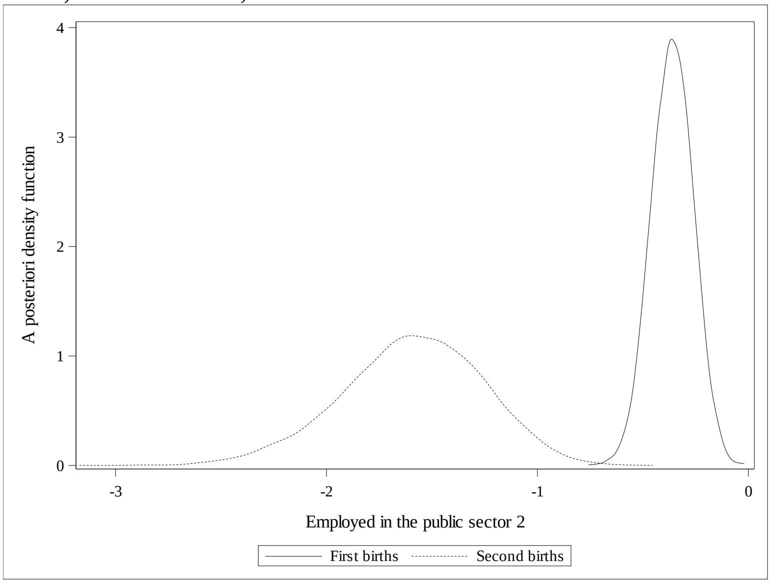


Fig. 2 The posterior densities for the characteristic *employed in the public sector 2* for first births and second births; data from GGS, Poland



In the next stage of research, in line with (Kruschke, 2015), we calculated the difference between relevant elements of the vectors of parameters for each step of the MCMC chains. In this way, the posterior distribution of credible values of the differences between these vectors was obtained. Next, HPD intervals were calculated for these differences. Whether the posterior means of the obtained differences between relevant elements of the considered parameter vectors are significantly different from zero was established on the basis of HPD intervals (Bolstad, 2007). For variables: *age group*11, *employment*12, *employment*22, *employment*32 and *parent relationship*, the posterior expected values of these differences were significantly different from zero. It can be interpreted that the influence of these factors on the time of the birth of first and second child was different.

Summary

In the face of the change in the fertility behaviours observed mostly in Central and Eastern European countries (Eurostat, 2018) it is important to answer the question: what factors influence the time of making a decision to give birth to the first child and what causes that fewer and fewer women decide to have a second child?

In this study, a comprehensive analysis was carried out involving the study of female fertility intentions and their fertility decisions. Such an analysis required the use of advanced econometric tools. In this work, discrete choice models and survival models in Bayesian approach were used. As a result of our analysis, it was found that the greatest parental intent was indicated by women from the second age group i.e. women aged 30 to 39. This phenomenon, according to the findings of (Bratti and Tatsiramos, 2012), may be the reason of such a low fertility as the fertility observed in Poland. In addition, our study shows that major parenting intentions are indicated by women with higher education and working. It confirms the observed changes in the impact of these factors on fertility in recent years (Matysiak and Vignoli, 2013). Interestingly, the material situation of the respondent's household did not influence neither the intention nor the actual fertility decision. An important element of the study was to consider in survival models both the variability of the considered features over time and their varying impact on the phenomenon under consideration. For the employment variable, the key moment was the age of 30. At the beginning of their careers, women were more likely to give birth to both their first child and their second child - than later. To sum up, based on our analysis, it can be concluded that there are statistically significant differences between the impact of the examined factors on the birth of the first and second child. In particular, this applies to such features as age group, employment and parent relationship. It was obtained *inter alia* that having employment has a greater impact on the decision to have a first child than a second child.

References

Billari, F. C., Kohler, H. P., Andersson, G., & Lundström, H. (2007). Approaching the limit: Long-term trends in late and very late fertility. *Population and development review*, 149-170.

Bolstad, W.M. (2007). Introduction to Bayesian statistics. Wiley & Sons, New York

Bratti, M., & Tatsiramos, K. (2012). The effect of delaying motherhood on the second childbirth in Europe. Journal of Population Economics, 25(1), 291-321.

Brzozowska, Z., & Mynarska, M. (2017). Fertility intentions and their realisation: Insights from the Polish generations and gender survey (No. 16/2017). Vienna Institute of Demography Working Papers.

Congdon, P. (2006). Bayesian Statistical Modelling, Wiley & Sons, New York.

Cox, D.R. (1972)Regression models and life tables (with discussion). Journal of the Royal Statistical Society. 34, 187-220.

Cox, D. R. (1975). Partial likelihood. Biometrika, 62(2), 269-276.

Cox, D. R., & Oakes, D. (1984). Proportional hazards model. Analysis of survival data, 91-111.

Enderlein, G. (1987). Cox, DR; Oakes, D.: Analysis of Survival Data. Chapman and Hall, London–New York 1984, 201 S.,£ 12,–. Biometrical Journal, 29(1), 114-114.

Eurostat: Fertility statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Fertility_statistics (2018, June 7).

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing, 7(1), 57-68.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. Chapman and Hall/CRC.

Generations and Gender Programme (GGP). http://www.ggp-i.org/

Ibrahim J.G., Chen M.H., Sinha D. (2001). Bayesian Survival Analysis, Springer-Verlag, New York.

Kruschke, J.K. (2015). Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan. Academic Press / Elsevier, Waltham

Matysiak, A., & Vignoli, D. (2010). Employment around first birth in two adverse institutional settings: Evidence from Italy and Poland. ZfF–Zeitschrift für Familienforschung/Journal of Family Research, 22(3).

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior in: P. Zarembka (ed.) Frontiers in Econometrics. New York: Academic Press,105–142.

Sobotka, T., & Testa, M. R. (2008). Attitudes and intentions toward childlessness in Europe. In People, population change and policies (pp. 177-211). Springer, Dordrecht.