Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Demographics for Epidemiology Forecasting

Promotor :
Prof. dr. Niel HENS

Promotor :
Prof. OLIVIER LEJEUNE

Marcelius Atanga

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Demographics for Epidemiology Forecasting

Promotor :
Prof. dr. Niel HENS

Promotor :
Prof. OLIVIER LEJEUNE

Marcelius Atanga

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Modeling Male and Female Demographics for Epidemiology Forecasting

Marcelius Atanga

Internal supervisor: Prof. dr. HENS Niel

External supervisor: Prof. LEJEUNE Olivier

September 10, 2014

Contents

Certification

This is to certify that this report was written by Marcelius Atanga under our Supervision.

*Internal supervisor:*Prof. dr. HENS Niel

*External supervisor:*Prof. LEJEUNE Olivier

Signature ....

Date...............

Signature ....

Date................

..

Acknowledgments

I express my sincere thanks and gratitude to my supervisors Prof. dr. HENS Niel and Prof. LEJEUNE Olivier for their guidance, patience, tolerance and cooperation; the entire teaching staff of the department of Biostatistics, administrative staff of UHasselt,and Ensoy chellafe for their contributions and encouragements. Exspecially Prof.Joost Weyler for his understanding and moral support.

..

Abstract

Demographic forcasting is a very important tool to support societal decisions and a fundamental approach to the understanding of human society. In order to understand demography, demographers study the components of demographics namely mortality,migratiom and fertility.These components are then used to make demographic projections both for short and long term purposes.This project focuses on modeling male and female demographics for epidemiology forecasting. In order to achieve this different statistical models were fitted on the data for Belgium demography obtained from Eurostat database, with the aim to develop models that best capture the trends in the evolution of the demographic components for more accurate projections.our results show that,the composition of three exponential components (Exp2b) was the best model for mortality compared to a single exponential component (Exp1a) and the composition of two exponential components (Exp1b) for both male and female.The sum of an exponential and two Gaussian curves (Exp2G) performs best compared to the sum of an exponential and three Gaussian components (Exp3G), and the sum of an exponential and modified Gaussian curves (ExpMod2G) for both male and female populations as well. The sum of three Gaussian components (Gaussian 3) was considered the best for fertility compared to polynomial of degree four and polynomial of degree five in terms of BIC. Females have a higher life expectancy at birth (80.52years) compare to (75.07years) for males, there are more men on average approximately 51% in the total population than females approximately 49%, female migrate more with a migration peak between 18-22Years of age with an average of about 600-800 migrants per year

as compared to the male with a migration peak between 22-25years of age with an average of 400-600migrants per year. We observed that there seem to be gradual but persistent shift to the right in the peak of fertility curve for both males and females from age approximately 25 years in 1965 to to approximately 31years in 2012.

1 Introduction

1.1 Literature Review

Demography originally considered to be “science of human population” is a fundamental approach to the understanding of human society. The primary objectives are to ascertain the number of people in a given area, with respect to some time frame, determine what change have taken place from a previous census, and to estimate the future trends of population changes. Hence it is the statistical study of human populations (Leeson, 2011). According to Leeson, such study involves the measurement of the so-called demographic components such as fertility, mortality and migration in well-defined populations and understanding them is as important as measuring them accurately.

With advances in technology the ability and capacity to collect and store data more accurately is enhanced, thus it is reasonable to deduce that statistical literature will offer a continuous accumulation of knowledge. According to Keilman (2008) one can make a reasonable assumption that this progress has led to more accurate demography forecasts. Thus societal decisions that depends greatly on demography forecast such as pension planning, future health care and the spread of diseases, socio-economic and political projections has been improved upon. According to Ensoy (2013) Central to any demography forecast are the assumptions made about population dynamics and simplistic assumptions for migration, birth(fertility), and death(mortality) (proportional to population size) are untenable thus fueling need for more flexible models. “While forecasts are surprise-free, reality is not”. Hence the rapid fall In fertility in many western countries in the 1970s took demographers as a suprise (Keilman, 2004). The complexity with which demography components and their evolution are measured and projected pose serious call for concern. In the face of these challenges mathematical models for population dynamics have been developed dating back from Malthus model to Lotka-McKendrick model (Kim, 2003). As stated by Manuelli (2007) the important links between economic development and population growth are the most obvious focus, but other demographic topics, such as urbanization and migration, the composition of the population by age and sex, and family formation, are also relevant to development objectives. In this regard the accuracy of demography (or its components) projection cannot be over emphasis. In the following Section the objective of this project is given , this will be followed by the data description in Section 2, in Section 3 the statistical methodology used is outlined, results obtained are summarized in Section 4, and in Section 5 a general discussion and some conclusions are made.

1.2 Objective

The aim of this project is to develop a two-sex demography model that will perform accurate population projections over decades with a minimal set of parameter inputs. In this regard the different components; fertility, mortality and migration are considered separately.

2 Data

The data used for this study is taken from the Eurostat data of Belgium

$http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.$ The total population is the number of people as of January 1. for that year, the age at death is considered to be years from birth up to last birth day. Only live births by mother's age reached during the year are considered. For mortality and fertility only data from 1965-2012 is considered and data for new borns male and female was approximated. Migration (Human movement) here is considered to be the difference between the total number of emigrants(leaving Belgium with the intend to settle permanently in another country abroad) immigrants(entering Belgium with intend to settle permanently or as a future citizen).The data considered for migration is from 1965-2006, since migration data was availiable only up to 2006.

..

3 Methodology

In this project, the mortality , fertility and migration models for aggregated population (male and female) suggested by Ensoy (2013) are considered on males and females separately. Guided by some basic exploration plots, the mortality model is fitted in two stages, first over age and then over years, while the migration and fertility models are fitted in one stage.

3.1 Mortality

Though over the years several statistical models have been developed for capturing mortality and projections of future mortality Hollmann (2000), Deaton (2004), and Lee-Carter (2007). These models come with a lot of setbacks such as untenable assumptions leading to inaccurate demographic descriptions. Thus, the need to fine tune these models or develop new variants to accurately capture human mortality and hence improve population projections for both short and long term. According to Ensoy (2013) describing mortality over age group presents a population of seemingly three subpopulations, suggesting that the sum of 3 exponential functions out performs all other attempted mortality models for the aggregated population ignoring gender. We try to verify if this is true for the males and females population separately .In this regard we consider three different models for population mortality, these models are fitted for the aggregated population, male and female populations separately. The best model based on BIC is then considered for stage two. The models considered for the first stage are :

$$Model1(Exp1a) : \mu_a = \exp(\beta_0 + \exp(\beta_1 + \beta_2 a))$$

$$Model2(Exp1b) : \mu_a = \exp(\beta_0 + \exp(\beta_1 + \beta_2 a) + \exp(\beta_3 + \beta_4 a))$$

$$Model3(Exp2b) : \mu_a = \exp(\beta_0 + \beta_1 a) + \exp(\beta_2 + \beta_3 a) + \exp(\beta_4 + \beta_5 a)$$

Simplistic assumptions have proven to be untenable for example the most popular Lee-Carter model assumes a constant mortality pattern for each age group accross years (Lee-Carter ,1992) and (Booth, 2006). This model is popular in the sense that it has been judged by many to be appriopriate in all most all cases and more complex models are not very different in results (Booth, 2006). To account for variabilty accross the years Ensoy (2013) introduced the two stage model, fitted first over age and then over the years. Fitted models from stage two are compared with the actual parameters obtained from stage one for the different years.

3.2 Migration

Migration which is considered in this context as the net flow of people in to a country with the intend to stay or of becoming future citizens has been earmarked by many demographers as the most complex element of the three demography elements(mortality,fertility ,and migration). Stemming from the fact that migrations is a composition of two distinct component namely emigration (people going out) and immigration (people entering) with each of these component depending on several other facts that do not always overlap, makes it very difficult to develop a statistical model that captures the proper evolution and trends over time. This can be attributable to the lack of time series of reliable and detailed data that represent the actual migration flow (Ensoy, 2013). According to Greenwood (2006), this is also due to the fact that migration models are abstract depictions and simplifications of real-world processes that may or may not be expressed mathematically, such as political strife, religious differences and other social phenomena that leads to voluntary as well as involuntary movements. Unfortunately most or all migration models begin with the assumption that migration is a voluntary human act, on the contrary most migration is involuntary (Greenwood, 2006). However in order to be able to address the elements of emigration and immigration in population forecasting, De Beer (1997) found consistency between time series forecast of total immigration, emigration and net migration. Due to the fact that migration depends on many factors such as economy, social security, religion, political stability, age, gender, marital status, type of migration, distance from origin to destination, and more with complex evolutions, most migration models have too many parameters leading to difficulties in forecasting. However most recent authors including Keilman (2004), Rogers (2005), and Ensoy (2013) have developed alternative models with reduced set of parameters offering better possibilities for forecasting. In this project we focus on the models suggested by Ensoy (2013) for the aggregated population and compare with for the total population, male and female subpopulations separately. In this regard the models considered are :

$$Model4(Exp2G) : M_a = \beta_1 \exp(-\beta_2 a) + \beta_3 \exp\left[-(\tfrac{a-\beta_4}{\beta_5})^2\right] + \beta_6 \exp\left[-(\tfrac{a-\beta_7}{\beta_8})^2\right]$$
$$Model5(Exp3G) : M_a = \beta_1 \exp(-\beta_2 a) + \beta_3 \exp\left[-(\tfrac{a-\beta_4}{\beta_5})^2\right] + \beta_6 \exp\left[-(\tfrac{a-\beta_7}{\beta_8})^2\right] + \beta_9 \exp\left[-(\tfrac{a-\beta_{10}}{\beta_{11}})^2\right]$$
$$Model6(ExpMod2G) : M_a = \beta_1 \exp(-\beta_2 a) + \beta_3 \exp\left[-(\tfrac{a-\beta_4}{\beta_5})^2\right] + \beta_6 \left\{\exp\left[-(\tfrac{a-\beta_7}{\beta_8})^2\right] - \beta_9\right\}$$

In this scenario, migration for each age group is considered constant over the years.

3.3 Fertility

Fertility in this context is considered to be the rate at which women give birth. As oppose to fecundity which is the potential for reproduction (contrary to sterility).According to Manuelli (2007). Fertility rates varies significant across countries and regions. This variability is attributable to life expectancy, years of schooling, income, output per worker. Though the study of fertility models date as far back as 1933 in the works of Pascal Whelpton, more has been done to this effect but the approach is still unsatisfactory for example ,does not consistently reflect uncertainty for forecast variable (Lee, 1999).Because of these many statistical agencies in recent years attempted to compute stochastic population forecast with predictive intervals (Hanika, 1997). According to Ensoy (2013) this stems from the fact that fertility and births are non-stationary processes with structural changes in the fertility trajectory, changing age patterns and the complex association between the two. This complexity can be linked to the role of latent variables such as years of school, income, output per work, life expectancy, migration, just to mention a few. An attempt to address this complexity has been made by using time series methods on dynamic factor models Ortega (2005). Lee-Carter use a similar method to deal with structural changes by specifying long term mean value of total fertility and imposing limits Lee (1993). Here the sum of three Gaussian models as suggested by Ensoy (2013) for the aggregated population together with two other models are used to capture the fertility rate stratified by gender. The models considered are:

$$Model7(Gaussian3) : \varphi_a = \beta_1 \exp\left[-\left(\frac{a-\beta_2}{\beta_3}\right)^2\right] + \beta_4 \exp\left[-\left(\frac{a-\beta_5}{\beta_6}\right)^2\right] + \beta_7 \exp\left[-\left(\frac{a-\beta_8}{\beta_9}\right)^2\right]$$

$$Model8(Polynomial4) : \varphi_a = \beta_0 + \beta_1 a + \beta_2 a^2 + \beta_3 a^3 + \beta_4 a^4$$

$$Model9(Polynomial5) : \varphi_a = \beta_0 + \beta_1 a + \beta_2 a^2 + \beta_3 a^3 + \beta_4 a^4 + \beta_5 a^5$$

4 Results

4.1 Data Exploration

Results from data exploration suggest that female children between age zero to one in Belgium have a higher survival rate ranging from 97-99.5% compared to those of males ranging from 97-98.5% for males, mean while there are more men on average approximately 51% in the total population than females approximately 49% except for those above eighty years of age, where we observe higher female population compared to male (see Figure 7.9-7.10 Appendix A) . We observe from data that fertility rate for male is higher than that for female using approximated data (see Appendix B), this explains why though women survival rate at birth is higher, there are more males in the general population than females. On average we observe that more females migrate than males and at a slightly younger age.

4.2 Mortality Results

4.2.1 Stage One

From Table 4.1, Model 3 outperforms the other two models for all three populations. Mean while for all three groups, there seems to be a gradual improvement in performance over time as from 1980 forward , this may be attributable to advances in data collection. Thus we can say that the sum of three exponential functions which suggest varying magnitude in parameters(amplitude and slope) but of similar shape curves is the best fit for mortality for aggregated population as suggested by Ensoy (2013) and for the males and females sub populations separately as shown in our results.

Table 4.1: Comparison of the different candidate models for mortality by BIC for selected years.

		BICs					
Pop	Mean Function	1966	1971	1981	1991	2001	2011
Total	Model1(Exp1a)	1709.636	1708.832	1697.901	1680.995	1671.368	1648.22
	Model2(Exp1b)	1299.792	1268.904	1259.28	1188.631	1218.289	1148.87
	Model3(Exp2b)	*1186.072*	*1203.004*	*1236.713*	*1171.815*	*1204.576*	*1123.53*
Male	Model1(Exp1a)	1590.948	1588.208	1577.953	1560.787	1554.22	1530.86
	Model2(Exp1b)	1180.939	1186.886	1190.811	1129.031	1139.624	1047.44
	Model3(Exp2b)	*1120.88*	*1138.4*	*1171.309*	*1122.826*	*1128.524*	*1031.4*
Female	Model1(Exp1a)	1544.541	1543.342	1530.297	1512.464	1499.472	1477.32
	Model2(Exp1b)	1125.678	1088.193	1057.783	971.5769	965.2652	966.251
	Model3(Exp2b)	*966.7573*	*1004.529*	*1017.118*	*936.161*	*951.9542*	*934.073*

4.2.2 Stage Two

Considering Model 3 as the best model from stage one, we start by inspecting the parameter plots as shown on Figures 4.1-4.3 for the total, male and female populations. Different distributions were imposed on the parameters over time, this is to account for variability of model parameters $(\beta_0, ...\beta_5)$ over time (Ensoy, 2013).

$$E(\beta_0) = \gamma_0 + \gamma_1 * year - - - (A)$$

$$E(\beta_1) = \gamma_0 - - - - - - - - - - (S)$$

$$E(\beta_2) = \gamma_0 + \gamma_1 * year - - - (A)$$

$$E(\beta_3) = \gamma_0 + \gamma_1 exp(\gamma_2 * exp(\gamma_3 * year)) - - - - - - - - (S)$$

$$E(\beta_4) = \gamma_0 + (\gamma_1 - \gamma_0)/(1 + \exp(-\gamma_2(year - \gamma_3))) - -(A)$$

$$E(\beta_5) = \gamma_0 + (\gamma_1 - \gamma_0)/(1 + \exp(-\gamma_2(year - \gamma_3))) - -(A)$$

From Table 4.2,β_0, β_2,and β_4 are the amplitudes of the exponential functions component. Similarly β_1, β_3,and β_5 are the slopes for the exponential functions component. Due to the highly non linear nature of the functions structures we observe some high variability for some of the second stage parameters.

Table 4.2: 2nd-stage model parameter estimates for mortality base on Model 3.

Stage 1	Stage 2	Totpop-Estimates(SE)	Malpop-Estimates(SE)	Fempop-Estimates(SE)
β_0	γ_0	-3.6901 (0.0241)	-3.5424 (0.0270)	-3.8260(0.0253)
	γ_1	-0.0444 (0.0008)	-0.0447(0.0009)	-0.0444(0.0009)
β_1	γ_0	-2.622 4(0.0515)	-2.902 5(0.0618)	-2.4292(0.0589)
β_2	γ_0	-7.5009 (0.0250)	-7.4587 (0.0349)	-7.5530(0.0340)
	γ_1	-0.0406(0.0009)	-0.034 4(0.0012)	-0.0522(0.0012)
β_3	γ_0	-0.0103 (0.0031)	-0.006 8(0.0029)	-0.0170(0.0038)
	γ_1	0.0817(0.0055)	0.0720(0.0049)	0.0944(0.0066)
	γ_2	-5.3453 (1.3237)	-6.5482(2.0127)	-6.6964(2.1159)
	γ_3	-0.0941 (0.0121)	-0.1077(0.0154)	-0.1058(0.0157)
β_4	γ_0	-9.8877(0.0916)	-12.5638(0.1231)	-10.4654(0.1304)
	γ_1	-13.727 7(0.1943)	-9.5233(0.0638)	-16.8716(0.4200)
	γ_2	0.1590(0.0205)	-0.2032(0.0264)	0.1225(0.0142)
	γ_3	30.968 3(0.8640)	33.7574(0.7101)	33.6857(1.1597)
β_5	γ_0	0.0955(0.0009)	0.0920(0.0007)	0.1016(0.0013)
	γ_1	0.1286(0.0076)	0.1198(0.0020)	0.1585(0.0038)
	γ_2	0.1823(0.0275)	0.2247(0.0359)	0.1371(0.0182)
	γ_3	30.8300(0.8947)	29.0800(0.7688)	33.1832(1.1483)

However a plot of the predicted and the obtained parameters from stage one show that the imposed distributional structure truly captures the distribution of the parameters over the years Figures 4.1 -4.3. Figures 7.11-7.12 Appendix A, shows the observed versus predicted mortality for the male and female populations, from which we observed that this mortality model captures the trends better in the younger age, with greater variability for persons of ages eighty years and above. Figures 7.11-7.12 Appendix A, also seem to suggest that as year goes by the mortality model draws closer to the observed mortality for all three populations as the ratio curve move closer to the reference line. Further investigation can be done to check why the curves wiggles in a similar manner for all populations, although history and literature suggest that this may be due to the effect of influenza after the first world war (1918 Influenza), known as the Mother of All Pandemics (Taubenberger, 2006), how ever this can not be justified since this is not a cohort study, but rather a phenomenon of very old age where the mortality stops to increase and may even decrease this suggest that at the age of eighty and above the mortality trend is not captured properly by our exponential model. In general, our model suggest that men have a higher mortality rate than women for persons above sixty years of age (see Figure 7.13, Appendix A). We also observe that women have a higher life expectancy at birth (80.52) compare to (75.07years) for men and (77.76 years) for the general population (Figure 4.4).

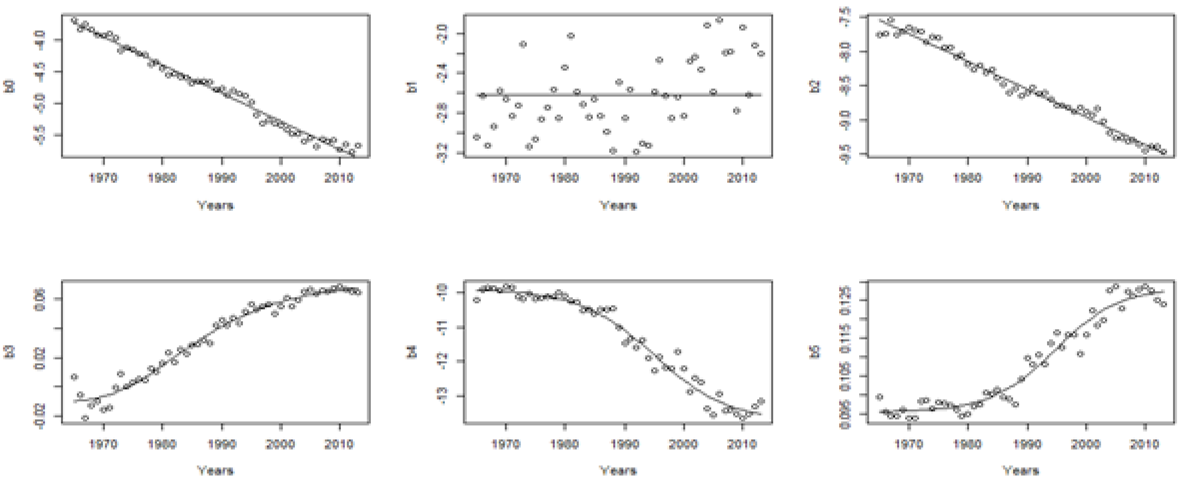

Figure 4.1: Yearly mortality parameter estimates distribution based on 1st-stage Model 3 (points) and predictions based on 2nd-stage models over time for the total population.

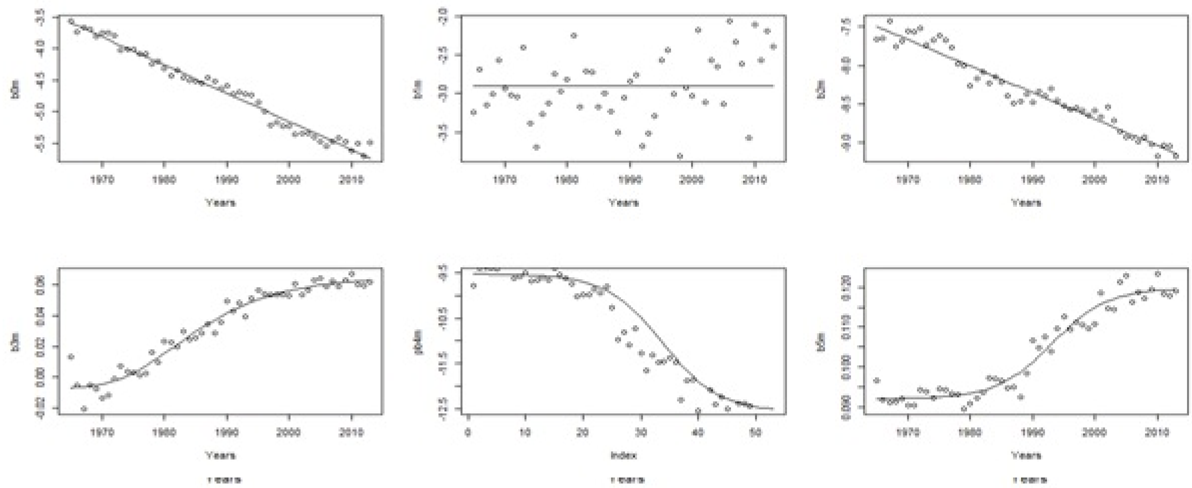

Figure 4.2: Yearly mortality parameter estimates distribution based on 1st-stage Model 3 (points) and predictions based on 2nd-stage models over time for the male population.

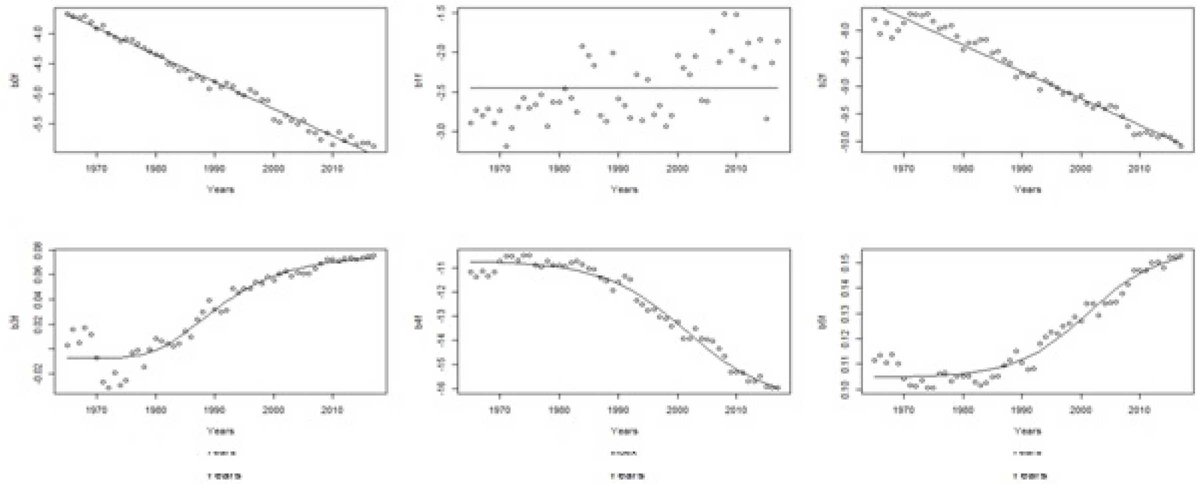

Figure 4.3: Yearly mortality parameter estimates distribution based on 1st-stage Model 3 (points) and predictions based on 2nd-stage models over time for the female population.

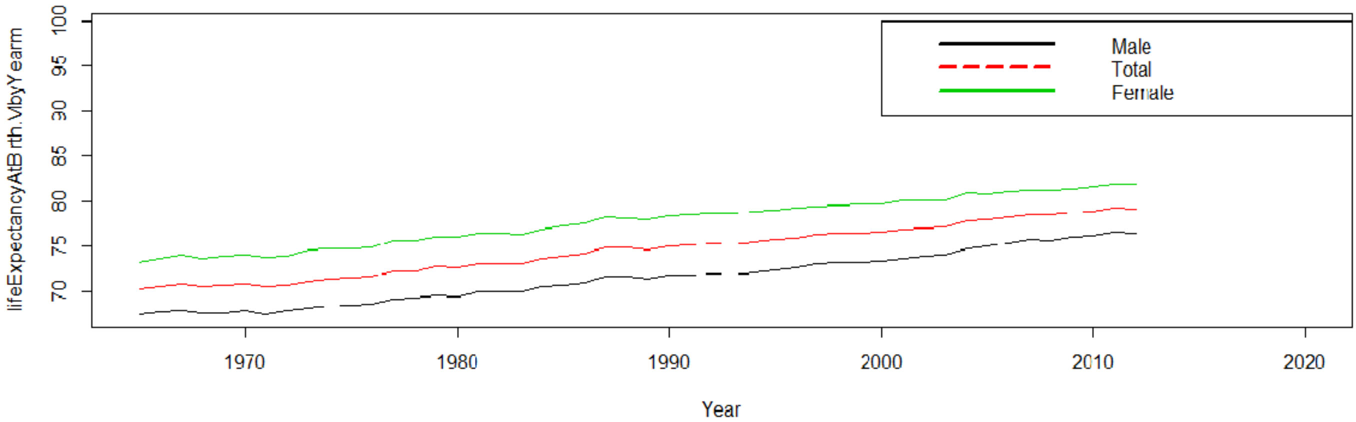

Figure 4.4: Life Expectancy for male and female over the years.

4.3 Migration Results

From Table 4.3, Exp2G performs best compared to Exp3G and ExpMod2G for the total and male populations for almost all the years. Meanwhile there seems to be a persistent competition for the female population with Exp3G emerging. In general Exp2G was selected as the best model for migration for all three populations. Comparing the fitted and observed values show that Exp2G is flexible enough to capture the age-specific distribution of net migration.

Although there seems to be some variability over the years, it is assumed constant.

Table 4.3: Comparison of different candidate models for migration using BIC for selected years for three different populations (for the Total, Male, and Female) populations separately.

Pop	Mean Function	1990	1995	2002	2003	2004	2005
Total	Exp2G	1118.818	1111.278	1142.015	1151.958	1146.24	1132.661
	Exp3G	1141.149	1111.311	1153.184	1155.202	1151.853	1141.149
	ExpMod2G	1097.399	1126.787	1146.979	1149.299	1147.17	1136.446
Male	Exp2G	1019.675	1032.426	1065.6	1061.956	1054.857	1019.294
	Exp3G	1022.844	1022.155	1074.557	1073.748	1063.005	1017.515
	ExpMod2G	1016.886	1037.044	1069.898	1064.293	1057.513	1019.259
Female	Exp2G	1029.558	1011.41	1029.821	1035.499	1044.883	1065.2
	Exp3G	1005.928	1021.31	1074.557	1034.667	1038.049	1014.298
	ExpMod2G	1007.609	1015.748	1022.584	1028.205	1045.956	1067.275

From Figure 4.5 we observed that in Belgium, females migrate more than males and also females migration attend its peak at a younger age than males, while at very tender age (0-5years) of age more males seems to be migrating compared to females. In general the migration bell-shaped curve for Belgian population has its peak between 19 -24 years of age with an average of 1200-1400 migrants per year, the female population has its peak between 18-22Years of age with an average of about 600-800 migrants per year, and the male population has its peak between 22-25years of age with an average of 400-600migrants per year.

Fitting a one stage model, parameter estimates for the model Exp2G are presented on Table 4.4. This seems to suggest that the age at which the two peaks for migration are reached are very close precisely (20.4 and 26.7) for total population, (22.3 and 27.6) for males and (19.7

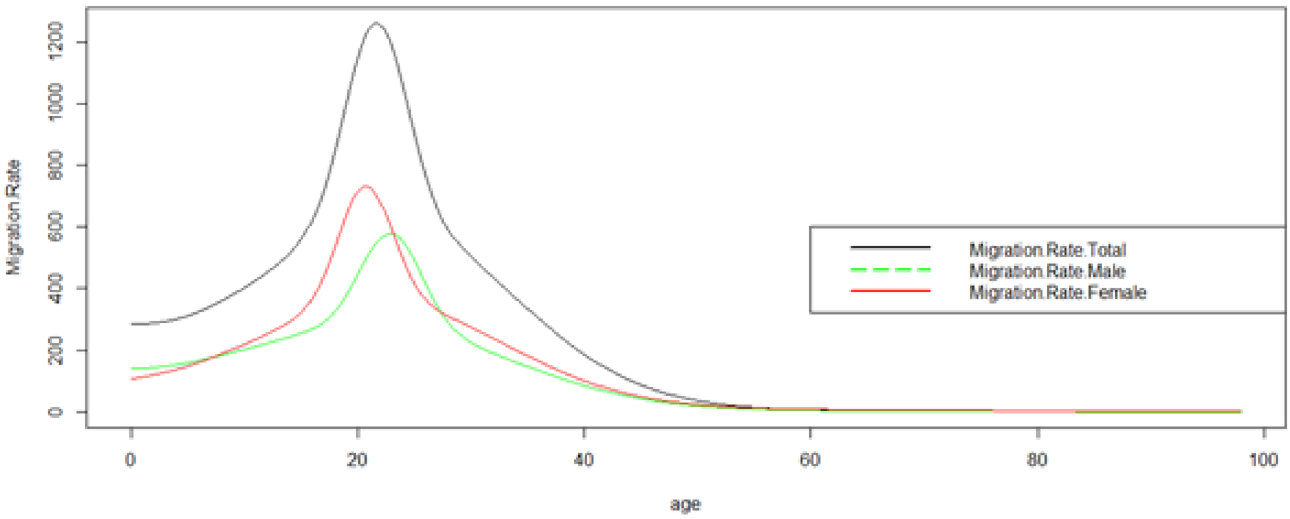

Figure 4.5: Migration plots for total population, male and female subpopulation.

and 27.8) for females for the Belgian population. In fact this is not a surprise as most young people between the ages of eighteen and twenty-four are obsessed with their education and from twenty-five to thirty are hunting for jobs or traveling for work or marriage. These estimates seems to be very stable for all three populations but much more stable for females than males as we observe lower standard errors for females compared to males and on the bootstrap confidence intervals (see Figure 7.14-7.15 Appendix A).

Table 4.4: Parameter estimate for the best migration model

	Total Pop	Male	Female
Parameter	Estimates(S.E)	Estimates(S.E)	Estimates(S.E)
b1	632.5424 (61.5593)	307.2533 (30.2928)	248.7894 (15.7356)
b2	0.1205 (0.0378)	0.1024 (0.0396)	0.05461 (0.0092)
b3	556.1965 (86.8112)	249.4611 (71.6993)	417.2211 (46.1565)
b4	20.4631 (0.2439)	22.3462 (0.4491)	19.7300 (0.1383)
b5	4.4003 (0.5667)	5.7476 (0.9946)	3.8195 (0.3549)
b6	588.9485 (66.2005)	215.9054 (65.1109)	273.5979 (21.3855)
b7	26.7288 (1.4398)	27.6464 (2.8498)	27.8456 (1.1275)
b8	11.9890 (1.0939)	12.3197 (1.8449)	9.9672 (1.2462)

4.4 Fertility Results

From Table 4.5, the sum of three Gaussian (Gaussian 3) out performs the other two models in terms of BIC for the total population, male and female subpopulations. Thus Gaussian 3 was selected as the best model for fertility model.

Table 4.5: Comparison of BICs for fertility models

Population	Mean Function	1965	1975	1995	2000	2010
Total						
	Gaussian 3	502.3173	455.4493	481.6933	390.1114	384.4763
	Polynomial 4	661.1942	646.8651	631.3162	640.2981	655.7179
	Polynomial 5	634.2833	635.7704	617.7349	617.3666	637.2215
Male						
	Gaussian 3	561.1493	460.44	399.767	461.54	379.728
	Polynomial 4	614.7863	598.146	582.5109	591.7418	607.5973
	Polynomial 5	586.4339	581.7431	566.4462	573.8841	590.8531
Female						
	Gaussian 3	576.243	462.544	449.712	460.781	367.425
	Polynomial 4	610.4681	596.041	580.5661	589.0238	603.9728
	Polynomial 5	583.702	579.5861	564.0282	571.786	586.7687

where , β_1, β_4,and β_7 define the amplitude or maximum fertility rate of each model component. The parametersβ_2, β_5,and β_8 are the mean age where the fertility rate is at its peak for each of the respective components. The parameters ,β_3, β_6,and β_9 define the standard deviation of each component.

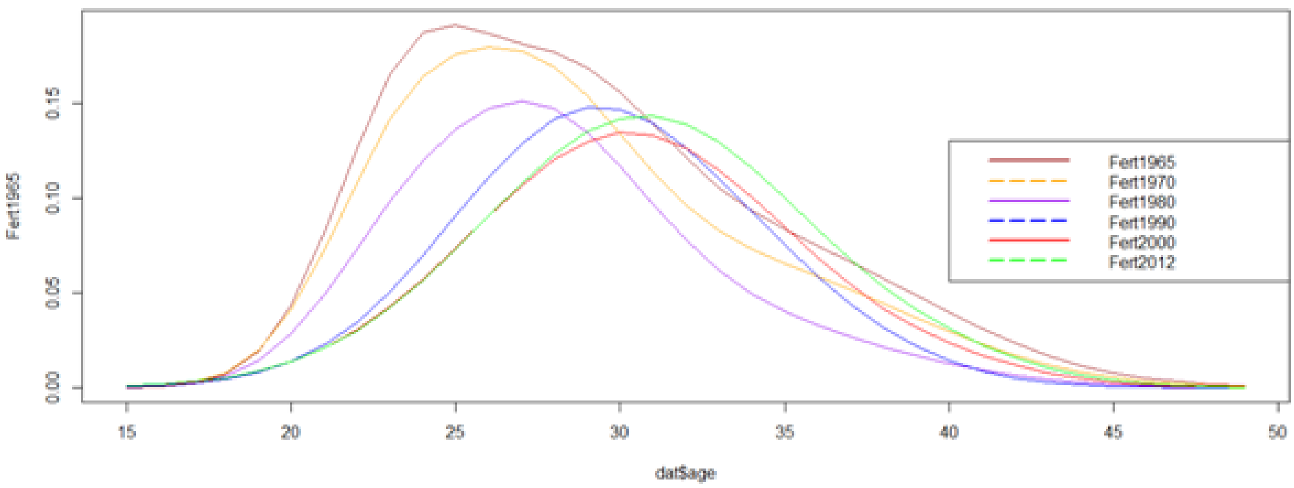

Figure 4.6: Fertility Rate for Selected Years –Total Population.

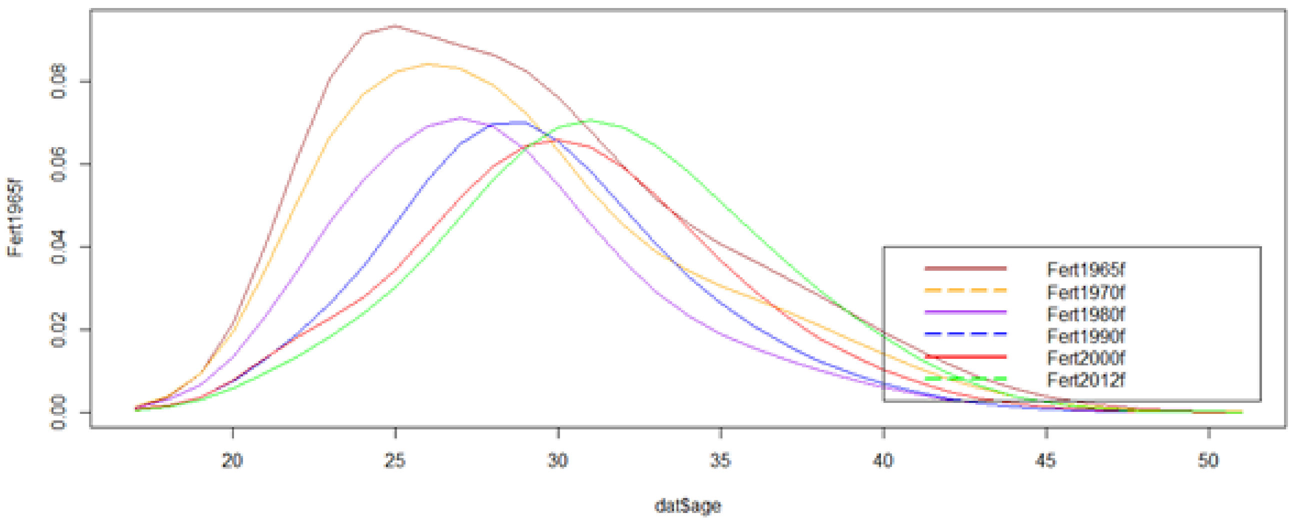

Figure 4.7: Fertility Rate for Selected Years –Female.

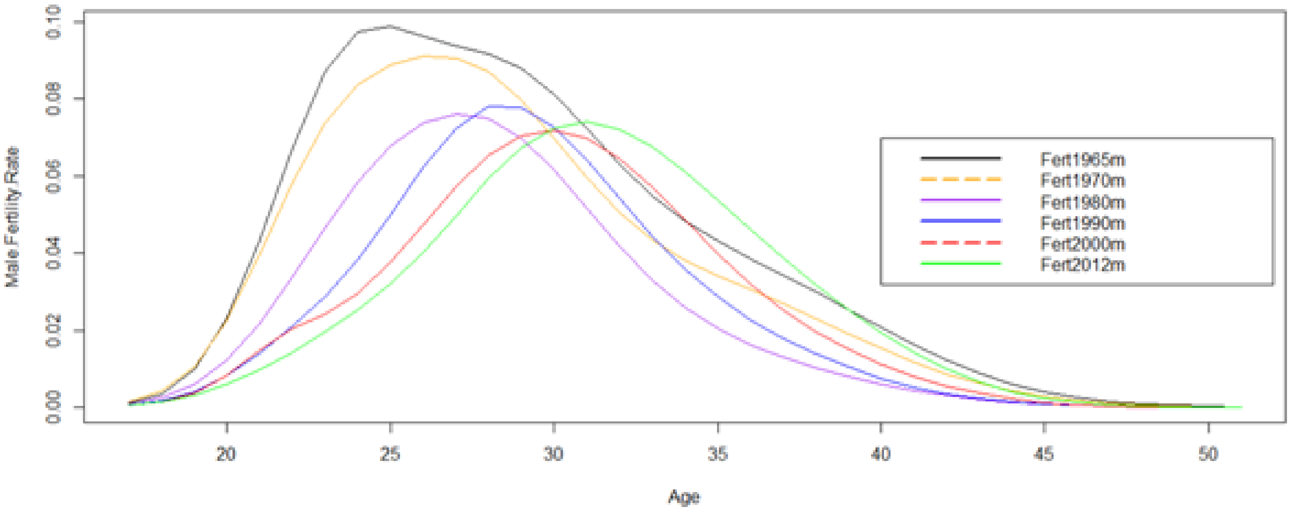

Figure 4.8: Fertility Rate for Selected Years –Male .

From Figure 4.6-4.8 we observe that there seems to be gradual but persistent shift to the right in the peak of fertility curve for both males and females , meaning more Belgian women give birth at later ages over the years. Also the same plots show a persistent decrease in fertility rate over the years for both sexes. Figure 7.17-7.19 Appendix A show the distribution of $\beta_1, ..., \beta_9$ for the total population, male and female subpopulations obtained over the years.

..

5 Discussion

The aim of this project was to model male and female demographics for epidemiology forecasting.The three components of demography were considered (mortality, migration, and fertility).The data was extracted from Eurostat demography database and our interest was on the Belgian population.We obtained data for the total number of deaths and migrants for each year per age for both male and female. For fertility we found only data for total number of new borns, we then calculated the number of newborn males and females as fraction of the males and females at age zero to one and accounting for the mortality for each year while ignoring migration. Though many studies on modeling demography for population projections have been conducted over the years, we forcused mainly on the models of Ensoy (2013) while accounting for gender.In total nine models were fitted. For mortality, we found that the composition of three exponential component (Exp2b) was the best model for mortality compared to a single exponential component (Exp1a) and the composition of two exponential component (Exp1b) for both male and female (see section:methodology Model1-Model3).The sum of an exponential and two Gaussian curves (Exp2G) performs best compared to the sum of an exponential and three Gaussian components (Exp3G), and the sum of an exponential and modified Gaussian curves (ExpMod2G) for both male and female populations as well(see section:methodology Model4-Model6). The sum of three Gaussian components (Gaussian 3) was considered the best for fertility compared to polynomial of degree four and polynomial of degree five in terms of BIC(see section:methodology Model7-Model9). Generally, we can conclude from our results that, females have a higher life expectancy at birth (80.52) compare to (75.07 years) for males. There are more men on average approximately 51% in the total population than females approximately 49%, men have a higher mortality rate than women for persons above sixty years of age, female migrate more with a migration peak between 18-22 years of age with an average of about 600-800 migrants per year as compared to the male with a migration peak between 22-25 years of age with an average of 400-600migrants per year. we observe that there seems to be gradual but persistent shift to the right in the peak of fertility curve for both males and females from age approximately 25 years in 1965 to to approximately 31years in 2012 Further investigation can be done to check why the curves wiggles in a similar manner for all populations at age 80years of corresponding to mortality of approximately 23 %. There is also need for further research on how to obtain more accurate data for newborns for both male and female populations(see Appendix B).

6 References

- De Beer, J., 1997. The effect of uncertainty of migration on national population forecasts: The case of the Netherlands. Journal of Social Statistics 13, 227-243.
- Deaton, Angus and Christina Paxson. 2004. Mortality, Income, and Income Inequality Over Time in the Britain and the United States. Technical Report 8534 National Bureau of Economic Research Cambridge, $MA : .http : //www.nber.org/papers/w8534$.
- Greenwood .M.J , 2006. "Modeling Migration." Encyclopedia of social Measurement 2, 726-734 .
- Hanika, A., Lutz, W. and Scherbov, S., 1997. 'Ein probabilistischer Ansatz zur Bevölkerungsvorausschätzung für Österreich'. Statistische Nachrichten 12, 984–988.
- Hollmann, F.W., T.J. Mulder and J.E. Kallan. 2000. "Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100." Working Paper 38, Population Division, U.S. Bureau of Census.
- Keilman, N., Pham, D.Q., 2004. Time series based errors and empirical errors in fertility forecasts in the Nordic countries. International Statistical Review 72, 5-18.
- Kim, M.Y. and Selenge,T., 2003.Age-time discontinuous galerkin method for the Lotka-Mckendrick equation. Commun.Korean Math.Soc 3,569-580.

- Lee, R.D., 1993. Modeling and forecasting the time series of us fertility: Age distribution, range, and ultimate level. International Journal of F 9, 187-202.

- Leeson,G.W., 2011. The Importance of Demography,Population Ageing.Springer Science 4,1–4.

- Manuelli, R. and A. Seshadri, 2007a "Human Capital and the Wealth of Nations," mimeo, University of Wisconsin-Madison.

- Rogers, A., Castro, L.J., Lea, M., 2005. Model migration schedules: Three al- ternative linear parameter estimation methods. Mathematical Population Studies 12, 17-38.

- Ortega, J.A., Poncela, P., 2005. Joint forecasts of southern european fertility rates with non-stationary dynamic factor models. International Journal of Forecasting 21, 539-550.

- Taubenberger, J.K, Morens, D.M., 2006. 1918 influenza: the mother of all pandemics.National Institutes of Health, Bethesda -USA.

7 Appendix

7.1 Appendix A

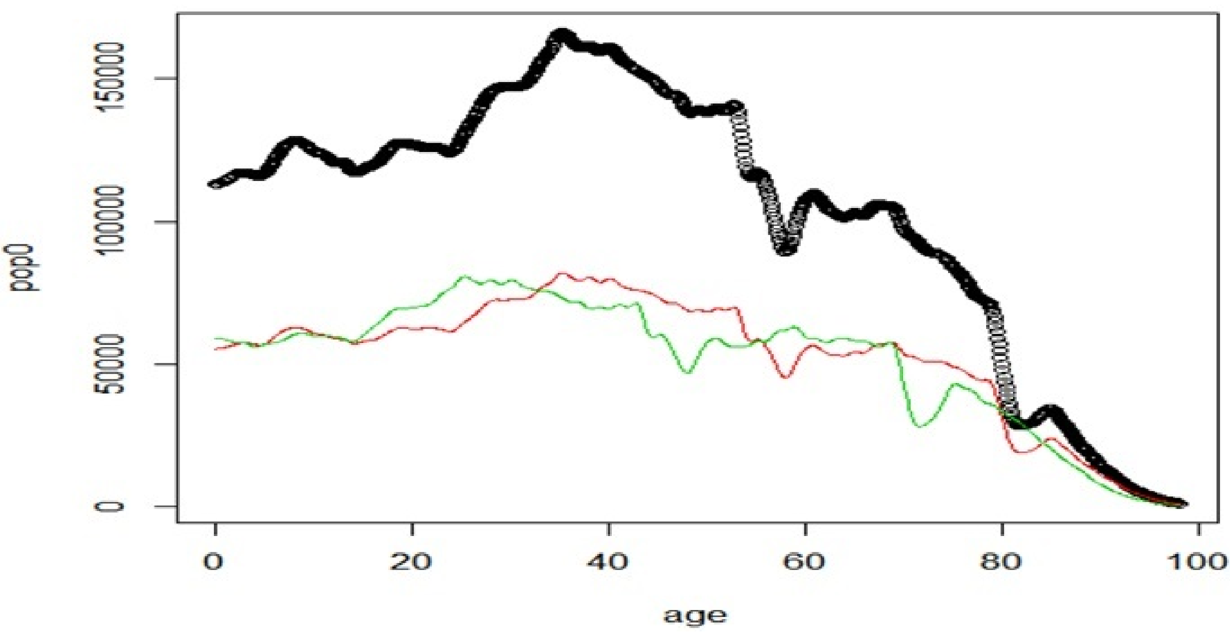

Figure 7.9: Mean Population distribution for the total poulation black points, female population red line and male population green line.

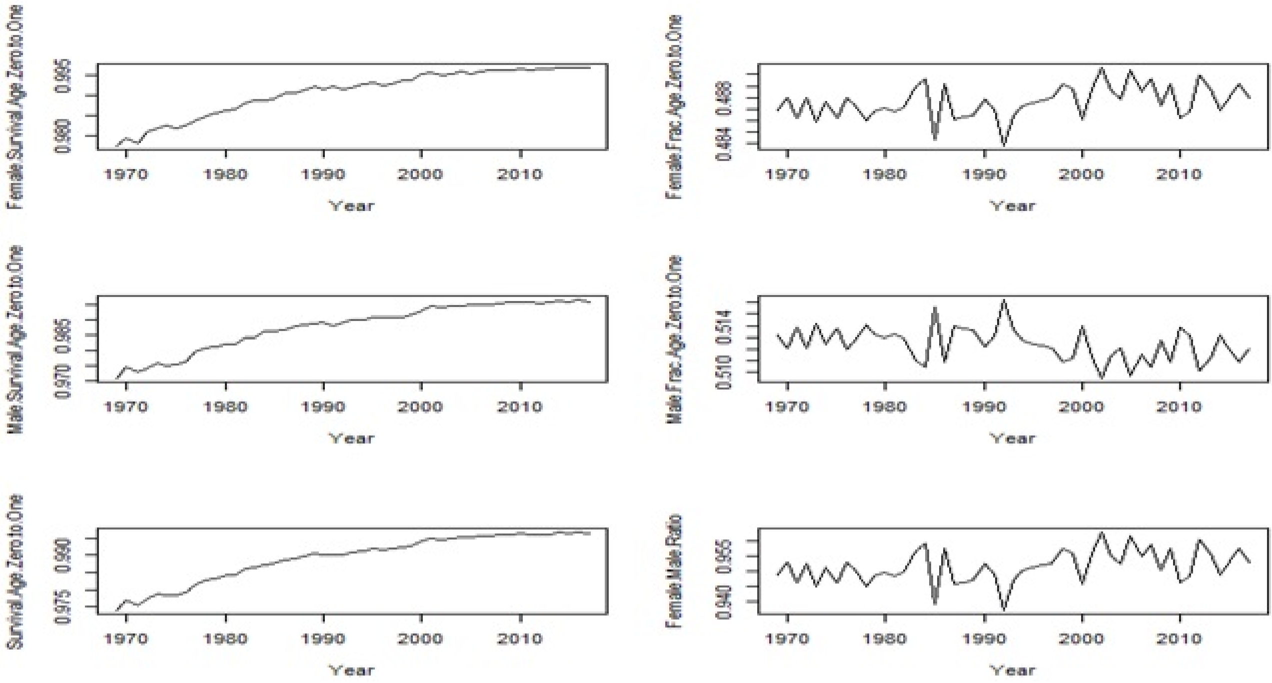

Figure 7.10: Survival rate for age group zero to one and Fraction of female and male populations for all age groups.

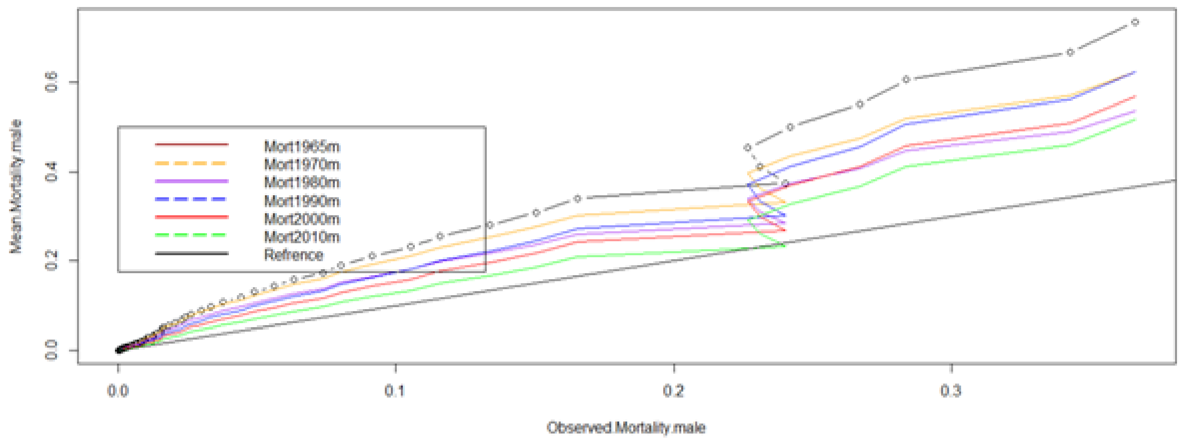

Figure 7.11: Observed vs estimated mortality for male population.

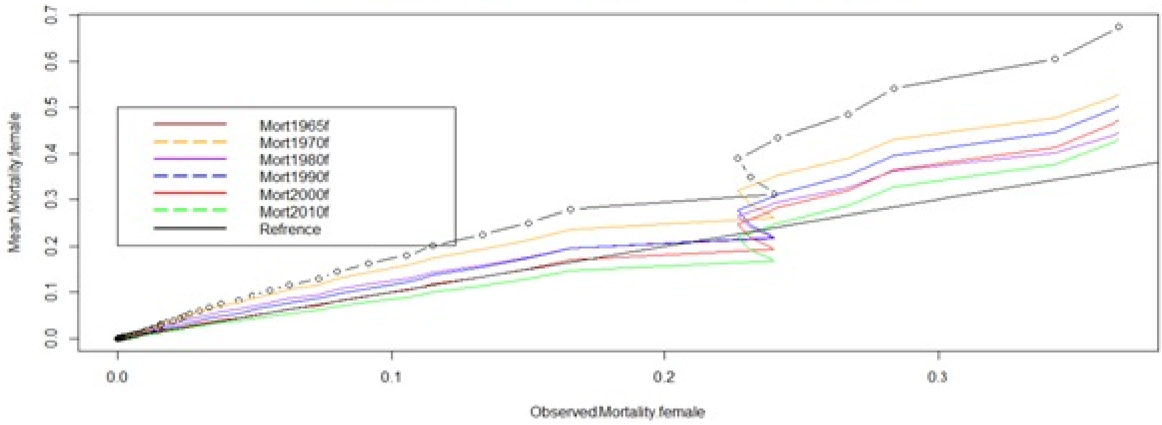

Figure 7.12: Observed vs estimated mortality for female population.

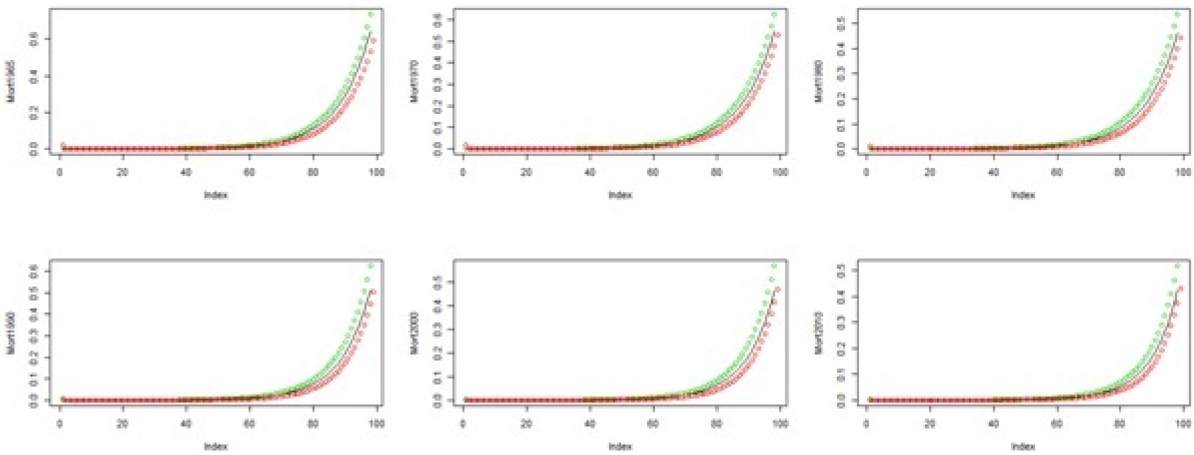

Figure 7.13: Mortality rates for total population dark line, female population red points and male population green points for selected years .

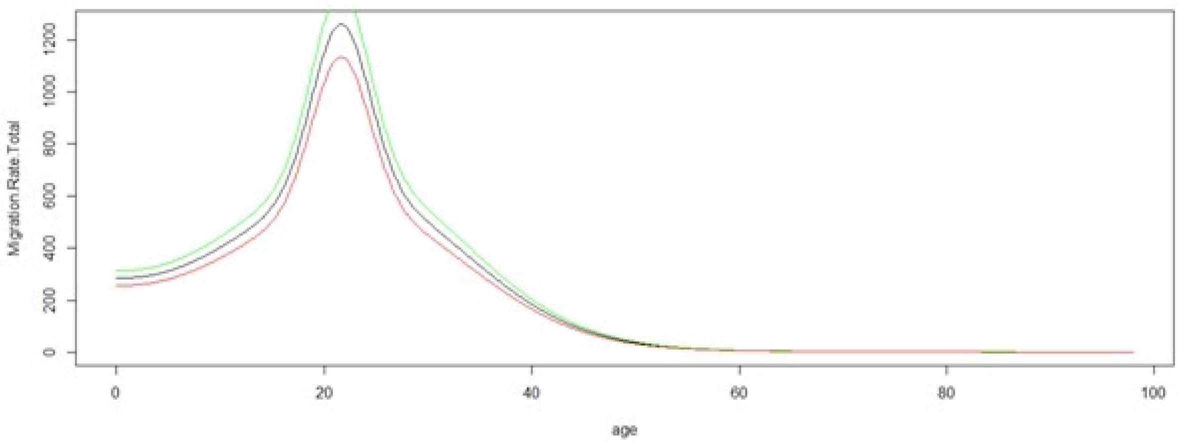

Figure 7.14: Migration for total population with bootstrap confidence intervals.

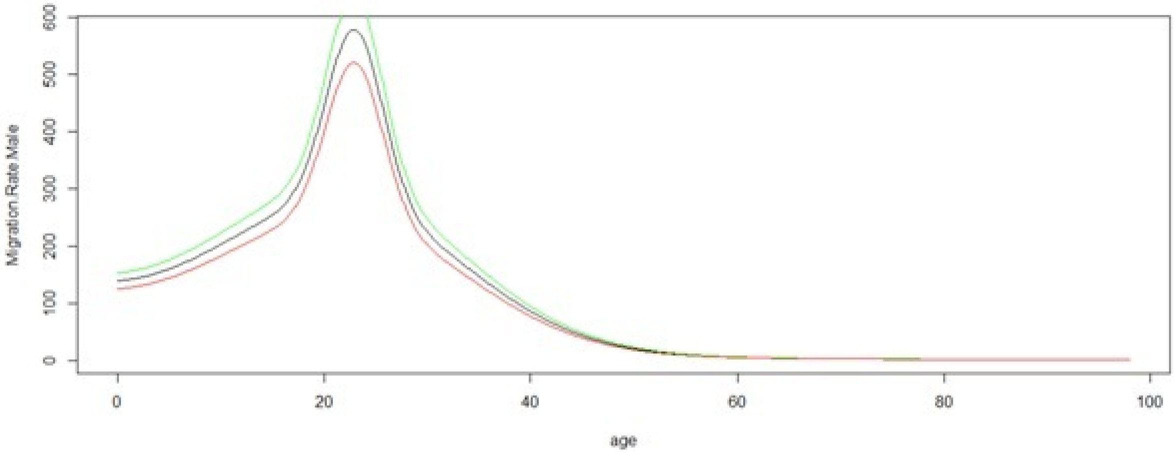

Figure 7.15: Migration for male population with bootstrap confidence intervals.

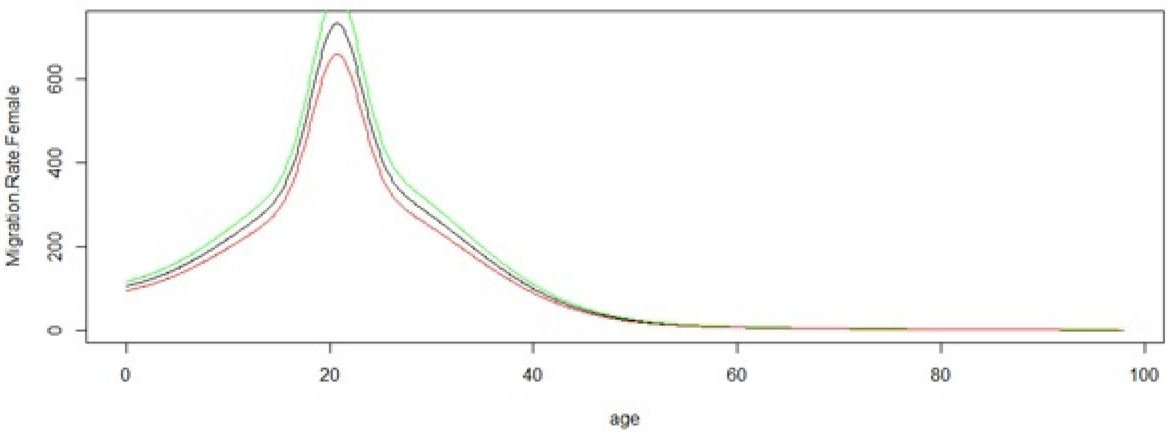

Figure 7.16: Migration for female population with bootstrap confidence intervals.

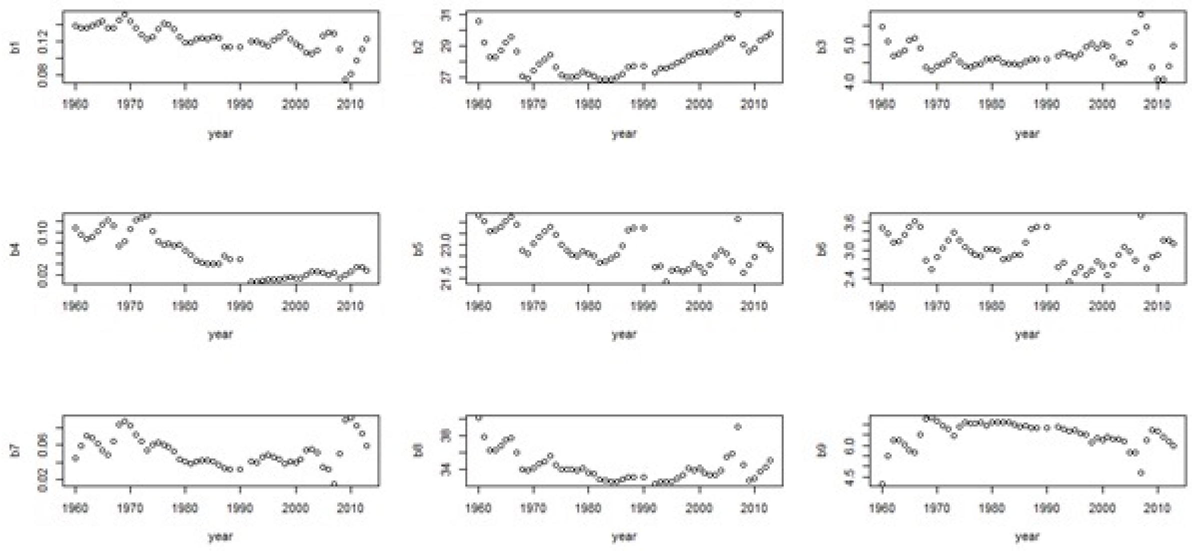

Figure 7.17: Fertility parameter estimate distribution for total population from model stage one.

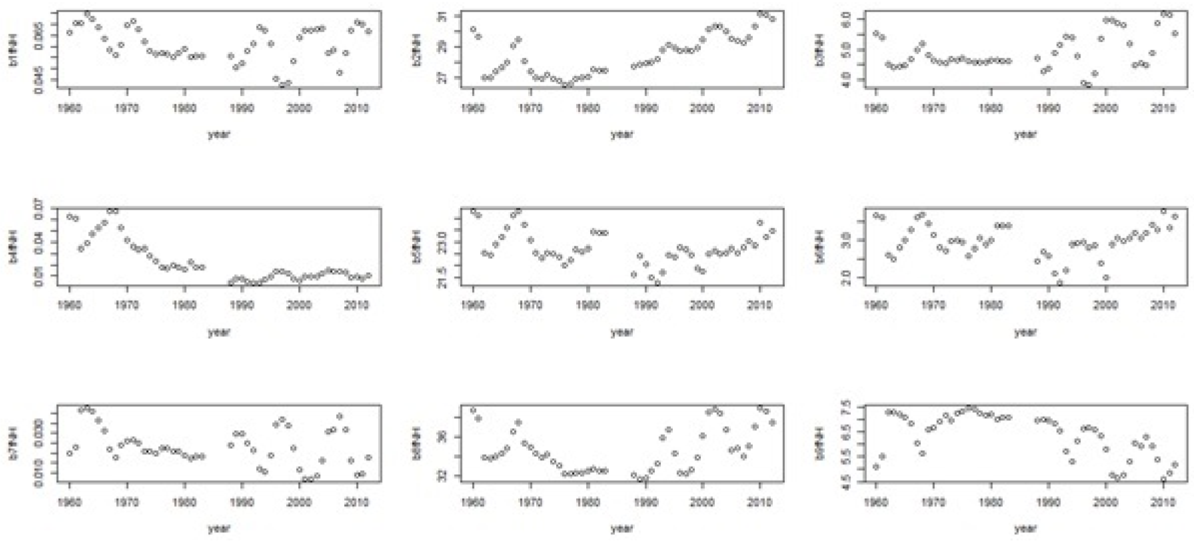

Figure 7.18: Fertility parameter estimate distribution for female population from model stage one.

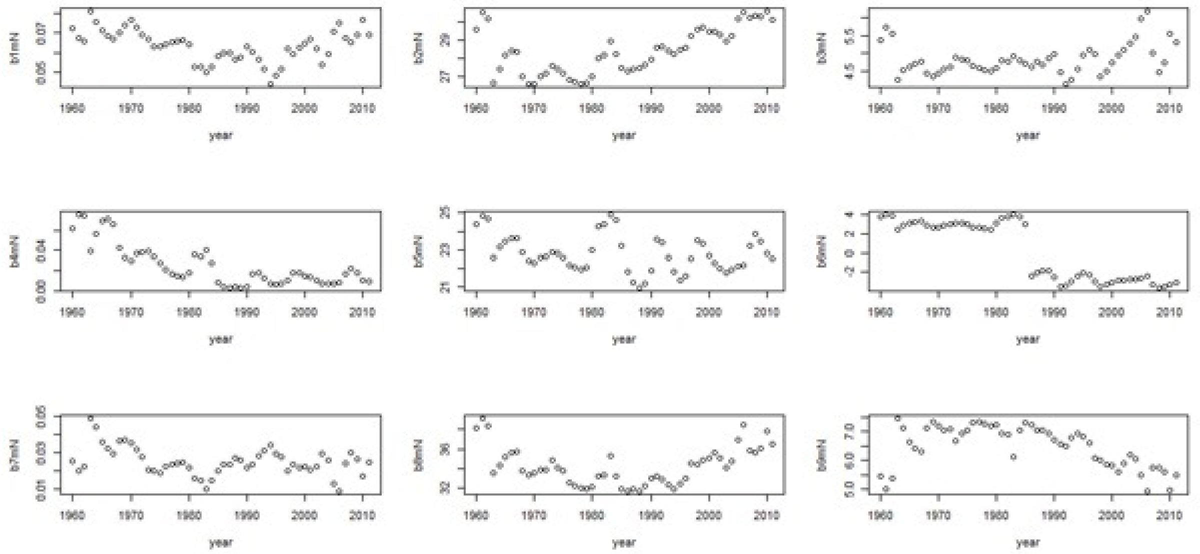

Figure 7.19: Fertility parameter estimate distribution for male population from model stage one.

...

7.2 Appendix B

Given that data was not avaliable for new borns for male and female populations, different approximations where made with the aim of obtaining data that is as close as possible to observed data.

The total age-specific fertility rate in year t is computed as $\varphi_{a,t} = \frac{B_{a,t}}{N^f_{a,t}}$, where $B_{a,t}$ is the total number of births in year t to females of age a and $N^f_{a,t}$ is the female population size of age a at January 1, year t. the total age-specific fertility rate for each year is computed directly from Eurostat. The "female" age-specific fertility rate in year t is naturally defined as $\varphi^{\mathrm{f}}_{\mathrm{a,t}} = \frac{B^f_{a,t}}{N^f_{a,t}}$, where $B^f_{a,t}$is the number of female births in year t to females of age a. Likewise, the "male" age-specific fertility rate in year t is defined as $\varphi^{\mathrm{m}}_{\mathrm{a,t}} = \frac{B^m_{a,t}}{N^m_{a,t}}$, where $B^m_{a,t}$ is the number of male births in year t to females of age a. We have the equivalent conservation relations $\varphi_{a,t} = \varphi^{\mathrm{f}}_{\mathrm{a,t}} + \varphi^{\mathrm{m}}_{\mathrm{a,t}}$ and $B_{a,t} = B^f_{a,t} + B^m_{a,t}$. For recent years, we can directly compute these partial age-specific fertility rates since Eurostat provides births by sex.

For previous years, for which $B^f_{a,t}$. and $B^m_{a,t}$. are not provided by Eurostat, we made different assumptions:

Method 1:

We assume a constant rate of 0.49 for females and 0.51 for males for each year independent of mothers age.

Method 2:

Assuming that mortality and migration can be neglected in the first age class, and assuming that the proportions of female and male newborns is an approximation independent of the age of the mother, we may write : $\varphi^{\mathrm{f}}_{\mathrm{a,t}} \approx \frac{\rho^f_t}{\rho^f_t + \rho^m_t} \varphi_{a,t}$ and $\varphi^{\mathrm{m}}_{\mathrm{a,t}} \approx \frac{\rho^m_t}{\rho^f_t + \rho^m_t} \varphi_{a,t}$, where $\rho^f_t = \frac{N^f_{0,t+1}}{B_t}$ and $\rho^m_t = \frac{N^m_{0,t+1}}{B_t}$, given that $B_t = \sum_{a=0\ldots\max} B_{a,t}$ is the total number of births in year t and that $N^f_{a,t}$ and $N^m_{a,t}$ are the female and male populations size of age a at January 1, year t.

Method 3:

Here we use the same approach as in method 2 but accounting for mortality while ignoring migration. Thus we the new born for males and females increasesd by a fraction as follows: $\frac{B^f_{a,t}}{S^f_{0,t}}$ where, $S^f_{a0,t}$ is the survival rate for female at time t for children age zero.

$\frac{B^m_{a,t}}{S^m_{0,t}}$ where, $S^m_{0,t}$ is the survival rate for male at time t for children age zero.

Other improvements could be considered such as : $B^s_{a,t} = \frac{B^s_t}{B^f_t + B^m_t} B_{a,t} = \frac{N^s_{0,t+1} + D^s_{0,t} - M^s_{0,t}}{N_{0,t+1} + D_{0,t} - M_{0,t}} B_{a,t}$

where: $s = f, m$ $N_{a,t} \equiv N^f_{a,t} + N^m_{a,t}$ is Population size of age a at January 1, year t,

$N_{0,t}$ is Population size of age 0 at January 1, year t,

$D_{a,t} \equiv D^f_{a,t} + D^m_{a,t}$ is Deaths in year t of people of age a, $D_{0,t}$ is Deaths in year t of people of age 0, $M_{a,t} \equiv M^f_{a,t} + M^m_{a,t}$ is Net migration in year t of people of age a, $M_{0,t}$ is Net migration in year t of people of age 0.

This last approch was not implemented, but could be a better approximation compared to the other methods.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Demographics for Epidemiology Forecasting

Richting: **Master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2014**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Atanga, Marcelius

Datum: **10/09/2014**