https://doi.org/10.26219/heal.aueb.5600

ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

MODELING THE AGE SPECIFIC FERTILITY PATTERN

By

Dimitrios G. Kosmas

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
2005

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Modeling the age specific fertility pattern

By

Dimitrios Georgiou Kosmas

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
May 2005

https://doi.org/10.26219/heal.aueb.5600

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΗΣ ΕΙΔΙΚΗΣ ΚΑΤΑ ΗΛΙΚΙΑ ΓΕΝΝΗΤΙΚΟΤΗΤΑΣ

Δημήτριος Γεωργ. Κοσμάς

ΔΙΑΤΡΙΒΗ
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Μάιος 2005

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

MODELING THE AGE SPECIFIC FERTILITY PATTERN

Dimitrios G. Kosmas

Approved by the Graduate Committee

A. Kostaki	S. Psarakis	I. Papageorgiou
Assistant Professor	Assistant Professor	Lecturer
Thesis Supervisor	Members of the Committee	

Athens, July 2005

Michael Zazanis, Professor
Director of the Graduate Program

https://doi.org/10.26219/heal.aueb.5600

DEDICATION

To my father

ACKNOWLEDGEMENTS

I would like to express my gratitude to all who have supported me in all my decisions in life, regardless of their own opinion about them. I would like also to thank my supervisor Anastasia Kostaki for all the help she offered in the completion of this work.

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

VITA

I was born in 1981 in Athens Greece. I entered the department of mathematics in the University of Athens in 1999 and I received my degree in 2003. The same year I was accepted in the Master's Program in statistics in the department of statistics, Athens University of Economics and Business.

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

ABSTRACT

Dimitrios Kosmas

MODELLING THE AGE SPECIFIC FERTILITY PATTERN

May 2005

The age specific fertility pattern exhibits a typical pattern, common in all human populations. Many formulae have been proposed in order to describe this pattern. In recent years, a hump has been presented in the early years of the age specific fertility curve creating new facts and giving a new shape in the fertility pattern. This thesis provides a presentation an evaluation and a comparison of the various parametric models proposed. In order to evaluate and compare their adequacy in describing the fertility pattern, we fit them to a series of empirical data sets.

ΠΕΡΙΛΗΨΗ

Δημήτριος Κοσμάς

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΗΣ ΕΙΔΙΚΗΣ ΚΑΤΑ ΗΛΙΚΙΑ ΓΕΝΝΗΤΙΚΟΤΗΤΑΣ

Μάιος 2005

Οι ειδικοί κατά ηλικία δείκτες γεννητικότητας εμφανίζουν μια τυπική μορφή, κοινή σε όλους τους ανθρώπινους πληθυσμούς. Για την παραμετρική μοντελοποίηση της μορφής αυτής έχουν προταθεί διάφορα μοντέλα. Πρόσφατα παρουσιάσθηκε μια διαμόρφωση στο μέρος της καμπύλης που αφορά τις νεαρές ηλικίες, η όποια αλλάζει τα μέχρι τώρα δεδομένα. Η μελέτη αυτή παρουσιάζει τα διάφορα μοντέλα που έχουν προταθεί για την περιγραφή της κατά ηλικία γεννητικότητας και εφαρμόζει τα τρία θεωρούμενα ως καλύτερα σε μία σειρά εμπειρικών δεδομένων με σκοπό την αξιολόγηση και την σύγκρισή τους.

https://doi.org/10.26219/heal.aueb.5600

TABLE OF CONTENTS

https://doi.org/10.26219/heal.aueb.5600

LIST OF TABLES

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

LIST OF FIGURES

https://doi.org/10.26219/heal.aueb.5600

Chapter 1

Introduction

One of the goals of demography is to measure several population figures and processes, such as mortality, migration, nuptiality and fertility. Fertility is the demographic phenomenon related to the reproduction of the population. The typical pattern of births by age of women is one of the subjects of special interest, due to its use in predicting the future size and age composition of a population.

This thesis has the purpose to present and compare all the possible ways that are presented in the bibliography for parametric modeling of the typical pattern of the age specific fertility. These models try to describe the age pattern of fertility by a set of parameters. A review of parametric models for fertility is given in the next chapter. The Chapter 3 gives the statistical background that is needed in order to understand the procedures described at the following chapters.

Each of the following chapters, from four to eight, describes one specific model that has been proposed for this purpose. Chapter 4 describes the efforts made to model fertility by polynomial functions. Chapter 5 describes a widely used model proposed by Coale and Trussell, which took their name. Chapters 6 and 7 describe the efforts to model fertility by the density function of some known distributions. Chapter 8 describes the Gompertz function in modeling fertility. Chapter 9 presents an evaluation and a comparison of most of the previously mentioned models, while Chapter 10 presents the results of applications on empirical data. Finally in Chapter 11 some concluding remarks are presented.

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

Chapter 2

Modeling fertility

2.1 Fertility models

There have been many attempts to model many population processes such as mortality, fertility, nuptiality and migration by parametric functions. By parametric modeling of fertility, we mean to find a parametric model, which fits well the pattern of fertility by age of women in human populations. The fertility rate is defined as the number of births, which occur from women of a given age divided to the total of women of the same age being in the population of a period. We can also examine fertility rates in a cohort, which mean to examine the fertility rates at different ages of women born in a given year.

There are many models proposed for modeling the age pattern of fertility. The first attempt to model fertility was made by Wicksell (1931) with pearsonian type I and III curves. Hadwiger (1940) came next and proposed the density function of inverse Gaussian, which took his name, for this purpose. Polynomial functions were also proposed by Brass in 1960. Mitra (1967) proposed the reduction of the number of parameters for the pearsonian type I curve from five to three. Coalle and Trussel also proposed a model in 1974, which was more empirical than the others were; its parameters were constructed to have a demographical interpretation and also gave an explanation of fertility process.

Several attempts were made to fit the Gompertz function a well known function to demographers from modeling mortality. Gompertz function was first used for this purpose by Wunsch and Martin in 1967. Romaniuk and Tanny used it for prediction with disappointing results so Murphy and Nagnur proposed some modifications in 1972. Heather Booth in 1984 applied some

https://doi.org/10.26219/heal.aueb.5600

ideas presented by Brass in 1974 in transforming the age scale of the Gompertz function to improve its fit.

For a long period of time no new models were proposed because the existed variety was considered adequate to describe the age pattern of fertility. In 1981, a comparative study by Hoem et al. (1981) provided an evaluation of the accuracy of most of the existed models. A hump presented in fertility pattern, shown in data from United Kingdom and Ireland, indicated a transformation of the fertility pattern, which stopped the inactivity in this field and in 1999 a new model, a mixture of two Hadwiger functions, was proposed by Chandola et al. (1999), in order to describe the changed pattern of fertility. Non parametric techniques such as splines were also proposed for this purpose by Mcneil and Trussell (1977).

Two different procedures have been followed to model age specific fertility curve. According to the first one, it is suggested to model fertility by a distribution or a function that has a similar shape with that of the fertility pattern. Afterwards, in order to see how good this selection of a distribution of a function is, the model was tested for its good fit. The demographic interprentation of the parameters of the model, when this procedure is followed is not of first interested but after the model is fitted, it could be explored whether there is any demographic interpretation of the parameters. The constriction of Most of the models presented in this thesis follow this procedure.

According to the second procedure, we try to find the mechanism of age specific fertility and describe it with a model. Therefore, with this procedure we try to find the factors that affect fertility by age, model fertility so as to depend on the suggested factors, produce a parametric function and then check the good fit of the function on a variety of available data. The construction of model that have proposed by Coale and Trussell follows this procedure and tries to interpret fertility through the marital process and the use of birth control in the population. These two procedures try to reach the same goal but give priority to different benefits of modeling fertility.

The models that are presented in the bibliography can be separated into two categories. One category is those models that describe age specific fertility rates and the other that describe the proportional age specific fertility rates. The proportional

https://doi.org/10.26219/heal.aueb.5600

age specific fertility rates show the proportion of the total births in a population, which occur from women of a given age. A model, which describes the proportional age specific fertility rates, can also describe the age specific fertility rates with only the addition of one parameter, which would represent the total fertility level.

2.2 The need of modeling fertility

It has been noticed that age-specific fertility has a typical pattern common in all human populations. Age-specific fertility rates starts from zero at a young age then increase to reaches a peak between the age of twenty and thirty and then decrease to reach zero at the age of menopause with a slower rate than it started. This pattern gives the idea that it could easily be described by a family of functions, which would have the same behavior. Each special case would easily be defined by a set of parameters so one can reproduce the overall pattern of fertility by estimating the values of only a few parameters.

The number of parameters varies according to the proposed model, usually being three or four. The age-specific fertility pattern takes positive values from about age 15 to 45 or 50, that means that without modeling we would need thirty to thirty-five values to present the whole pattern. By modeling fertility, only estimates of the parameters are needed and thus an estimation of the whole pattern of fertility is available. In practice, the available data that are in most cases given in five-year age groups. The countries with available single-year fertility rates are only few and concerned developed countries. For these countries the pattern of fertility is not representing of the general pattern of fertility because developed countries usually have lower fertility that the others and from the data is obvious that the pattern is slightly different for different levels of total fertility.

https://doi.org/10.26219/heal.aueb.5600

2.3 Uses of modeling fertility

Parametric modeling of fertility is a useful tool in many applications in demography. It can be used for obtaining the whole pattern of fertility by only a few number of parameters. Another useful application is in fertility projections where we try to predict the number of births for a population in order to estimate its future size. Graduated rates can also be used instead of the empirical rates for calculating reproduction rates. Another application of parametric modeling is to represent fertility pattern of population where data are scanty of defective. We can represent with a more reliable way the values of fertility by using a model, obtained from patterns of fertility where we have complete and reliable data. Therefore, we manage to eliminate random errors and more accurately represent the general pattern of fertility. It can also been used to attain the single-year age specific fertility rates when only five-year fertility rates are available.

https://doi.org/10.26219/heal.aueb.5600

Chapter 3

Statistical background

3.1 Fitting a model

When we say parametric modeling of a pattern, which depends on x, we mean that we try to find a function depended on a number of parameters to represent this pattern. For this purpose, we use the empirical age-specific data. These data would represent the values of the pattern, let it be y_i, for a number of given values of x. In our case the data would be the age specific fertility rates y for each age x.

So we have (x_i,y_i) for i=1,2,...v where v is the number of ages considered and we try to find a function $f(x;\bar{\theta})$ where $\bar{\theta}$ is a vector with $\bar{\theta}=\{\theta_1,\theta_2,\ldots,\theta_n\}$ where n is the number of parameters. There are models where the value of f(x) is set to be exactly the same as the empirical one at the same age x, such as those produced by interpolation, and others which accept that the data used are distorted so the value of f(x) do not have to be the exact value given by the data y_i but near it. The first models are called reproducing and the second non-reproducing. The choice between the two forms should be based whether our data give us the exact values of the pattern so $y_i=f(x_i)$ or we accept the existence of an error in the data so we seek to find an f(x) with $f(x_i)=y_i+\varepsilon_i$, where ε_i is the error for the i value.

When we assume a $f(x;\bar{\theta})$ in fact we assume a model of the pattern. The next step is to fit the model. When we say to fit a model, we mean to find specific values of the parameters in order to represent better the pattern we want to describe. When we have chosen a $f(x;\bar{\theta})$ to describe a pattern. in reality we have chosen a family of functions that we believe that represent well the pattern but when we have specific data in our hands we want to find the one member of the family, that would describe best, the specific pattern. By estimating the values of the

https://doi.org/10.26219/heal.aueb.5600

parameters, we choose the exact member of the family of functions, which is better to represent the pattern.

The question automatically raised is the way of fitting a proposed model. By this, we mean which procedure we would follow to estimate the parameters of a model. An important role on selecting the method of fitting is what we have assumed about the pattern of the errors. The first part of the model the $f(x;\bar{\theta})$ is called the deterministic part of the model because it gives us an estimation of a specific value for the $f(x_i)$. Non-reproducing models come with the assumption that there is an error in the data. This error is usually assumed to have a stochastic pattern so it is called the stochastic part of the model. A full model would also make assumptions on the stochastic pattern of the error so $y_i = f(x;\bar{\theta}) + \varepsilon_\iota$ where ε_ι follows a distribution usually $\varepsilon_\iota \sim N(\mu,\sigma^2)$.

3.2 Estimation procedures

The most common method of fitting used is ordinary least squares but there are also many other ways of fitting a model. By the ordinary least squares method, it is tried to find the values of the parameters by which the sum of squares of the errors $Q=\sum_{i=1}^{v}\varepsilon_i^2$ is minimized. The values of the parameter that minimize Q could be attained, in general, by an iterative procedure but there are also models where the estimated values can be obtained from a simple set of equations.

Ordinary least squares contrary to weighed least squares give the same weight to all ε_i. Using the method of weighted least squares gives the option to consider the minimizing of certain ε_ι^2 more important than the others, so it is tried to minimize the $Q=\sum_{i=1}^{v} w_i\varepsilon_i^2$ were w_i is the weight it is decided to give to each residual. This method in general is proposed to be used when we assume that the errors $\varepsilon_\iota \sim N(\mu,\sigma_\iota^2)$ which means different σ_ι for each ε_ι.

Another method of fitting is the method of moments, which was wide used in the past. By this method the values of the parameters are estimated by equating the

https://doi.org/10.26219/heal.aueb.5600

moments of the $f(x;\bar{\theta})$ with the moments of the data. By the equations, attained from the use of equations, usually as many moments as the parameters, the values of the parameters are estimated. There are selections of $f(x;\bar{\theta})$ that the estimation of the parameters matches with the estimations attained by other methods but in general this method doesn't give us any mathematical justification why these estimates should be considered as the best. Estimated the values of the parameters, we actually have in hands an estimation of the considered pattern.

A method of fitting which has a statistical justification is the maximum likelihood method. By this method, we estimate the parameters of the model by values, which maximize the likelihood of obtaining the initial data. It is necessary in order to estimate the parameters to accept that ε_t follows a distribution and to calculate the likelihood of obtaining such residuals with respect to $\bar{\theta}$. At the next step, we maximize this likelihood with respect to $\bar{\theta}$ and we take these values of $\bar{\theta}$ as estimations for the parameters.

3.3 Goodness of fit tests

For judging the accuracy of a model in order to describe an empirical pattern we need a criterion. A wide used criterion is R^2 given as the percentage of the total dispersion of the empirical data that is explained by the model. By this, we mean that from the initial amount of dispersion of the empirical data we consider that has remained only the dispersion of the error and the rest is considered explained by the model. So the total dispersion is divided to the part that is due to the fitted values and the part that it due to the errors. The value of the criterion is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{\nu} \varepsilon_i^2}{\sum_{i=1}^{\nu} y_i^2 - \sum_{i=1}^{\nu} y_i}$$

From an other point of view R^2 is the square of the correlation coefficient between the fitted values and the initial values y_i. A consequence of all the above is that R^2 takes values between 0 and 1.

https://doi.org/10.26219/heal.aueb.5600

The choice of distribution for the error is connected with the way of testing the models good fit. If we assume that the error is normally distributed then we have in our hand a distribution for the sum of squares of the errors. This distribution is X^2 with v degrees of freedom, which gives as the option to calculate the likelihood of such an outcome if our assumptions were correct. If we assume a normal distribution for the error then the appropriate method of fitting is reducing the sum of squares and the appropriate test of good fit is by X^2 distribution. An advantage of this test is that we can attain a significance level for the proposed model.

A wide used criterion, which is directly connected with ordinary least squares, is mean square error. This criterion divides the square error to the number of parameters used in the model. It is used for the same data to attempt to fit more than one models to find which one achieves the best fit. Certainly, we can attain a perfect fit if we use as much as v parameters in our model. The fit would improve as we add more parameters so there are statistical test constructed to take into account also the number of parameters. Mean square error takes into account the number of parameters so it is useful for this purpose. The model, which reaches better values of the criterion we have set for a good fit would, is considered better. If we have used ordinary least squares for fitting the models then mean square error is the appropriate criterion to compare them

To all these methods of testing a model's good fit, we can add the graphical ways, by which we look with the eye whether the function proposed represents well the pattern. This way of course is very subjective because two different persons could have different opinions on the models good or poor fit.

https://doi.org/10.26219/heal.aueb.5600

Chapter 4

Polynomial functions

4.1 Introduction

The simplest forms of functions that have been proposed in order to fit the age-specific fertility curve are polynomial functions. Fitting a curve of a polynomial function means that we assume the relation $y=a_1x^n+a_2x^{n-1}+\ldots+a_n+\varepsilon$ between our variables, which are in our case the age-specific fertility rates and age. The number of the parameters of the model is depended on the grade of polynomial we choose to fit.

A polynomial of grade more than 1 either increases or decreases infinitely while x increase or decrease. Conversely, the age pattern of fertility reaches zero for both young and old ages and in advance it reaches the old edge of childbearing period so slowly that the rate of change could also be considered as zero. So in order to use polynomials for this purpose it would be better to use a more specific form of polynomials that would have a similar form with age specific fertility curve.

4.2 The model of Brass

A more specific form of polynomials was proposed by Brass(1960). The form of the polynomials was selected to comply with the properties of the curve of age specific fertility rates mentioned above. The form of the polynomials proposed is $f(x)=x(35-x)^2(a_n+a_{n-1}x+\ldots+a_0x^n)$ which has the property $f(0)=f(35)=f'(35)=0$, where $f'(x)$ is the derivative of the polynomial. In order to fit this polynomial, both reproducing and non-reproducing methods were used. When we take reproducing methods of fitting, we accept that the data are accurate and we try to find a smooth curve, which reproduce them. With non-reproducing methods, the fitted curve does not necessarily have to take the exact values of the data, but just to have the same

https://doi.org/10.26219/heal.aueb.5600

pattern. These methods are usually applied, when we believe that our data are submitted to an error. The difference is that in the first case, we try to fit smoother curves and in the second, we try to reduce the error of the suggested model.

At first non-reproducing techniques were used. A six-grade polynomial of the form mentioned above was selected in order to represent well the age-specific fertility rates curve. This means that four parameters had to be estimated in order to fit the model. The method of moments was used in order to fit thc model so four moments were needed. The model was tested on data from 12 countries, nine of them taken from the United Nations demographic yearbook of 1954. The data for the rest of the countries, which were Ukraine, Slovenia and Australia, were additionally included from other sources because they were available by single year values. The data from the United Nation Yearbook were available only in five-year age group apart from the data from England and Wales, which were available in single years. The shape of the fitted curve was more pointed than the actual curve and more flat at the sides. Table 1 shows all the observed and the graduated values of the model. In the paper is not mentioned any form of statistical testing for the goodness of fit of the model.

As far as reproducing methods are concerned, there is no point in testing the goodness of fit of a model because we assume that there is no error in our data. The only issue is to find a polynomial smooth enough to represent well the age specific fertility rates curve. The same general form of the polynomial was used to attain the same shape of the curve and the numbers of parameters needed was the only subject. The two-parameter model which is a four-grade polynomial was considered smooth enough to represent the values at most of the age range but for the lower ages a three-parameter model which is a five-grade polynomial was selected. The values of the parameters were estimated by the equations attained by demanding from the model to reproduce the values.

The choice between non-reproducing and reproducing methods must be made on the basis of the validity of the data we have in our hands. If we believe that the data give an accurate description of age specific fertility then only smoothing age specific fertility rates would be necessary. The author also suggests that the same method could have better results if different childbearing period is used. This period could be, 13-47 or 17-51 instead of 15-49 with five-year age groups shifted to the new start at the start of the childbearing period. In the paper is suggested with examples that the graduated fertility curves could be used in estimating child woman ratios.

https://doi.org/10.26219/heal.aueb.5600

Country	Year	Age-group in years								
			15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total
Taiwan	1951	O	339	1435	1748	1554	1130	659	173	7038
		G	352	1398	1793	1535	1112	680	168	7038
Puerto Rico	1950	O	498	1398	1302	1000	715	266	59	5238
		G	529	1293	1389	1046	624	296	61	5238
Ukraine	1926-27	O	216	1188	1294	1117	786	412	122	5135
		G	245	1099	1387	1104	737	447	116	5135
Virgin Islands	1950	O	769	1412	1265	861	567	190	27	5091
		G	767	1396	1278	906	511	204	29	5091
Panama	1950	O	62	1121	1057	696	411	131	41	4182
		G	621	1210	1087	700	56	164	35	4182
Jamaica	1951	O	550	1187	1010	676	448	166	29	4066
		G	564	1129	1055	716	392	175	35	4066
Netherlands	1951	O	67	485	904	809	552	220	20	3057
		G	66	504	858	847	550	208	24	3057
U.S.A.	1950	O	402	968	817	506	256	72	6	3027
		G	418	912	860	530	223	72	12	3017
Japan	1952	O	44	654	1030	745	387	115	7	2982
		G	51	658	990	786	382	105	10	2982
Slovenia	1948-52	O	110	647	810	669	466	201	23	2926
		G	132	594	827	713	439	188	33	2926
England and Wales	1951	O	105	631	671	44J	228	67	5	2152
		G	128	573	700	476	202	61	12	2152
Australia	1932-34	O	125	498	605	481	302	123	13	2147
		G	134	475	611	502	292	115	18	2147

Table 4.1 Observed and graduated specific fertility rates. (Births per thousand women over five-year age periods, observed (O) and graduated (G)) Brass(1960)

https://doi.org/10.26219/heal.aueb.5600

Chapter 5

The Coale Trussell model

5.1 Introduction

This model had been proposed from Coale and Trussel in 1974. It is not just a model, which tries to give a function, which fits well to age specific fertility data. It also tries to give an explanation of the fertility procedure. According to the model the f(a), where f(a) is the fertility rate of age a , is the product of two other functions G(a), r(a), which represent the proportion of the woman ever married and the pattern of marital fertility respectively.

$$f(a)=G(a)r(a)$$

Where

$$G(a):=\int_{a_0}^{a} g(x)dx$$

With

$$g(\alpha)=\left(\frac{0.19465}{k}\right)\exp\{\left(\frac{-0.174}{k}\right)(a-a_0-6.06k)-\exp[\left(-\frac{0.2881}{k}\right)(a-a_0-6.06k)]\}$$

By this way, the model regards the fertility procedure as the product of two independent procedures. The first is the marriage process and the second of marital fertility. The model is based on the assumption that we have a population in which the fertility procedure starts right after the marriage procedure, so the age of first marriage is considered to be the start of the reproductive process. This means that the model ignores the extramarital fertility and also divorces and deaths. However, it has been shown that the

https://doi.org/10.26219/heal.aueb.5600

model fits well also in populations where the assumptions are violated and there is a severe amount of births out of marriage.

5.2 Proportion of women ever married

For the calculation of the proportion of women ever married the model uses a model already proposed by Coale and Mcneel (1971) for nuptiality. This model assumes that the function g(a) represents well the age pattern of first marriage of women. In fact, g(a) represents the rate at which marriages occur in a population. From this initial model we use the function G(a) which represents the proportion of women ever married by the age of a. There is no analytical expression for G(a) but it can be calculated with numerical integration.

The parameter α_0 to represents the age at which marriages start and the parameter k is an indicator of how fast the population gets into the married state. The two parameters α_0 and k specify the location and the scale of the distribution of first marriage. Therefore, all the curves, which can be obtained from specific values of the parameters α_0 and k belong to the same location-scale family of distributions.

5.3 Marital fertility

The function r(a) in the product represents the age pattern of marital fertility. For the calculation of r(α) the Coale-Trussell model assumes that there is a natural way of fertility from which human populations departs from due fertility control so it introduces a parameter m, which indicates the level of the fertility control in the population.

For the calculation of natural fertility, the model uses a function introduced by Henri (1961). Denoting natural fertility at age α with the function n(α) which is the fertility rate of age α of a population with no birth

https://doi.org/10.26219/heal.aueb.5600

control, the model speculate that marital fertility r(a) departs from n(a) with the way which is demonstrated in the following formula.

$$\frac{r(a)}{n(a)} = M\exp\{m\nu(\alpha)\}$$

where, M is a scale factor expressing the ratio $\frac{r(a)}{n(a)}$ at an arbitrary value of α. Since only the age pattern of fertility is of our interest, parameter M would not play any role in the final model.

The parameter m shows the level of fertility control in the population. For m=0, we assume that there is no birth control in the population. As birth control we mean the level at which the population tries to control the natural procedure of fertility by any mean, such as condoms, peals, and abortion.

For calculating r(α), the calculation of two functions n(a) and ν(a) is needed. These functions have been calculated from empirical data in two steps. At the first step, six values of five-year age intervals above age 20 have been approximated. At the second step the single year values of the two functions are determined by freehand interpolation as well as, values for ages below 20 based on somewhat arbitrary but common sense principles which the paper does not specify.

The calculation of seven values of ν(α) had been made by Henri (1961). He calculated the arithmetic mean of ten fertility schedules, which he had considered to come from populations where no birth control take place. All schedules, which were based on surveys where age misreporting was widespread, were discarded in order not to distort the pattern of fertility. The reason for which Henri calculated n(a) from ages above 20 is that teenage marital fertility is affected from premarital conceptions with an irregular manner.

An analogues procedure was followed for the estimation of the seven values of ν(α). The data used for this procedure were schedules of marital fertility listed in the United Nations Demographic Yearbook for 1965. Again data, which were considered to be faulty either because of age misreporting or of any other reasons, were left out of the calculation. Finally, forty-three schedules were considered to carry an amount of information about the typical pattern of departure from natural fertility.

The procedure of calculating v(α) was the following. The parameter m was set to be 1 and $v_i(\alpha)$, which is v(α) for the i schedule, was calculated for each schedule separately.

So for m=1

$$v_i(a) = \log\left[\frac{r_i(a)}{Mn(a)}\right]$$

with i=1,2,...,43

Afterwards M was chosen so that $v_i(\alpha)=0$ for the age interval 20-24 v(α) was set to be the arithmetical average of the forty three values of $v_i(\alpha)$ for each of the seven age intervals. The values and the shape of n(α) and v(α) are shown in table 4.1 and figure 5.1 respectively

Age-group	20-24	25-29	30-34	35-39	40-44	45-49
n(α)	0.460	0.431	0.396	0.321	0.167	0.024
v(α)	0.000	-0.316	-0.814	-1.048	-1.424	-1.667

Table 5.1 Values of n(α) and v(α) for five-year age groups

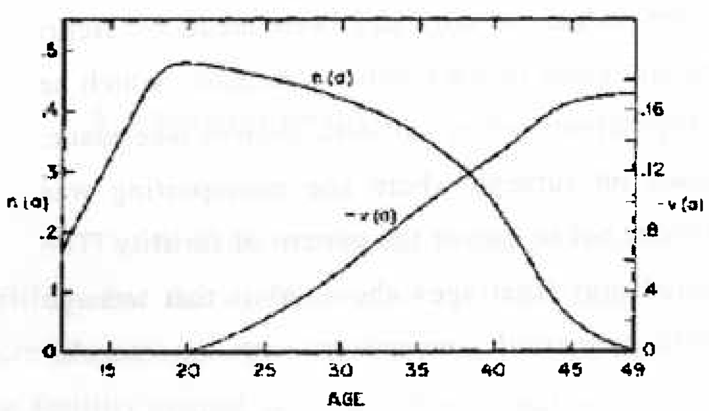

Figure 5.1 Values of n(α) and v(α) for single years Coale,Trussell(1974)

So the model turns to

$$f(a) = G(a)n(a)e^{mv(a)}$$

https://doi.org/10.26219/heal.aueb.5600

where the functions $G(\alpha)$, $n(\alpha)$, $v(\alpha)$ have been specified above.

The model comes with three parameters which are α_0, k, m. It is important to notice that this model has assumes that total fertility equals to one. This means that the three parameters of the model only give information about the way that fertility is distributed with age. If we want information about the total fertility rate, to the three parameters of the model we must add a forth one which specify the total fertility of the population. The shape of the model for various values of its parameters is shown at figure 5.2 and for two different values of m.

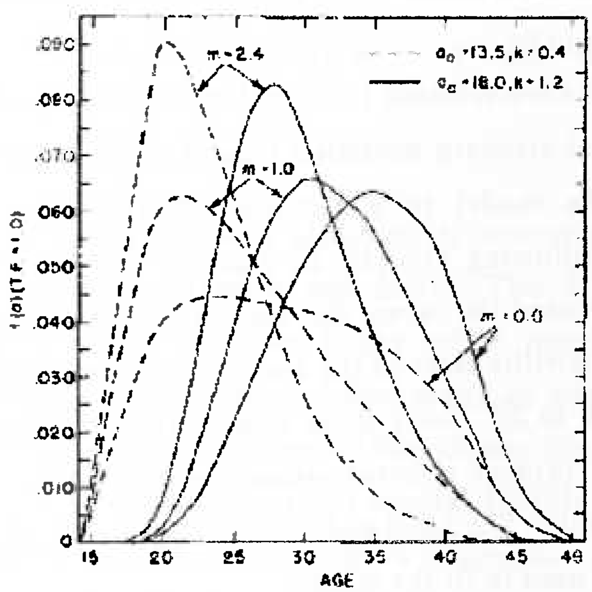

Figure 5.2. The shape of the fitted values of the Coale-Trussel model for various values of its parameters, all with total fertility equals to one. Coale,Trussell(1974)

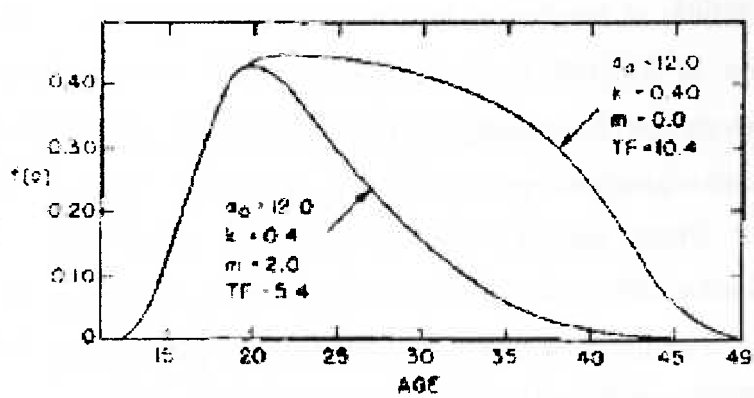

Figure 5.3. The shape of the fitted values of the Coale-Trussel model for two different values of m, and the other parameters fixed. Coale,Trussell(1974)

https://doi.org/10.26219/heal.aueb.5600

5.4 Fitting the model.

Because of the difficulty in calculating the values of the model, it had been considered necessary to calculate the fitted values of the model for a wide range of selected parameter values. For the starting age of marital fertility α_0 its value was allowed to range from 12.5 to 18. Parameter k of the model had been chosen to range from 0.2 to 1.7 where k=1 is considered to be the standard pace of marriage in the Swedish nuptiality schedule. A lower value of 0 and an upper of 3.9 were selected for the parameter m where m=1 is the average value of the forty-three schedules that have been used. All the 795 model schedules that have been tabulated by this way were normalized in order the sum of fertility rates to be 1.

The tabulated models were selected so as to produce mean ages from 24 to 34 at integral values and standard deviation from 4 to 7.5 at half-year intervals. In order to fit the model to a fertility schedule, Coale and Trussel(1974) proposed the following process. In order to find the optimal values of the parameters they used the mean, the standard deviation and the ratio of the average values of fertility rates of the age interval from ages 15 to 19 to the corresponding for 20 to 24. Using these three values, they find the best pattern through the 795 printed patterns whose corresponding values were as more similar as possible. The tabulated schedules were listed in an Appendix sorted by the values used to fit the model

5,5 Testing the model

The model had been tested for several fertility schedules and had shown a successful fit, which remained good even for the more extreme values of the patterns tested. These patterns were as far as the mean ages Hungary for 1970 and Sweden for 1891-1900 and as far as the lowest standard deviation was Japan for 1964. The good fit of the model is shown graphically for these three selected fertility schedules in Figure 5.4. As a measure of a good fit it was referred the fact that the absolute value of the area between the model schedule and the recorded rates is in each instance less than 2.5% of the total area of either curve.

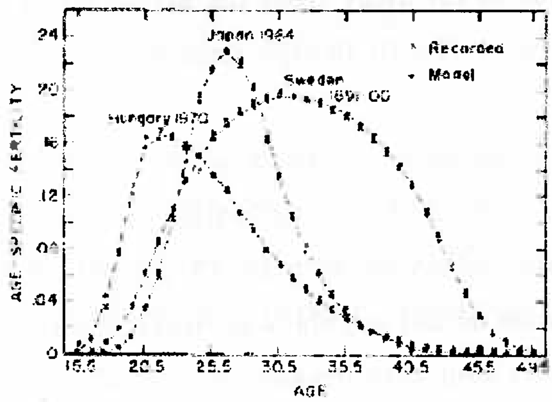

Figure 5.4 Age specific fertility rates of three selected fertility schedules fitted by the Coale-Trussell model. Coale,Trussell(1974)

The model fits well although its assumptions for non-existence of non-marital fertility divorces and deaths. This is no surprising because a population at which there are births before marriage resembles a fertility schedule at which the pattern of first marriage starts a little earlier than the age it actually does. The effect of illegitimate births at an older age can be considered as a slight increase of marital fertility. From the same point of view, divorces and deaths do not have a significant influence in the shape of the fertility pattern because they reduce the ever-married population in the same way that the function $v(\alpha)$ does.

Another assumption of the model, which has not been referred above, but arises by the way the model is defined is that it is addressed to cohort fertility rates. The decomposition of fertility to marital fertility and the proportion ever married is a consequence of viewing fertility as the result of a cohort moving through age. This would create the impression that the model would only fit well to cohort fertility age-specific rates and not to period ones. The lack of fit would become greater when conditions about the marring habits in a population changed though time.

The model was tested to fertility schedules independently of confronting to the model assumptions. As a result of the model assumption

https://doi.org/10.26219/heal.aueb.5600

violations is that the values at which the model achieves its best fit can be far away from the true values of the demographic figures that they represent. For example, the best fit of Japans fertility pattern for 1964 implies a mean age of marriage of 32.4 years. This value is far away from the actual mean age of marriage in Japan which is 24 years. The fit though remains good although the violations.

A final adjustment was proposed to be made on the model so as to overcome a weakness that it was noticed. The fitted values of the model is zero for all ages below α_0, so for values of α_0 over 15 we obtained zero fertility rates. This does not occur to actual populations, where fertility rates for over 15 years of age are always non zero because of a small amount of extramarital births, even when marriages start later. The final adjustment was achieved as follows: the value of fertility of exact age twenty and the cumulated fertility up to the same age were accepted as already defined by the model. The new values for ages below 20 were

$$f(\alpha)=R(\alpha-15)^n$$

with R and n chosen so as

f(20) be the value of the original model and

$R\int_{15}^{20}(a-15)^n da$ being the cumulated fertility up to the age of 20

At the Appendix of the tabulated values the chosen values for R and n for each fertility schedule with α_0 over 15 are available.

5.6 A variation of the Coale Trussell model by Page.

Page(1977) proposed a variation of the Coale Trussel model. What he proposed was a further decomposition of fertility rates. Specifically he

https://doi.org/10.26219/heal.aueb.5600

proposed that in the calculation of marital fertility, duration of marriage should also be taken into account. The final model that he proposed was

$$m(\alpha,d,t)=L'(t)\upsilon(\alpha)e^{(d-2)\,b(t')}$$

Where m is the function of fertility rates, α is the age, d is marriage duration and t is the time period of reference

$L'(t)$ is a parameter that specifies the level of fertility at time t,

$\upsilon(\alpha)$ is a standard age pattern resembling that of natural fertility of Henri(1961)

and $b(t')$ is a parameter characterizing the latent level of fertility control for the cohort married at time (t-d)

Page (1977) concluded to this final form of the model started proposing a simpler model that suggested the decomposition of marital fertility to two independent patterns. The first one is the age pattern while the second one is the pattern of marriage duration. The initial model proposed was

$$m(\alpha,d,t)=L(t)\ \upsilon(\alpha,t)\ u(d,t)$$

where $L(t)$ is the general level of marital fertility averaged through all marriage durations and ages.

$\upsilon(\alpha,t)$ is a factor of the age effect at age α and at time t and

$u(d,t)$ is a factor of the marriage duration effect after d years of marriage and at time t.

The model simplifies the factors of fertility assuming that there is no interaction between marriage duration and age. This assumption, though unrealistic, was implied by the pattern appeared in data from Sweden as well as from England and Wales. The data from these countries were the only available which gave fertility rates differentiated by age and by marriage duration.

The conclusion by examining the age pattern of fertility for each marriage duration separately, was that the data exhibit the same pattern of age

https://doi.org/10.26219/heal.aueb.5600

across all marriage durations. This fact is shown in Figure 5.6. This pattern remained the same for all marriage durations with the difference of becoming lower for higher marriage durations. The pattern was quite similar to the pattern of natural fertility introduced by Henri(1961). The pattern was not possible to be examined for large marriage durations and low ages. The aggregate age pattern of fertility would be a weighted average of the patterns for specific fertility. The weights would depend on the marriage pattern of the population.

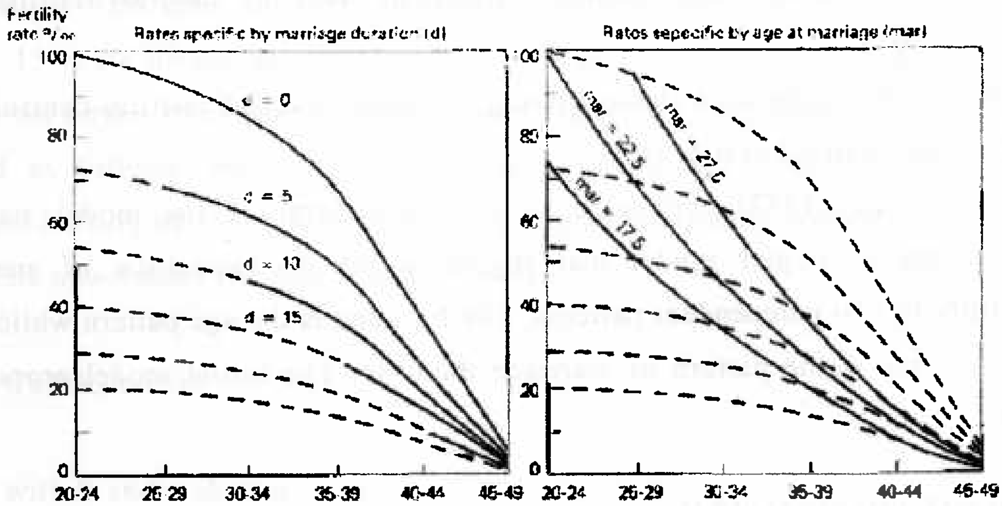

Figure 5.6 Age structure of fertility when fertility depends only on age or marriage duration. Page(1977)

In order to fit of the model a robust method was used since the prior purpose of the model was to discover the underlying patterns of marital fertility. A good fit was not a first priority so a qualitative method was considered as more appropriate for fitting the model. The method in particular ignored the residuals, which had extreme values relative to the rest. The fit was accomplished with the minimization of the squared residuals that had remained. As a consequence the fitted values were not the mean values of all the values but a trimmed mean of them. A trimmed mean though for some values, as those of young woman with large marriage duration, were not available. This difficulty was resolved by several iterations.

To test the goodness of fit of the model there is no formal statistical test which arises directly from the method used to fit the model. A visual

https://doi.org/10.26219/heal.aueb.5600

representation of the residuals was enough to demonstrate the good fit of the model. The R^2 statistic was used as an indicator of the goodness of fit. This statistic is not appropriate to test the goodness of fit because the method of fitting used, does not maximize R^2.

The values of R^2 are extremely high and exceed the value of 0.96 for urban areas and 0.98 for national data. If a least squares method was used the value of R^2 would be even higher as for example for the period 1931-35 the value of 0.972 would have raised to 0.994. The adjusted value of R^2 according to the number of parameters specified remains high even when a large number of parameters are specified. A large amount of unexplained variation is attributed to residuals of marriage duration zero. This fact become obvious if we exclude duration zero from the model, then the value of R^2 exceeds 0.99 for almost every period.

Another control that have been made to evaluate the fit of the model is about the pattern of residuals. The fit was exceptionally well for an early period 1911-1950 but for the later period the fit was somewhat lower so the data were separated into two categories.

For the first category the only new problem detected was for women who married unusually early. It seems to be a tendency for high fertility right after an early marriage, which is made obvious with high positive residuals. The already known problem for duration zero was again detected.

For the later period, the fit is not that well. The problem starts at 1950 and even earlier (1940) for the urban areas. This lack of fit according to Page (1977) is likely to be due to the fact that a part of the population had more that one marriage. This interpretation seems very likely because at that period there is a start in having more than one marriage and it seems reasonable that a second marriage would be less fertile than the first. Unfortunately the order of marriage is not available for the data so the assumption couldn't be tested. It should be noticed that in data as those from England and Wales, which refer only to first marriage, this problem does not exist, a fact that sustains the interpretation.

Another interesting result of the analysis was that the period parameter t in the age pattern $\upsilon(\alpha,t)$ was considered dispensable. The pattern of $\upsilon(\alpha,t)$ was similar for every t with a slight difference in steepness. This

https://doi.org/10.26219/heal.aueb.5600

difference is not evident in data from England and Wales, which indicates a possible similarity of the reasons. If we exclude the parameter t from the model the proportion of variance explained, R^2, drops only by two percent. In the final model the standard age pattern $\upsilon(\alpha)$ was used instead of the $\upsilon(\alpha,t)$.

An analogous procedure of standardizing the period effect of marriage duration effect was attempted but didn't meet with success. A standard pattern of marriage duration effect was discovered though for cohorts. Cohorts exhibit a declining fertility rate for all ages, which is only affected by period effects. After these remarks the model was modified as follows.

The parameter L(t) was substituted from L′(t) which represents the average value of fertility for marriage duration two years. Marriage duration two years was selected in order for L(t) not to represent the average value of fertility for all marriage durations but a value at the peak of fertility.

The parameter u(d,t) was substituted from u′(d,t) which represents the effect of duration d measured as the ratio of fertility rate typical for woman at duration d to the rate of woman at duration two. Afterwards u′(d,t) was substituted with $e^{(d-2)\, b(t')}$ which was assumed that it demonstrates well the way duration effect declines.

For the calculation of the model $u(d,t)=e^{(d-2)\, b(t')}$ it had been shown that it had a good fit and the statistic R^2 for most of the times remained over 0.95. The parameter b it can be considered as a indicator of fertility control as it shows how fast fertility declines with duration of marriage. The fitted value of this parameter keeps rising for later years where as we know fertility control rises as well.

5.7 Comparing Coalle Trussell model with its variation.

For the comparison of the two models first we had to calculate the age specific fertility rates from the variation of the Coale-Trussell model proposed by Page(1977). This had been made with the following procedure. First the values of u′(d,t) and L′(t) were calculated according to the model

$m(\alpha,d,t)=L'(t)\upsilon(\alpha)\ u'(d,t)$. The next step was to calculate the average value of b for each marriage cohort by the $u'\ (d,t)$ values. After that the values of $u'\ (d,t)$ were reestimated so as to take the values implied by b. The already known problem of disproportionaly high fertility at duration zero was coped with adding the amount of 0.3 to the logarithm of the estimate $u'\ (0,t)$. This number is the average discrepancy observed. The new values of $u'(d,t)$ as well as the values of $L'(t)$ and $\upsilon(\alpha)$ were used to calculate the fitted values of $m(\alpha,d,t)$.

The age and duration specific fertility rates were weighted accordingly to the number of women of specific marriage duration that exist in the population at all ages. Since the only data, that marriage duration was available at each age, were those used from the beginning, those were also the data for which the two models were compared. The general result of the comparison was that the second model generally estimates better the age specific fertility rates. The only point where Coalle Trussel model provides a better estimation is for the 20-24 age group.

[illegible]

5.7 Compa[illegible]

For the c[illegible]
age specific fertility [illegible]
proposed by Page(1977) [illegible]
First the values of $n(a,$[illegible]

https://doi.org/10.26219/heal.aueb.5600

Chapter 6

Pearsonian curves

6.1 Pearsonian type I curve

There are many distributions proposed for modeling age specific fertility. One of them is the beta distribution, whose density function common form is given by the following formula.

$$h(x) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}(\beta-\alpha)^{-(A+B-1)}(\chi-\alpha)^{A-1}(\beta-\chi)^{B-1} \quad \text{for } \alpha<\chi<\beta$$

the total number of parameters of the distribution is four. We will use an more general form of this distribution to model age specific fertility. Personian type I curves had been proposed for this purpose. The form of the function that was used by Mitra (1967) was

$$y = y_0(1+\frac{x}{a_1})^{m_1}(1-\frac{x}{a_2})^{m_2} \quad \text{for } -\alpha_1<\chi<\alpha_2$$

According to this model, the origin of the age axis is set to be at the mode of the age specific fertility, which is y_0. Two additional restrictions were set to be $\frac{m_1}{a_1}=\frac{m_2}{a_2}$ and $a_1+a_2=35$. The second one was derived by the desire for the range of fertility capability to be from 15 years of age to 50.

This equation represents a pearsonian type I curve. This curve has a tendency to increase for the negative region of x, reach its peak at zero and decrease for positive values of x. Its skew is depended by the relative size of

https://doi.org/10.26219/heal.aueb.5600

a_1 and a_2, for $a_1=a_2$ we have a symmetrical curve. Beta distribution is a special case of this curve. This curve has the same shape as beta but it is not restricted for the total fertility to be one, so we have one more parameter to specify. Considering that fertility has a standard point where it starts and a standard point where it stops the parameters of the model can be reduced to 3, one determining the total fertility rate and the other two determining the shape of the distribution.

6.2 Fitting the model

The fitting method that was used was based on the knowledge of the two first moments of the age distribution of fertility and the gross total fertility, which is the sum of all observed age specific fertility rates. The two first moments are used to estimate the parameters m_1,m_2,a_1,a_2 and the gross fertility rate for the estimation of y_0. It have been managed to use only the two first moments instead of four by adding the two restrictions mentioned before. By this way we only need to determine two of the parameters m_1,m_2,a_1,a_2 in order to attain the whole set of the four parameters.

The estimation of the parameters was attained by equating the mean and the variance of a pearsonian type I distribution with those of the age distribution of births and using the restrictions as equations. The estimation of y_0 was attained by equalizing the sum of observed and graduated distributions.

Some alternative ways of estimating the values of a_1, m_1 and y_0 were proposed by the authors. Their performance was tested, based on how close they come to the estimation obtained with the first way. For the parameter a_1 was proposed the use of only the two first age groups. This proposal was based on the idea that the speed in which marriages increases at peak ages of fertility should be crucial in determining the peak age of fertility. Following this notion, a new parameter was used named ρ with value

$$\rho=\frac{m_{25-29}-m_{20-24}}{m_{20-24}}100,$$

https://doi.org/10.26219/heal.aueb.5600

where m_{25-29} is the proportion of woman married in ages 25-29 and m_{20-24} is the proportion of woman married in ages 20-24

The parameter a_1 is proposed to be estimated from the value $22+\rho$, which gives values close to the values of a_1 attained by the first mentioned method of fitting. This method of estimating the parameter a_1 is expected to give better estimates at high level fertility countries where fertility is not distorted and the pattern of marriage plays the central role in the pattern of fertility.

The previous way of estimating a_1 was used to estimate the parameter m_1. From the estimation of a_1 an estimation of m_1 was attained by adding 2.5 to the estimated value of a_1. The success of this estimation was not as good as for the parameter m_1, using always the same criterion. Only for nine out of thirteen countries, the estimated values were reasonably close to the fitted values with the first method.

The set of the two estimated parameters, which is enough to construct the relative age specific fertility rates, was tested to see how well the model fits and new Δ values were calculated for the new estimated values of the parameters. The new Δ values were, for all cases but two, larger than the old ones but were considered satisfactory apart from two countries where the values of Δ were extremely higher.

6.3 Testing the model

The model was tested on data of fifty countries from the seventy-two ones available on the United Nations study on fertility patterns. The number of fifty countries was considered large enough for the purpose of testing the fit of the model. The data used were available in five-year age groups. The value used to test the good fit of the model was the index of differential composition, which was symbolized by Δ. This is the sum of the absolute value of the differences between the observed and the expected group fertility rates, expressed as a percentage of gross total fertility. The selection of Δ as a

https://doi.org/10.26219/heal.aueb.5600

measure of goodness of fit has the disadvantage that we get a significance level for the goodness of fit.

The values of Δ ranged from 1.01 to 9.75. A test was performed in order to check whether the curve fits better to non-distorted fertility patterns from fertility control than the rest. High fertility patterns were considered to represent non-distorted fertility patterns. From the nineteen countries, which demonstrated values of gross fertility total more than 1000, eighteen had values of Δ less than 5%; the only exception was Taiwan with Δ value of 5.02%. On the other hand, all patterns with gross total fertility less than 600 have values of Δ higher than 3%. These values suggest that this kind of curve fits better to high-level fertility patterns whose shape has not been distorted from birth control. For high fertility schedules, the fitted curve has in common with the actual curve more than 90%. This value for low fertility schedules was never less than 80%.

6.4 Indications for distorted pattern of fertility

In general, the distortion of the natural fertility pattern due to fertility control in mainly at the sides of the age range. This is because women select to make children at ages in the middle of the age span. This means that in populations with high fertility control the fertility of the younger and older ages becomes lower and therefore the variance of fertility reduces. Hence, smaller variances are related with smaller gross fertility total. The data showed that sixteen out of twenty countries with gross fertility total less than 600 had variances less than 35. This number for the nineteen countries with gross fertility total more than 1000 was zero; fourteen of these countries had variances more than 45 but none of the first group countries.

Another indicator of distorted fertility may be a small contribution of the youngest age group to the total fertility something that may occur by delayed marriages. The distortion of the age pattern of fertility it was also tried to been studied from the point of view of what percentage of total fertility belongs to the age group where fertility reaches its peak and where

https://doi.org/10.26219/heal.aueb.5600

this peak lays. In order to find where the peak is placed, the fitted values were used, because the data were available only in five-year age intervals.

The age where fertility peaks was compared to the gross total fertility in order to examine their relation. The value of fertility rate at the age where fertility peaks was expressed as a percentage of total fertility, which was calculated as five times the gross total fertility. It was found that to countries with high gross fertility total the range of the percentage of modal fertility was from 4.6 to 6.0 a smaller range than the range for all countries, which was from 4.6 to 8.6.

6.5 Pearsonian type III curve

Another distribution proposed for modeling age specific fertility is gamma function that its distribution density function is given by the type

$$h(x) = \frac{1}{\Gamma(b)c^b}(x-d)^{b-1}\exp\{\frac{-(x-d)}{c}\} \quad \text{for } x>d$$

The parameter d represents the lower age of childbearing

The parameters c , b do not have a direct demographic interpretation. The density function of a gamma distribution forms a pearsonian type III curve which in its more general form does not have to be integrated to one.

An comparative evaluation of the performance of this curve to most of the presented curves is given in Chapter 9, thus a particular presentation of the fit of this curve in this chapter was not considered necessary.

https://doi.org/10.26219/heal.aueb.5600

Chapter 7

Hadwiger function

7.1 Introduction

Another distribution proposed for modeling age specific fertility is the inverse Gaussian. The density function of an inverse Gaussian distribution was named after Hadwiger who was the first who proposed its use for demographic purposes. The type of Hadwiger function is

$$h(x) = \frac{ab}{c}\left(\frac{c}{x-d}\right)^{\frac{3}{2}} \exp\{-b^2(\frac{c}{x-d} + \frac{x-d}{c} - 2)\} \text{ for x>d}$$

If we set $a = \frac{1}{\sqrt{\pi}}$ and d=0 we have the density of inverse Gaussian distribution.

7.2 A model based on Hadwiger function

A model that was based on Hadwiger function was proposed by Chandola et al (1999). They proposed a mixture of two Hadwiger functions in order to model age specific fertility. This model was especially designed to deal with a problem recently aroused in the age specific fertility curve. Some later patterns from United Kingdom and Ireland exhibit a hump at the early years, which destroys the fit of all the other proposed fertility models. The suggested model assumes that the reason for this bulk is the existence of a heterogeneous population so they used a mixture model of two Hadwiger

https://doi.org/10.26219/heal.aueb.5600

functions. Mixture models assume that what we observe is due to the existence of separate procedures. These separate procedures act independently with a probability of either one of them to occur. Modeling age specific fertility rates with a mixture model, it may increase the number of the parameters, but it keeps the explanation of the procedure, for a homogenous population, in only few parameters.

The mixture Hadwiger model was compared to the simple Hadwiger model on 1994 data. The simple Hadwiger model was found to have a good fit on the Danish, French, Austrian and Swedish data but the fit was not satisfactory for the data from United Kingdom and Ireland because of the hump mentioned before. At these cases, the mixture model of two hadwiger functions was capable of describing the shape of the age specific fertility curve. This model suggests that there is heterogeneity in the population, which creates the hump.

7.3 An interpretation for the heterogeneity

The first step in order to detect the causes for the heterogeneity is to detract the two separate independent procedures. If two different subpopulations were found that, their fertility has the same shape with the two different procedures, then a strong indication for the reason for the heterogeneity is available.

The smaller of the two procedures, for the United Kingdom and Ireland, was found to represent the one forth of the total fertility. Chandola et al.(1999) suggested that the two separated procedures could correspond to marital and non-marital fertility. The shape of marital and non-marital fertility for these two countries resembled that of the two separate procedures, the smaller group for the non-marital fertility and the larger group for the marital procedure. There are though some problems with this suggestion, such as an underestimation of the right tail of the non-marital distribution, which comes together with an overestimation of the size of marital fertility for the

https://doi.org/10.26219/heal.aueb.5600

larger age groups. The two procedures do not represent well the size of marital and non-marital fertility but they represent well their age structure.

In order to support the hypothesis that the two mixtures represent marital and non-marital fertility, the analogous patterns for the other countries , which do not exhibit the hub in the early fertility year, ought to be checked for an analogous structure. For these countries, the relation between the relative size of the two mixtures and the ratio of marital to non-marital fertility was tested. For this purpose, Chandola et al.(1999) checked the linear relation between the mixing parameter and the illegitimacy ratio and found that indeed there was a linear relation between them. The parameter R^2 for this relation was equal to 0.751. It is important to notice that there are countries with higher level of non-marital fertility than England and Ireland, which does not exhibit the hump.

A logical question raised from these results is why countries with higher level of non-marital fertility do not also exhibit a hump in the age pattern of fertility. If the reason for a hump is a high level of non-marital fertility then all the countries with higher non-marital fertility than England and Ireland would also exhibit a hump. There is though a difference in non-marital fertility for these countries because they do not show high teenage fertility rates as United Kingdom and Ireland does. A solution to this problem could be found if we check the age patterns of marital and non-marital fertility for the countries that not exhibit the hub. It is obvious that the age patterns of both marital and non-marital fertility for these countries are similar; a fact that can explain why countries with high illegitimacy ratio have smooth fertility curves so the simple hadwiger model fits well.

If we center our attention to the previous year's data of these countries, we can see that this hub was created since 1970s in United Kingdom and since 1980s in Ireland. Before these years, a simple hadwiger model would be enough for a good fit but after these years, the age patterns for these countries started to change. This change was mainly due to the decrease of the fertility rates at all ages apart from the younger age groups, which remained constant. This movement in the age pattern of fertility has created the hump.

https://doi.org/10.26219/heal.aueb.5600

The association between the illegitimacy ratio and the mixture parameter was tested for these past periods data as well. The parameter R^2 was equal to 0.9109 showing that there was also a linear relation. This linear relation shows that while non-marital fertility was becoming higher then the mixing parameter was also becoming higher. The lack of fit could suspect us that non-marital fertility is not the only factor that influences the mixing parameter. the reason for the heterogeneity could be found if from non-marital fertility, is excluded that fertility which come from couples whose behavior is similar to that of married ones then what remains could be the reason for the hump.

7.4 Interpretation of the parameters

The parameters of the simple model have being shown, by recent studies, contrary to previous beliefs, to have a demographic interpretation. This interpretation does not hold for the mixture model as well. The way that this interpretation was shown was by finding correlations between the parameters of the fitted model on a population and some demographic terms of the same population. The parameter a of the model was found to be correlated with total fertility and it represented about 0.56 of the total fertility. The evidence for this correlation is the value of R^2, which was equal to 0.9895. Parameter c is linearly correlated with the mean age of childbearing which was shown by the value of R^2 being equal to 0.9896. As far as for the parameter b, it was found, that the amount $\frac{ab}{c}$ was linearly correlated with modal age of childbearing. The value of R^2 for this correlation was also high and specifically 0.9891. A demographic interpretation of parameter b can be extracted by its participation in the amount $\frac{ab}{c}$.

It could be expect that these interpretations would hold for each mixing part of the hadwiger function but this is not the case. For example, the parameter c of the one mixing part would represent the mean age of marital

https://doi.org/10.26219/heal.aueb.5600

motherhood and the parameter c of the other mixing part would represent the mean age of non-marital motherhood. For this hypothesis, there is evidence only for the marital case but there is no evidence for the non-marital case.

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

Chapter 8

The Gompertz function.

8.1 Introduction

One of the functions used for modeling age specific fertility is the Gompertz function whose general form is

$$F(x) = FA^{B^x},$$

where F, A and B are parameters with the restrictions, $0<A<1$ and $0<B<1$.

The first use of this function in Demography was in describing the age pattern of mortality. Later its use was also proposed for describing the age pattern of fertility. The shape of this function is not appropriate for modeling age specific fertility rates. Its form, which is monotonic with x, is only appropriate for modeling cumulative age specific fertility rates. Due to the fact that $F'(x)$ reaches its peak for $x=0$, because $F''(0)=0$ something which certainly not occur in fertility rates, a location transformation for x, which represent age, is used. With this transformation, fertility reaches its peak at a non zero value of x. Therefore, the form of the Gompertz function used for modeling cumulative fertility is

$$F(x) = FA^{B^{x-x_0}},$$

where x_0 is an arbitrary origin of the age scale.

There are models proposed, which make use of this form to the Gompertz function to fit the cumulative age specific fertility curve but there are other authors believe that this location transformation is not enough to create a model with a good fit. They consider that a further transformation of the age scale necessary to achieve this purpose for the model. For this purpose a transformation of the age scale have been proposed based on some principles proposed by Brass (1971) previously used for modeling mortality.

https://doi.org/10.26219/heal.aueb.5600

Brass (1971) had proposed for modeling mortality that age scale should be transformed so as to determine a standard form of mortality $h_s(x)$. A standard form of mortality would be determined by empirical patterns and the rest of the parameters of the model would specify any variations of the standard pattern. Heather Booth (1984) applied these principles and tried to find an appropriate transformation for the age scale so as to represent a standard pattern of cumulative fertility. The model proposed can be expressed as a linear relation between the double logarithm of cumulative fertility and the transformed age scale, that is

$$Y(x)=a+bY_s(x),$$

where $a=-\log(-\log(A))$, $b=-\log(B)$, and $Y_s(x)$ is the transformed age scale, while A and B is the original parameters.

This equivalent form of the model can be attained with a double logarithm of the original model, that is

$$Y(x) = -\log(-\log\frac{F(x)}{F})$$

where F(x) is the cumulative fertility.

Heather Booth tried to determine the standard pattern of fertility by finding the best transformation of the age scale $Y_s(x)$ which would achieve the best fit for the model. The model was written in this form, to make obvious that a transformation of the age scale, which makes $Y_s(x)$ linearly related to the double logarithm of the cumulative fertility rates is the best choice. The untransformed pattern of cumulative fertility rates has this form only to the center of the childbearing range, but at the sides its shape changes, losing the linear relation and creating a lack of feet. The question is to find the appropriate transformation to correct the problem at the sides.

For the calculation of the standard pattern of fertility, high fertility patterns were used. This choice has led to a model more suitable for high fertility schedules. The high fertility schedules used were not selected from high fertility schedules observed in real populations but were produced from the Coalle-Trussel model. Coalle-Trussel model which has a lot of empirical elements, has three parameters named a_0, k and, m. The patterns used were produced by setting values of a_0 from 10 to 15 in steps of 0.5 years. The low values for a_0, which in the model represents the age at which marriage procedure starts, were selected because high fertility schedules occur in population where marriages start at a low age. The selected patterns had mean values between 27 and 29 and standard deviations below 6,75. The values of the other parameters of the Coale-Trussel model were automatically determent to be $0.1<k<1.3$ and $0<m<1$. The value of parameter m was further restricted to $m<0.6$ in order to attain more likely fertility schedules. A connection of the values of a_0 and k was attained by imposing the value of 21 for the singulate mean age at marriage. By this way, the equation $\bar{M} = a_0 + 11,37k = 21$ restricts k for given values of a_0.

Finally a set of thirty-three high fertility schedules were selected for the calculation of the standard pattern of fertility. Because for this calculation the differences of their transformed values were needed, the following process was followed. The schedules were transformed into Y(x) for five year age groups and the differences ΔY were attained. Their averaged value for the three central age groups from 25-39 were considered to be the standard values of $\Delta Y_s(25\text{-}29)$, $\Delta Y_s(30\text{-}34)$, $\Delta Y_s(34\text{-}39)$. In order to calculate the ΔY_s for the rest age groups, at the tails, a weight was given to the schedules with high fertility at younger or older ages. As a result of his construction the model fits better to fertility schedules were a significant amount of the total fertility is located at the tails than to those which not. About half of the thirty-three fertility schedules were used for the calculation of the values of ΔY_s for each tail, 17 for the young ages and 16 for the old ages. The two younger age groups were combined for each age pattern and only the patterns, whose age specific fertility rate for the lower age group was more than 0.15, were used. The age groups from 35 to 49 were combined for each pattern and those patterns that had age specific fertility rate for 35-49 more than 0.21 were selected. Although the age group 35-39 had already been considered to belong to the center, it was used for the calculation of the late age groups. This inclusion in the combination had been made so as the last age group to incorporate more than ten per cent of the total fertility.

In order to bring these values of ΔY to the same level with those at the center adjustment factors were needed. These factors were used to eliminate the use of only a part of the thirty-three schedules. The factors were calculated as the ratio of the average value of the thirty-three schedules of $\Delta Y(25\text{-}39)$ to the value of those schedules used in the calculation. So the value $k_1=0.99135$ was used for the adjustment of the values $\Delta Y(15\text{-}19)$ and $\Delta Y(20\text{-}24)$ and the value of $k_2=1.02287$ was used for the adjustment of $\Delta Y(40\text{-}44)$. A fixed point is required in order to transform the values of ΔY_s into $Y_s(x)$. This point which is the origin of the standard x_{0s} was arbitrary chosen to be $Y_s(30)=0.7$ a rough average of the thirty-three schedule values of $Y(30)$.

8.2 Testing the model

The model was tested for data of low fertility even though it was designed for high fertility patterns. This was chosen because the only good quality data available are those for developed countries, which usually have low fertility. This choice will not cause any problem in the test of the model although the model is tested under different assumptions than those it was made for. The method of fit used was ordinary least squares that were attained by an iterative procedure.

The test was made with both simulated and real data that presented consecutive cohort periods. At first simulated data were attained by a modification of Barrett simulation model. Five fertility schedules were simulated representing consecutive stages of a declining fertility period. The results were that the model tends to underestimates parameter F but no more than 1%. In the case of fitting to fewer points, instead of an underestimate an overestimate of less than 6.2% is obtained.

As far as parameters a and b are concerned we have the following results. There is a decrease on both a and b when F increases and vice versa, with only few exceptions were a and b change to opposite directions. This is the reason why higher estimation of F is associated with lower estimations of b and vice versa. For the parameter a, this relation is rather obvious because lower estimates of a mean that a smaller proportion of total fertility F is reached at age x_{0s}. Large values of parameter b indicate a low variance of fertility, which mean lower values of fertility rates at young and old ages. Therefore, higher fertility patterns are associated to a later and slower pattern of fertility. In both cases an error in estimation of one parameter should be accompanied with the error in the estimation of the other parameter which brings the pattern of fertility back to opposite direction.

Apart from simulated data, the model was also tested on actual data. These were ten complete cohort data from Sweden covering the period 1870/1 to 1915/16 at five-year intervals. For the cohorts 1870/1 to 1900/1 the errors obtained, were of the same size with those of the simulated data but the errors for later cohorts were larger. As far as the signs of the errors, no regular pattern was detected. For the first five cohorts variable errors of both signs were obtained indicating a separate declining pattern of fertility for each age group. Oppositely, the cohorts of 1895/6 and 1900/1 exhibit negative errors, which shows late fertility is higher than expected. This happens because of the period effect in the cohorts. Its age group fertility is affected differently from the period effects in the cohort. This period signs a transition from a period of declining fertility to an increasing fertility period, which creates this image. The following three cohorts exhibit much larger errors with positive signs. The positive signs show that later fertility is lower from what expected. Therefore, it is noticed a general trend for overestimation of F when early fertility is higher than late fertility and an underestimation of F when late fertility is higher than that of early years.

For the cohorts 1910/11 and 1915/16 larger estimates for the parameters b were obtained when incomplete data are used than the estimate obtained using complete data, which indicates higher variances and lower level of fertility. Oppositely for the cohorts 1895/6 and 1900/1 when the model is fitted for incomplete data estimates for b are obtained, which indicate smaller variances and smaller levels of fertility.

https://doi.org/10.26219/heal.aueb.5600

In general negative values of a indicate a later pattern of fertility than that in the standard and values of b more than one indicate a more peaked pattern of fertility than the standard. From the fitted values of a and b a transition through time is obvious. The pattern of fertility changes towards the early years until 1900/1 and afterwards to a more peaked pattern of fertility. The pattern of fertility may change towards early years until 1900/1 but this change does not mean an increase in early fertility during these years but an increase in the proportion of early fertility to the total fertility. An increase in both a and b mean that fertility rises and decreases simultaneously, rises for younger ages and decreases for later ages.

Other efforts had also been made to fit the Gompertz function without an additional transformation of the age scale. Murphy et al. (1972) use the method of selected points. A comparison between the method of selected points and ordinary least squares was made in order to find the best method of fitting. The model used for comparison of the two fitting methods is the initial model, the one which has only the transformation of the origin of the age scale. The comparison had been made on both period and cohort data. Cohort data were covering the period from 1911 to 1945 and period data were from 1926 to 1969. The comparison was made on the basis of which fitting method reduces more the square error and how much. Ordinary least squares were found to reduce the sum of squares up to 90% in comparison with the selected points method.

In order to compare the errors between models fitted to different data a standardization of fertility rates was considered necessary. This was due to the fact that different cohorts or periods can exhibit quite different part of their population to belong to each age group, so errors of each age group should be taken into account accordingly to the participation of the age group to the population. One standard population was used for all cohorts and another one for all periods. The one chosen for cohorts was the, by the year of the research, last available official life table for Canadian women, which was for the period 1965-1967. As far as period data are concerned, although period patterns had not changed drastically, it was chosen for the standard population the age distribution of Canadian women in the 1966 Census graduated from five-year age group, using Sprague's multipliers.

8.3 Goodness of fit

At first the goodness of fit was tested for the cohort data. In specific for the 1911 cohort the error was of 2.8% of the fitted values. As a measure of comparison for this value, in the paper is referred, that for the same cohort if the technique of approximation, suggested by Keyfitz, to the sampling error of a standardized rate is used then the net error would be such as 2,47%. For this cohort, and also for the cohorts until 1920, our data are complete at least till the age of 45. The net error for those cohorts is between 1.62 and 2.87.

For the cohorts until 1920, as previously mentioned, data until the age of 45 are available for all cohorts, but for the later cohorts only truncated data are available. The fit of the model is better in the center than in the sides. This fact lead us to expect better fits for data with one side missing, such us truncated data. In fact, if the model is fitted to the cohorts, which there are available the fertility rates till the age of 30 then the errors are between 1.36 for the 1921 cohort to 0.0001 for the 1931 cohort. For these cohorts more than 60% of the whole fertility is known judging from previous cohorts.

Another measure of goodness of fit used in the paper was the gross error. The values of the gross error for the cohorts until 1935 are between 2.5% and 14%. To select between the measures the use of the fitted curve should be taken into account. If modeling fertility is needed in order to use the whole set of fitted values, then the most appropriate measure is the net error. If on the other hand it is only needed a part of the fitted values such cases as transforming cohort data to period then the gross error is the more appropriate one. The large values of gross error make the Gompertz function inappropriate for making estimates for cohorts and translating the results into period rates, a projection technique which was at the time of the research considered standard.

The results from the fitted curves for period data were the following. The net error for the periods from 1926 to 1969 was 0.72 the lowest and 1.87 the largest. The same numbers of the gross error were 7% the lowest and 12.02% the largest. The sum of squared error ranged from 0.0053 to 0.0417. In the data is obvious an improvement in the fit for the period data as we

https://doi.org/10.26219/heal.aueb.5600

move to more recent periods. The reason for this improvement, since there are no truncation problems, it can be infered that it is the fact that we move to a period were most fertility lies at the part of the function where we have better fits, that is at the centre of it.

8.4 Interpretation of the parameters.

As we have already said Gompertz function involves three parameters F, A, B. In addition to these parameters, there is parameter x_0 which is the origin of the age axis. The demographic interpretation of parameter F is simple and rather obvious. The Gomprertz function has asymptote F, so F is the value that cumulative fertility reaches asymptotically, which means that the demographic interpretation of parameter F is the total level of fertility.

For the parameter A we notice that for $x=x_0$ we have

$$Y = F\ A^{B^{x-x_0}} = F\ A^{B^0} = F\ A^1 = FA$$

So the completed fertility at age x_0 is Y=FA and because A takes values between 0 and 1 it is obvious that A shows the proportion of the total fertility completed at the age of x_0. This age is usually taken as $x_0=24$ so A to demonstrate how much of the whole fertility is completed at the younger age groups, since the age of 24.

The demographical interpretation of parameter B is not simple or obvious. To find an interpretation the derivative of the Gompertz function has to be used which is

$$F'(x) = FA^{B^{(x-x_0)}}\ B^{x-x_0}\ \ln(A)\ln(B)$$

As far as F(x) represents cumulative fertility rates its derivative represents age specific fertility rates. The shape of this function exhibits a unique maximum which is the peak of the fertility rates. Another property of the function is that the values of the fitted fertility rates decrease for larger values of B inside an area around the maximum. This property shows us that parameter B can be seen as an indicator of how much of the whole fertility is concentrated around the peak.

Gompertz function has also the property that at the point that its derivative reaches its peak the value of cumulative fertility rates is known and it is $F(x) = \frac{1}{e}$. This mean that by modelling cumulative fertility by Gompertz function it is assumed that by the time fertility reaches its peak, 37% of total fertility is completed. This assumption does not cause any problems with the the model good fit. Another property of the Gompertz function, which has to be noticed, has to do with the way its derivative reaches zero for older ages. This property is that for high values of parameter B the fitted curve approaches zero at an age above the end of the human childbearing period.

Coming back to period data it is noticed that the value of B rises as we move to more recent periods and this may give an additional explanation why we have better fits at those periods. Patterns which have lower fitted B values is logical to have better fits because there is no problem in the way that the function reaches zero at the end of childbearing period.

8.5 The Makeham function

A suggestion to deal with the poor fit of the Gompertz function is to add an extra parameter. Following this notion, Makeham function was also tested for modelling cumulative age specific fertility rates. The form of Makeham function is given by the formula

$$Y = FS^{t}A^{B^{t}}$$

The Makeham function is the Gompertz function with the addition of the extra parameter S in the form of S^t. This function arises from the Gompertz function if a stable force of fertility is assumed along with the

https://doi.org/10.26219/heal.aueb.5600

pattern that the Gompertz function suggests. The fit is better , as expected because a family of curves which is larger and contains all the previous curves for S=1 is fitted .This additional parameter is something that should be avoided as it may improve the fit but it has the weakness of having to calculate an additional parameter. The idea of modelling is to have available an estimation of the whole set of age specific fertility rates by giving only a few values, that of the parameters, so the better fit of Makeham function don't tempt us to go from three parameters to four.

Another drawback is that by this addition, the interpretation of the parameters that they had before is lost and the new set of parameter has no obvious demographic interpretation. For example parameter F could not anymore be considered as the level of fertility because the function don't reaches anymore the value of F asymptotically but it increases limitlessly for S>1 or reaches zero for S<1. This property has the implication that the fitted values for the cumulative rates of some Canadian data turned down before the age of 49, which implied negative fertility rates.

https://doi.org/10.26219/heal.aueb.5600

Chapter 9

Comparative studies

9.1 Comparing most of the models

A comparison of a wide variety of models, both parametric and non-parametric, has been made by Hoem et al. (1982). The models tested were the Coale-Trussell the gamma density the Hadwiger, the Beta density and also a splines model, two versions of polynomials and two of Brass's procedures. The method used for fitting and the data used were common for all models, to be able to make comparisons between the models. The data were the age specific fertility rates for Denmark for the years 1962-1971 for the ages 15-46. The method used for fitting was minimizing the sum of square error.

Most of the models fitted in this paper have already been presented in this thesis. The two forms of polynomials used, were the one mention above that Brass proposed, little changed, which was $f(x)=(x-15)(47-x)^2(a_4+a_3x+a_2x^2+a_1x^3)$ and the other allowed for to the fixed parts of the latter polynomial to be parameters so $f(x)=(x-a)(b-x)^2(a_n+a_{n-1}x+\ldots+a_1x^n)$. The selection of b=47 for the first form of polynomial was later shown from the fitted values of the second form not to be the best one possible. A selection of b=45 or b=46 would have been more appropriate. This choice would come closer to the original proposal of Brass whose original model considered a difference between a and b of 35 years.

Two Brass's procedures were used, one making use of the transformation $Y(x)=-\ln(-\ln(x))$ and the other making use of the transformation $Y(x)=\frac{1}{2}\ln\frac{x}{1-x}$. The first form of polynomial, of the two mentioned above, was used to represent the untransformed age specific fertility rates. A selection of another function with a better fit was considered to make things too easy. At a previous chapter, we have presented the first of the Brass's procedures but with the derivative of the gompertz function

https://doi.org/10.26219/heal.aueb.5600

used to represent the age specific fertility rates instead of the first form of polynomials that it is used here

9.2 Results of the comparison

The majority of the models were fitted separately for the proportional age specific rate and the total fertility level. The best fit in terms of reducing the sum of square error was obtained by Coalle Trussel model and the gamma density function. The models had very good accuracy and fitted the data very well. Because the parameters of the gamma density do not have any direct demographic interpretation, the authors suggested the Coale Trussell model for use in forecasting populations. The next best function was the Hadwiger function, which has very good fits and come right after those of the two first models. The Hadwiger function was proposed to be used in cases where the use of the two first functions is inconvenient.

These three functions give better fits than the beta function and both Brass procedures. For the Brass's procedures, this was not considered surprising because Brass himself did not claim anything more than a fair representation of the age specific fertility pattern. For the beta function, though this fact was surprising enough to suspect that the algorithm used had reached a local minimum instead of a global one. More so, other authors had reached comparable fits to this and without fitting by minimizing the square error, but further research was not made because Hoem et al. (1982) had already reached to satisfactory results by other models.

As far as the fit of the polynomials and the splines is concerned, a slight different form of fitting was used which was reducing the sum of squares for the whole pattern of fertility including both the proportional age pattern and the total level of fertility. The six-parameter polynomial fits better than the four-parameter polynomial as expected because all four parameters polynomials are constructed to be a special case of the six-parameter form. The fit of the four parameters polynomials was as poor as the Brass procedures. The six-parameter polynomial gives very good fits but not as well as those of the hadwiger function, which has only four parameters. In advance, the calculation of the six parameters for the polynomial was found inconvenient.

https://doi.org/10.26219/heal.aueb.5600

As far the splines are concerned, the form tested was cubic splines with three knots ξ_i, which can be represented as $g(x)=\sum_{k=0}^{3} a_k x^k + \sum_{k=1}^{3} b_k (x-\xi_\kappa)_+^3$

Where $t_+=t$ for $t>0$ and $t_+=0$ for $t \geq 0$ and

For this model we have in total ten parameters $\alpha_0,\alpha_1,\alpha_2,\alpha_3,b_1,b_2,b_3,\xi_1,\xi_2,\xi_3$

The fit of the model is excellent something that is expected due to the large number of its parameters and the general properties of splines. The worst fit of the splines model, which is achieved in 1962, is better than any other fit achieved by any of the models at any year. The best fit of any of the rest models, is achieved in 1971 by the hadwiger function. For the reduction of the number of parameters, fixing the knots is proposed at the ages 20, 24, 29 something that would reduce the number of parameters to seven. The fit of a seven-parameter splines model was not tested, a model that has an additional drawback that its parameters lack of any demographical interpretation. A model such that could be a useful instrument, when only the shape of the age specific fertility pattern, is of interest.

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

Chapter 10

Applications

In order to evaluate and compare the various models performance we fit it in a set of empirical data sets. Our data concern the age specific fertility rates of Northern Ireland for each year during the period 1974-2001 as given by the General Register Office for Northern Ireland. Throughout this period, the pattern of age specific fertility rates exhibits a hump in the early years of childbearing period, which is made obvious in recent years. Figure 10.1 shows the observed fertility rates for selected years. During the examined period, the total level of fertility was declining steadily, as shown in figure 10.1, mainly influencing the fertility rates of younger ages than those of describing the fertility pattern of the later reproductive ages, which roughly remained on the same level.

The goodness of fit of the models is influenced by the bulk that recently appeared in age specific fertility curve. The Coale-Trussell model, gamma density curve and the Hadwiger function was used. To compare the fit of these models we used the R^2 statistic since all these models have the same number of parameters. The values of adjusted R^2 the sum of squares of the error and the mean square error were also calculated for each model.

Coale-Trussell model shows a very good behavior for the period 1974-1989, with the R^2 statistic reaching values greater than 0.99. However, at 1990 the values of R^2 start to decrease. After 1996, the values of R^2 fall bellow the value of 0.95 reaching its lowest value at 2000 being equal to 0.9078. For fitting the model, the Matlab Trust-Region Reflective Newton algorithm was used. The numbers of parameters of the Coale-Trussell model are four, three for the proportional age specific fertility pattern and one for the total fertility pattern. Table 10.1 presents the values of R^2, the adjusted R^2, the sum of squares of the errors and the root of mean square error for each year for the Coale-Trussell model.

The gamma density function exhibits the same behavior throughout the examined period but it shows a greater resistance in the change of fertility pattern of later years. The R^2 statistic retains values over 0.99 until 1993 periods and never falls under the value of 0.95 for all periods. After 1985 the gamma density achieves better

https://doi.org/10.26219/heal.aueb.5600

fits than the Coale-Trussell model. The difference in the goodness of fit is minor until the 1989 but it becomes significant for the later years. We used a four parameter version of gamma density curve, in the three parameters version described before with the addition of a forth describing the total level of fertility. The Matlab Trust-Region Reflective Newton Algorithm was used for estimating the parameters. Table 1 presents the values of R^2, the adjusted R^2, the sum of squares of the errors and the root of mean square error for each year for the gamma density model.

The Hadwiger function shows a similar behavior as the gamma density function. At early periods, Hadwiger function shows a slight inferior fit than the gamma density function, an inferiority which deteriorates through time resulting to a slight superior behavior than the gamma density function at later periods. Therefore, the Hadwiger function can be considered as the most resistant to the change of fertility pattern given the fact that at early years it exhibits a worse fit than the gamma density and the Coale-Trussell model. Hadwiger function achieves the best behavior in terms of R^2 statistic values than the other two models after 1992. The Matlab Trust-region Reflective Newton algorithm was used for estimating the four parameters of the model.

The results of the early periods confirm the findings of Hoem et al. (1982) who conclude that the Coale-Trussell model and the Gamma density achieve the best fits in modeling age specific fertility rates. Next after these two models, according to the same paper, comes the Hadwiger function. The paper of Hoem et al. (1982) was presented at a period when a bulk had not appeared in any of age specific fertility patterns available. In our appliances, we can verify the same results in our data until the period of 1989 where no bulk in the young ages of childbearing period is obvious. At these periods, the Coale-Trussell model and the gamma density achieve the best fits with none of them to mark out. Hadwiger function attained in all cases the worst fit of the three models until 1985. This year can be consider as a start of a transition period because by that year Hadwiger function comes second in goodness fit with gamma function first and Coale-Trussell third. After the period of 1992 Hadwiger function takes first place followed by the gamma density and last of the three comes the Coale-Trussell model. Table 4 shows the values of R^2 statistic for all the models fitted. Table 10.5 shows the sum of squared errors of each of the models fitted. Figures 10.2-10.4 shows the observed and the fitted values of the 1974, 1993 and 2000 schedules.

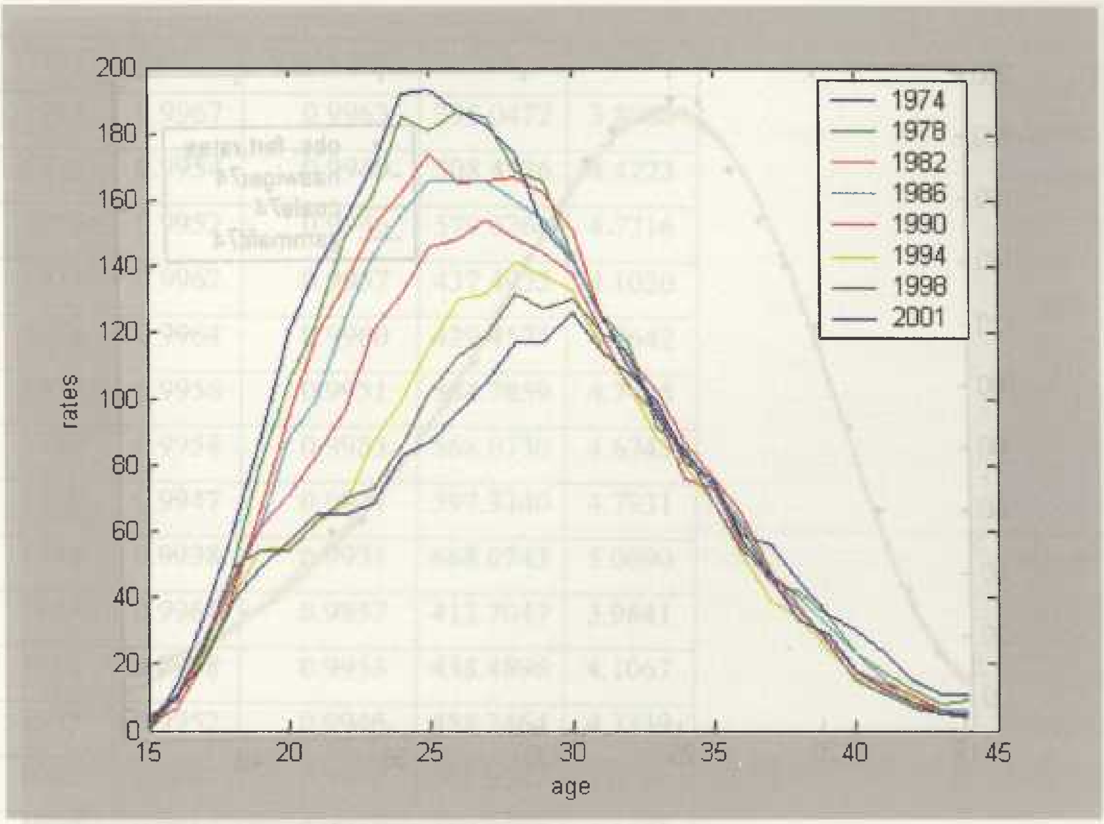

Figure 10.1. Fertility rates for the years 1974,1978,1982,1986,1990,1994,1998,2001

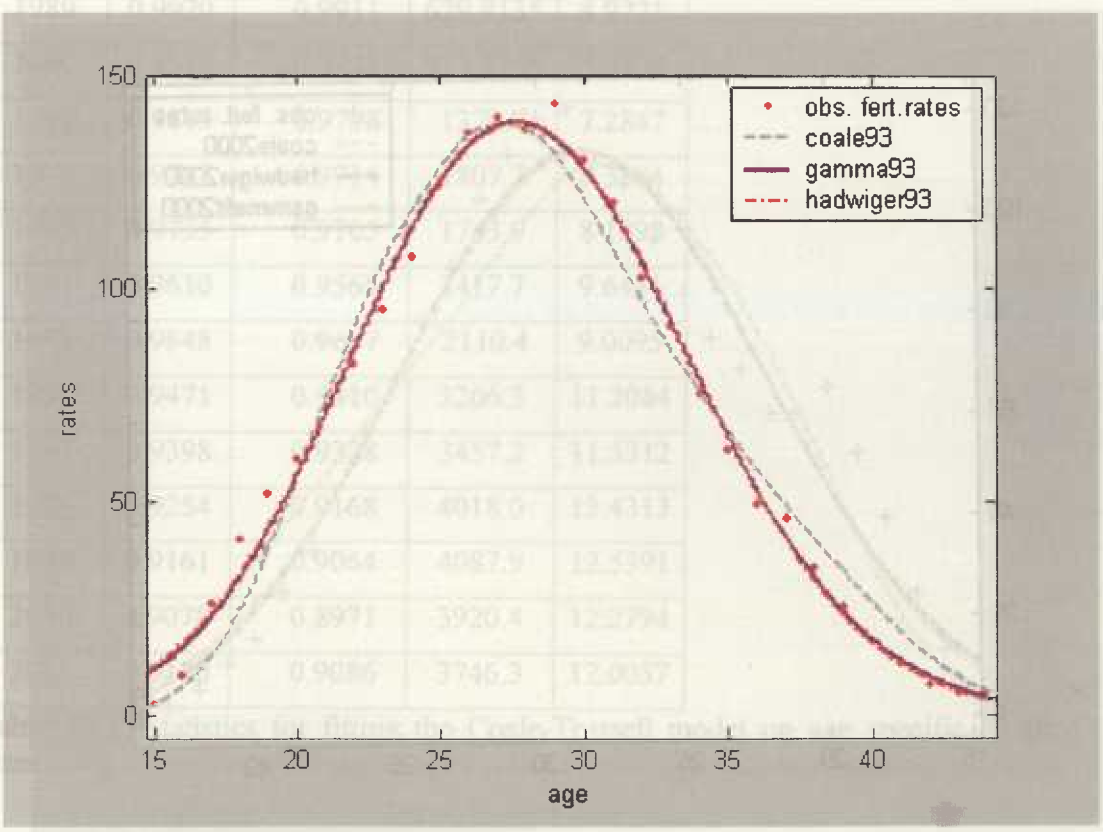

Figure 10.2. Modeling age specific fertility rates for Northern Ireland of 1993

https://doi.org/10.26219/heal.aueb.5600

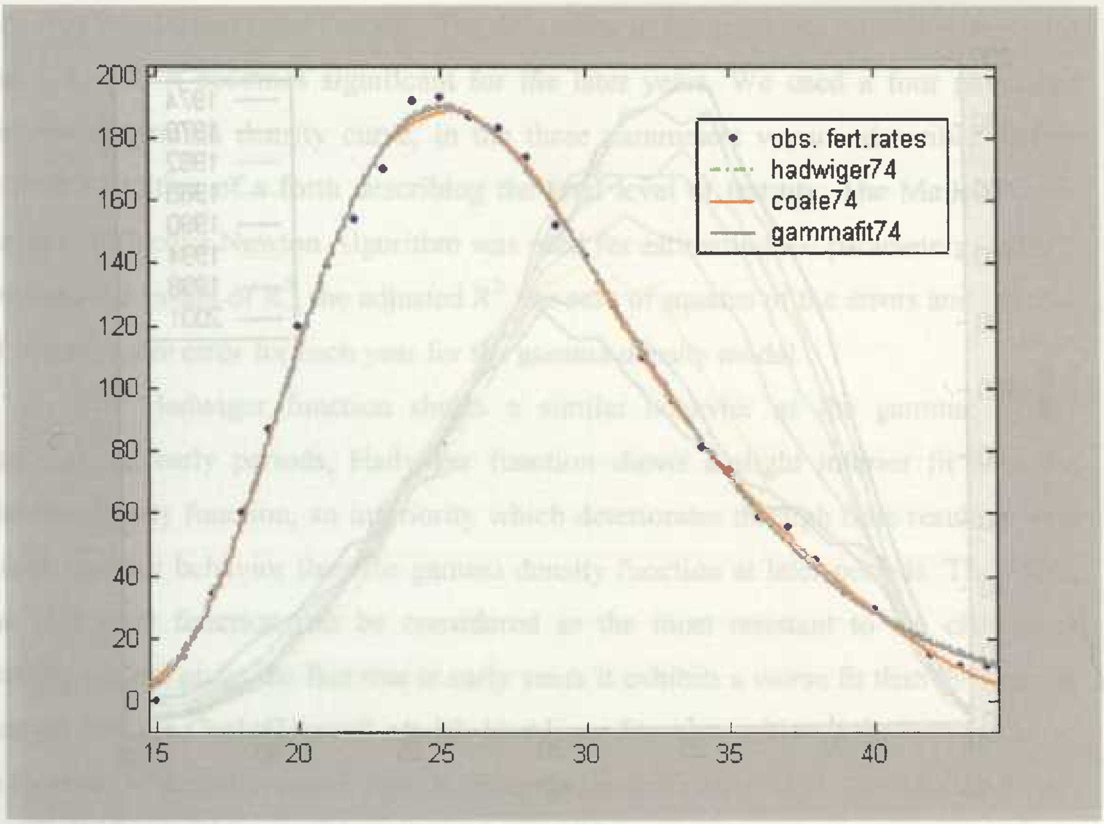

Figure 10.3. Modeling age specific fertility rates for Northern Ireland of 1974

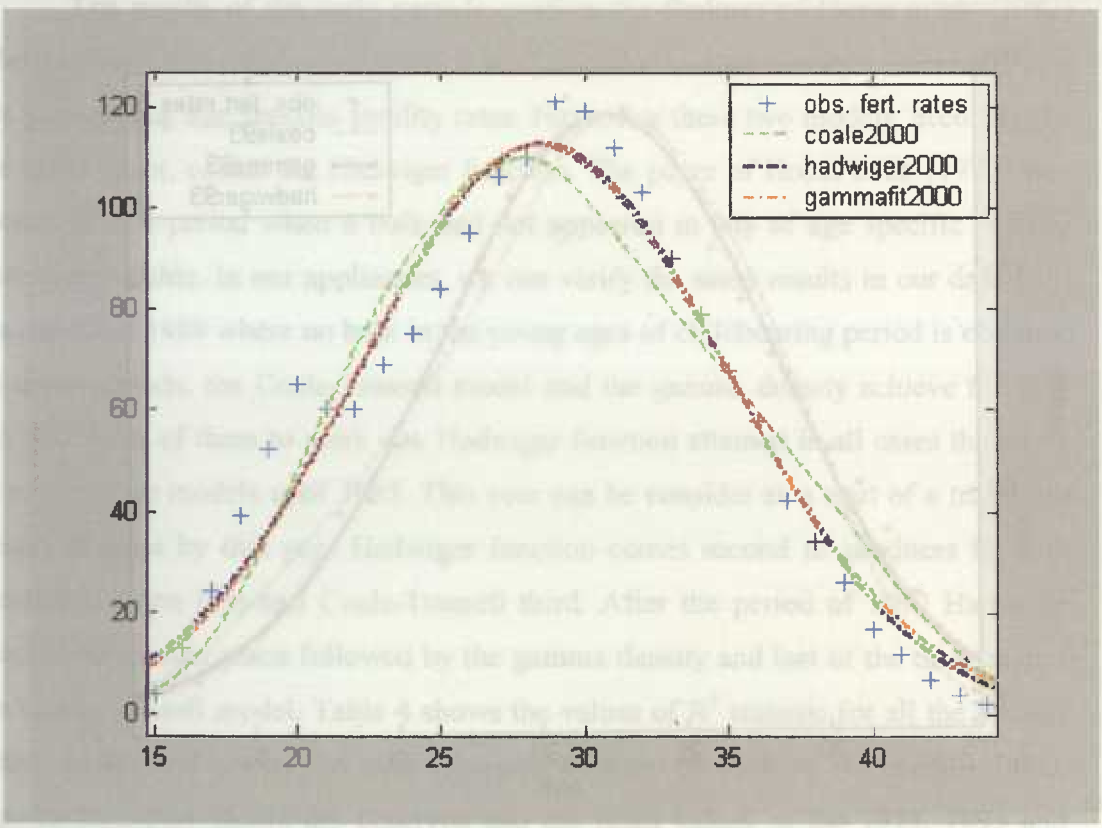

Figure 10.4. Modeling age specific fertility rates of 2000 for Northern Ireland

https://doi.org/10.26219/heal.aueb.5600

Year	R	R adjusted	Sse	rmse
1974	0.9967	0.9963	395.0472	3.8980
1975	0.9954	0.9949	508.4736	4.4223
1976	0.9952	0.9946	579.6264	4.7216
1977	0.9962	0.9957	437.4935	4.1020
1978	0.9964	0.9960	429.4575	4.0642
1979	0.9956	0.9951	584.7859	4.7425
1980	0.9958	0.9953	568.0730	4.6743
1981	0.9947	0.9941	597.3240	4.7931
1982	0.9938	0.9931	668.0743	5.0690
1983	0.9961	0.9957	412.7047	3.9841
1984	0.9958	0.9953	438.4896	4.1067
1985	0.9952	0.9946	488.3464	4.3339
1986	0.9964	0.9960	351.0207	3.6743
1987	0.9952	0.9946	457.7927	4.1961
1988	0.9918	0.9909	738.6746	5.3302
1989	0.9920	0.9911	629.9135	4.9221
1990	0.9885	0.9871	913.5761	5.9277
1991	0.9819	0.9798	1379.7	7.2847
1992	0.9743	0.9714	1807.7	8.3384
1993	0.9735	0.9705	1743.9	8.1898
1994	0.9610	0.9565	2417.7	9.6431
1995	0.9648	0.9607	2110.4	9.0095
1996	0.9471	0.9410	3266.3	11.2084
1997	0.9398	0.9328	3457.2	11.5312
1998	0.9254	0.9168	4018.0	12.4313
1999	0.9161	0.9064	4087.9	12.5391
2000	0.9078	0.8971	3920.4	12.2794
2001	0.9180	0.9086	3746.3	12.0037

Table 10.1. Statistics for fitting the Coale-Trussell model on age specific fertility rates.

https://doi.org/10.26219/heal.aueb.5600

Year	R	Radjusted	SSE	rmsqe
1974	0.9969	0.9965	370.9419	3.7772
1975	0.9944	0.9937	627.3000	4.9119
1976	0.9951	0.9945	589.3737	4.7611
1977	0.9956	0.9951	497.9388	4.3762
1978	0.9963	0.9958	452.2148	4.1705
1979	0.9953	0.9948	622.0257	4.8912
1980	0.9945	0.9939	744.3498	5.3506
1981	0.9950	0.9946	567.2977	4.5838
1982	0.9934	0.9926	717.1547	5.2519
1983	0.9961	0.9956	413.8602	3.9897
1984	0.9955	0.9950	463.0218	4.2200
1985	0.9957	0.9953	433.3126	4.0824
1986	0.9970	0.9966	296.0042	3.3741
1987	0.9970	0.9967	281.3549	3.2896
1988	0.9952	0.9946	434.1463	4.0863
1989	0.9956	0.9950	349.1183	3.6644
1990	0.9946	0.9939	431.4076	4.0734
1991	0.9906	0.9895	716.9482	5.2512
1992	0.9910	0.9904	631.4419	4.8360
1993	0.9918	0.9912	536.6308	4.4582
1994	0.9872	0.9862	794.7946	5.4256
1995	0.9885	0.9876	691.3476	5.0602
1996	0.9806	0.9792	1196.6	6.6572
1997	0.9781	0.9765	1258.4	6.8271
1998	0.9681	0.9658	1715.8	7.9717
1999	0.9584	0.9553	2027.9	8.6665
2000	0.9515	0.9479	2063.9	8.7430
2001	0.9607	0.9577	1798.1	8.1608

Table 10.2 Statistics for fitting age specific fertility schedules with the gamma density model.

https://doi.org/10.26219/heal.aueb.5600

Year	R^2	R^2 adjusted	SSE	rmse
1974	0.9965	0.9961	423.2883	4.0349
1975	0.9936	0.9928	715.7248	5.2467
1976	0.9948	0.9942	630.3813	4.9240
1977	0.9951	0.9945	559.0213	4.6369
1978	0.9958	0.9953	511.7128	4.4364
1979	0.9941	0.9935	796.0121	5.5332
1980	0.9941	0.9935	796.0121	5.5332
1981	0.9942	0.9935	654.0502	5.0156
1982	0.9925	0.9916	812.4232	5.5899
1983	0.9956	0.9951	469.0740	4.2475
1984	0.9950	0.9944	520.7204	4.4752
1985	0.9953	0.9948	473.3038	4.2666
1986	0.9967	0.9963	326.3645	3.5429
1987	0.9968	0.9964	305.0452	3.4253
1988	0.9951	0.9945	446.0175	4.1418
1989	0.9955	0.9950	351.2652	3.6756
1990	0.9945	0.9939	436.1442	4.0957
1991	0.9906	0.9895	718.4297	5.2566
1992	0.9911	0.9901	626.2200	4.9077
1993	0.9920	0.9911	525.4014	4.4953
1994	0.9886	0.9873	704.6524	5.2060
1995	0.9894	0.9882	633.3052	4.9354
1996	0.9842	0.9824	974.3936	6.1218
1997	0.9826	0.9805	1001.4	6.2060
1998	0.9740	0.9710	1402.3	7.3439
1999	0.9651	0.9610	1702.2	8.0913
2000	0.9591	0.9544	1738.2	8.1764
2001	0.9675	0.9638	1485.5	7.5588

Table 10.3 Statistics for fitting age specific fertility schedules with Hadwiger function.

https://doi.org/10.26219/heal.aueb.5600

Year	Coale-Trussel	Gamma	Hadwiger
1974	0.9967	0.9969	0.9965
1975	0.9954	0.9944	0.9936
1976	0.9952	0.9951	0.9948
1977	0.9962	0.9956	0.9951
1978	0.9964	0.9963	0.9958
1979	0.9956	0.9953	0.9941
1980	0.9958	0.9945	0.9941
1981	0.9947	0.9950	0.9942
1982	0.9938	0.9934	0.9925
1983	0.9961	0.9961	0.9956
1984	0.9958	0.9955	0.9950
1985	0.9952	0.9957	0.9953
1986	0.9964	0.9970	0.9967
1987	0.9952	0.9970	0.9968
1988	0.9918	0.9952	0.9951
1989	0.9920	0.9956	0.9955
1990	0.9885	0.9946	0.9945
1991	0.9819	0.9906	0.9906
1992	0.9743	0.9910	0.9911
1993	0.9735	0.9918	0.9920
1994	0.9610	0.9872	0.9886
1995	0.9648	0.9885	0.9894
1996	0.9471	0.9806	0.9842
1997	0.9398	0.9781	0.9826
1998	0.9254	0.9681	0.9740
1999	0.9161	0.9584	0.9651
2000	0.9078	0.9515	0.9591
2001	0.9180	0.9607	0.9675

Table 10.4. Values of R^2 for the fitted models for each year.

https://doi.org/10.26219/heal.aueb.5600

	Coale	Gamma	Hadwiger
1974	395.0472	370.9419	423.2883
1975	508.4736	627.3000	715.7248
1976	579.6264	589.3737	630.3813
1977	437.4935	497.9388	559.0213
1978	429.4575	452.2148	511.7128
1979	584.7859	622.0257	796.0121
1980	568.0730	744.3498	796.0121
1981	597.3240	567.2977	654.0502
1982	668.0743	717.1547	812.4232
1983	412.7047	413.8602	469.0740
1984	438.4896	463.0218	520.7204
1985	488.3464	433.3126	473.3038
1986	351.0207	296.0042	326.3645
1987	457.7927	281.3549	305.0452
1988	738.6746	434.1463	446.0175
1989	629.9135	349.1183	351.2652
1990	913.5761	431.4076	436.1442
1991	1379.7	716.9482	718.4297
1992	1807.7	631.4419	626.2200
1993	1743.9	536.6308	525.4014
1994	2417.7	794.7946	704.6524
1995	2110.4	691.3476	633.3052
1996	3266.3	1196.6	974.3936
1997	3457.2	1258.4	1001.4
1998	4018.0	1715.8	1402.3
1999	4087.9	2027.9	1702.2
2000	3920.4	2063.9	1738.2
2001	3746.3	1798.1	1485.5

Table 10.5. Sum of squared errors for all the fitted models for each year.

https://doi.org/10.26219/heal.aueb.5600

https://doi.org/10.26219/heal.aueb.5600

Chapter 11

Conclusions

A variety of models have been proposed for modelling the age specific fertility pattern, some of them met with success while some other did not perform as succesful. Parametric models such as Coale-Trussel model, Gamma and Hadwiger functions are capable of an adequate fit for the age specific fertility pattern. Non-parametric models such as splines can produce excellent results but they require the use of too many parameters. The general result though is that we are able to attain the pattern of fertility by using only three or four parameters. We can choose whether we would use a model with demographical interpretation of the parameters and of the mechanism of fertility such as Coale-Trussell model or use a simpler model which achieves the same objective but without having a straight demographical interpretation of its parameters or giving an explanation of the procedure.

New horizons for future research give the slightly changed pattern of fertility that was recently observed in England and Wales. The previously proposed models lost their good performance and therefore a need for new models had been raised. A mixture of Hadwiger functions seems to have a good performance and also gives an interpretation of the parameters, but has the drawback of involving many parameters. This gives a push for further research to produce new models to describe both the changed pattern of fertility and the previously observed one, without a significant addition of parameters.

Another field of further research could be made is the estimation procedures Many models, in their originally proposed formation, were fitted by the method of moments, which does not give any justification for the choice of parameter estimates. The comparative studies though, used ordinary least squares to fit most models, even those which were originally proposed with inferior methods of fitting. Weighted least squares is a method that would improve the fit of the existing models.

The reciprocals of the empirical fertility rates could accurately be used as weights. This will mean that large weights are given at the centre of the childbearing period and smaller at the sides resulting to more accurate fits at the centre of the period and less accurate at the sides. This kind of fitting

https://doi.org/10.26219/heal.aueb.5600

would be more useful for birth projections where larger error at the centre of childbearing period would mean larger error in the forecasting of future births so the error should be analogues to the size of the rates.

https://doi.org/10.26219/heal.aueb.5600

References

Brass, W. (1960). The graduation of fertility distributions by polynomial functions *Population Studies* 14 (148-162)

Booth, H. (1984). Transforming Gompertz's Function for Fertility Analysis: The Development of a Standard for the Relational Gompertz Function". *Population Studies* 38(3):495-506.

Chandola, T., Coleman, D. A., Hiorns, R. W. (1999). Recent European fertility patterns: fitting curves to "distorted" distributions. *Population Studies*, Vol. 53, No. 3, Nov 1999. 317-29 pp. London, England. In Eng.

Coale, Ansley and Trussell, T. James. (1978). Technical Note: Finding the Two Parameters that Specify a Model Schedule of Marital Fertility. *Population Index* 44: 203-213.

Coale, Ansley J. (1971). Age pattern of marriage *Population Studies* 25 (2): 193-214

Coale A.J. and **Trussell T.J., (1974).** Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations. *Population Index* 40:185-258.

Coale Ansley, Trussell James, (1978). The Development and Use of Demographic Models *Population Studies,* 50 (1996), 469-484

McNeil D.R., Trussell T.J., and Turner J.C., (1977). "Spline Interpolation of Demographic Data". *Demography* 14(2):245-252.

Dunson, D. B. and Weinberg, C. R. (2000). Modeling human fertility in the presence of measurement error. *Biometrics* 56, 288-292.

Farid S.M. (1973). On the pattern of Cohort Fertility *Population Studies*, Vol. 27, No. 1 159-168.

Gilje. E. (1969). Fitting curves to age-specific fertility rates. Statistical Review of the National Centus Bureau of Statistics of Sweden Third Series , Vol 118-134

Henz, Ursula; Huinink, Johannes. (1999). *Problems concerning the parametric analysis of the age at first birth. Mathematical Population Studies,* Vol. 7, No. 2, 1999. 131-45 pp. Amsterdam, Netherlands. In Eng. with sum. in Fre.

Hoem, J.M. Madsen D., Nielsen J.L., Ohlsen E-M, Hansen H.O., and B. Rennermalm, (1981) Experiments in Modelling Recent Danish Fertility Curves. *Demography* 18(2):231-244.

Mazur D.P. (1963) A demographic model for estimating age order specific fertility rates, *Journal of the American statistical association* 58 pp 774-788

D.R. McNeil, T.J. Trussell, and **J.C. Turner, (1977)**. "Spline Interpolation of

https://doi.org/10.26219/heal.aueb.5600

Demographic Data". *Demography* 14(2):245-252.

McIntosh, James (1999). An analysis of reproductive behaviour in Canada: results from an intertemporal optimizing model. *Journal of Population Economics*, Vol. 12, No. 3, Aug 1999. 451-61 pp. Berlin, Germany. In Eng.

Mitra, S. (1967).'The pattern of age-specific fertility rates *Demography* 4 894-906

Mitra, S., and Romaniuk, A. (1972). Pearsonian type I curve and its fertility projection potentials *Demography* 10 (3): 351-365

Murphy Edmund M.; Nagnur Dhruva N. (1972). A Gompertz fit that fits: Application to Canadian Fertility Patterns *Demography,* Vol. 9, No 1 (Feb., 1972), 35-50.

Page H.J. (1977). Patterns Underlying fertility schedules: A decomposition by both Age and marriage duration *Population Studies*, Vol 31, No 1 (Mar 1977),85-106

Pandey, Arvind; Dwivedi, S. N.; Mishra, R. N. (1998). A stochastic model for the study of last closed birth interval with some biosocial components. *Mathematical Population Studies,* Vol. 7, No. 1, 1998. 1-27, 109 pp. Amsterdam, Netherlands. In Eng. with sum. in Fre.

Romaniuk A. (1973). A three parameter model for birth projections *Population Studies* (27): 467-470

Schmertmann Carl P.(2003). A system of model fertility schedules with graphically intuitive parameters ,*Demographic Research* ,volume 9, article 5

Schmertmann, Carl P.(1999) Fertility estimation from open birth-interval data. *Demography*, Vol. 36, No. 4,. 505-19 pp. Silver Spring, Maryland. In Eng.

Shanmugam, Ramalingam. (1994). *Statistical modeling of fertility: review.* In: Studies in applied demography: proceedings of the 5th International Conference on Applied Demography, 1994, edited by K. Vaninadha Rao. 1996. 65-75 pp. Bowling Green State University, Department of Sociology, Population and Society Research Center: Bowling Green, Ohio. In Eng.

Schmertmann, Carl P.; Junqueira Caetano, André. (1999). Estimating parametric fertility models with open birth interval data. *Demographic Research*, Vol. 1, 1999. Rostock, Germany. In Eng.

https://doi.org/10.26219/heal.aueb.5600