MULTILEVEL LOGISTIC MODELS ON
CHILDLESSNESS AND FERTILITY BY EDUCATIONAL
FIELDS AMONG SWEDISH WOMEN BORN IN 1970-74

Submitted by
Sofie Wiklund Froby

A thesis submitted to the Department of Statistics in
partial fulfillment of the requirements for a two-year Master
degree in Statistics in the Faculty of Social Sciences

Supervisor
Katrin Kraus

Spring, 2017
This paper further develops earlier findings of variation between the Swedish educational fields regarding childlessness and number of births. It can be argued that the processes of being childless, having one and having two children are substantially different. The purpose of this paper is to investigate the three distinct processes and allowing for educational field specific random effects. An exploratory two-level logistic regression approach is utilized to investigate variations between educational fields.

When applying the two-level logistic regressions one of the main findings were that the effect of being married varied across educational fields regarding childlessness. This was partly explained by the educational field specific covariate. In the process of having one child there were no educational field differences. The effect of mean age at first birth on having two children was different across the fields. This is also partly explained by the educational field specific covariate, the expected annual average income. The results of this thesis support the notion that the processes of being childless, having one and having more than one child are different and that the processes differ between educational fields.
## Contents

1 Introduction 1

2 Background 2
   2.1 The Swedish educational system 2
   2.2 Fertility and education 4
   2.3 Research questions 6

3 Data 6
   3.1 Sample 6
   3.2 Variables 7
   3.3 Childlessness and the mean number of births by educational field 9

4 Methods 10
   4.1 Two-level linear model 11
   4.2 Two-level logistic model 13
   4.3 Estimation 16
   4.4 Covariance specification 18
   4.5 Sample size recommendations 18
   4.6 Centering of variables 18
   4.7 Analysis strategy 19

5 Results 20
   5.1 A two-level logistic model of childlessness 20
   5.2 A two-level logistic model of having one child vs. more than one 23
   5.3 A two-level logistic model of having two children vs. more than two 26

6 Conclusions 29

7 Discussion 31
1 Introduction

In the early 1990s Sweden faced a financial crisis which made the unemployment rate rise from 3 to 11 % (SCB, 2011a). Consequently, it was more difficult to get a job which resulted in that people started to educate themselves further at the University level. It was especially women born in the early 1970s who sought further education (SCB, 2015b). The impact of higher education level of women in relation to childlessness has been widely discussed (Bloom and Pebley, 1982). Women with a high education level tend to have a greater risk of being childless (Lappegård, 2000). Hoem, Neyer and Andersson (2006a,b) discuss the influence of the education field on the ultimate fertility. The field of education seems to influence the number of births a woman chooses to have. Particularly, the educational fields that are dominated by women seems to increase the absolute fertility rate (Hoem et al., 2006b).

In this paper, the relation of childlessness and number of births, on the one hand, and its’ variation across educational fields, on the other, is examined to validate and elaborate the results of Hoem et al. (2006a,b). It is reasonable to view childlessness, having one child and having at least two children as different processes. Infertility is likely one important factor for being childless, however it is irrelevant for the process of having one or more children. It is reasonable that a high age at first birth is an important factor for having only one child. Meanwhile the process of having two children might depend on other factors such as economic status. Hence, it can be argued that there are several distinct data generating processes for the number of children (Santos Silva and Covas, 2000). Consequently, three models will be fitted to describe the different processes. One analysis will model the probability of being childless, the second analysis will model the probability of having one child versus having more than one child and the last analysis will model the probability of having two children versus having more than two children. The outcomes in all three cases are binary, the models are fitted with a logistic link function. A natural consideration for this type of data is a two-part, or hurdle, model. In a hurdle model one part is a logistic model where childlessness is explained and the other part is a Poisson model that models the count data of one or more children. However, it has been argued that the Poisson model often suffers from under-dispersion when it comes to model the number of births (Wang and Famoye, 1997).

The aim of this paper is to see if there is any variation between women in different educational fields regarding being childless, having one child and having two children with an exploratory approach. More specifically, the goal is to describe the main characteristics of the
data. To account for educational field differences in the models a two-level analysis will be used. The first level will include individual specific covariates and the second level will include educational field specific covariates. Allowing for educational field-specific random intercepts and slopes the two-level approach can give a nuanced picture and deeper understanding of how the educational fields affects childlessness and the number of children. In addition, the two-level approach implies a hierarchal data structure where educational fields specific covariates as well as individual specific covariates can be used to explain the potential educational field differences and thus give a detailed picture of the processes.

The structure of the paper will be as following: Chapter 2 will give some background on the Swedish educational system together with the research questions. Chapter 3 will introduce the data and the variables that are included in this paper. In Chapter 4 the multilevel models will be explained and the choice of estimator discussed. In Chapter 5 the results of the analysis will be presented and in Chapter 6 the conclusions of the result is presented. Last, in Chapter 7 a discussion of the results is given.

2 Background

This chapter aspires to give a deeper understanding of the Swedish educational system. As mentioned in the previous chapter the aim of this paper is to investigate the variation between the educational fields regarding childlessness, having one and having two children. Hence, it is of great importance to explain the structure of the educational fields for understanding the analysis of this paper and its’ results. Some of the main factors that may cause the differences between the educational fields regarding childlessness and childbirths will be introduced and discussed. Last in this section the research questions will be presented.

2.1 The Swedish educational system

The Swedish educational system is divided into five educational levels: Primary School, High School, Lower-tertiary education, Upper-tertiary education and Postgraduate studies. The Primary School is a nine-year mandatory education for all children between 7-16 years of age. High school is often a three-year program directed to pupils at the ages 16-19. The Lower-tertiary education level is studies that last for 2-3 years after the High School level. The Higher-tertiary education level is studies that last at least 3 years or more after the High School level.
Postgraduate studies leads to a research degree (Hoem et al., 2006a).

At the High School level students can choose one of nine fields: Agriculture and Veterinary care, Education and Teaching, Engineering and Manufacturing, General programmes, Health and Welfare, Humanities and Arts, Natural Science and Mathematics, Services, and Social Science. The tertiary level is also classified into those fields. The only exception is General programmes, which only exists at the High school level (SCB, 2000). Each field is divided into subject areas, which are further divided into specifications according to the following description:

Main fields → Subject within fields → Specifications within fields

While the main fields consist of nine groups, the subject fields consist of 116 subgroups and the specification fields consist of 330 subgroups (SCB, 2000). To easier grasp the structure of the educational fields Figure 1 shows an example.

![Figure 1: The Health and Welfare field with a few subjects and specifications for an overview](image)

Note that the field of Health and Welfare contains several more subjects and specifications; the figure just presents some examples. In most of the specific fields the educational level
is inbounded but that is not always the case. For example, in the field specification Basic economics there can be several education levels: High School, Lower-tertiary, Upper-tertiary or Postgraduate. Meanwhile the field specification of internal medicine can only be obtained at the University level. The specifications within fields are of most interest in this paper regarding the variability between childlessness and fertility of the 330 educational fields.

2.2 Fertility and education

The association between childlessness and tertiary educational level has been widely discussed and investigated. A common finding is that tertiary educational level leads to higher childlessness (Bloom and Pebley, 1982). However, Hoem et al. (2006a,b) found that the Swedish women with the highest level of education (postgraduate level) do not have substantially higher levels of childlessness compared to lower educational levels in the same educational field. The authors also found that the mean number of births of highly educated women, even though they have a higher mean age at first birth, does not differ from women with lower educational levels in the same educational field. They suggest that those differences can to some degree be explained by the gender structure of the education field, the labour market connected to the educational field and social norms (Hoem et al., 2006a).

In recent decades differences in the gender structure through the educational levels in Sweden have been equalized (Hoem et al., 2006a). However, the Swedish educational system is still gender-segregated in terms of educational fields. Gender gaps between different educational fields have been very persistent over the decades. Even though several efforts have been made by the Swedish government to prevent gender-segregation between educational fields, women and men still choose differently regarding educational fields. The most female-dominated educational fields are: Education and Teaching, Health and Welfare, Humanities and Arts, and Services fields. Meanwhile men often choose: Engineering, Industry, Trade and Natural science fields (Melkas and Anker, 1998, p. 8-27). Female-dominated educational fields tend to have a more supportive environment for childbearing and encouragement of childbearing through education (Hoem et al., 2006a). Hence, childlessness would be expected to be lower within female-dominated educational fields and higher within male-dominated.

One other important factor that influences fertility choices and childlessness is the job prospects for the educational fields. Fields that lead to jobs in the public sector tend to have more security regarding childbearing and parental leave than the private sector (Hoem et al.,
Jobs in the public sector such as Nursing, Teaching and Welfare are female-dominated and the employers are more compatible of combine work and motherhood than general jobs in the private sector. Women in educational fields that lead to jobs in the public sector tend to have lower childlessness and higher mean age at first births than other educational fields (Hoem et al., 2006a). Also, educational fields that do not have an obvious occupation tend to have higher childlessness and lower mean number of births. For example, women that have degrees in Humanities and Arts or in General programmes tend to take longer to find a job after finishing their education. They also tend to have lower income and higher risk of unemployment than other educational fields. Because of the poor job prospects, motherhood may be postponed and childlessness may increase. Educational fields that lead to jobs where skill depreciation might be a problem if taking a long job break, also tend to have a higher childlessness (Hoem et al., 2006a). Since their skill in the occupation may be outdated after a long break from their jobs it can be a reason for higher childlessness or fewer births.

Social norms and social background influence the choice of educational field and fertility. Women that choose educational fields where social relationships and close contact with people are an important part of their profession, such as Teaching and Nursing, may be more family oriented than others. It is known that there is a high correlation between parents educational level and the education attainment of their children (Hahs-Vaughn, 2004). Individuals whose parents have lower income or lower educational level tend to in greater extent not attend tertiary education compared to those with highly educated parents. Individuals with low educated parents tend to be less prepared academically which leads to unfinished educations and drop outs. Hence, this might lead to a lower income which might affect the number of children a women chooses to have (Hahs-Vaughn, 2004).

Given the discussion above it is clear that a lot about childlessness and fertility still needs further attention. Given the known differences between the educational fields it is interesting to find what additional differences across the educational fields that could be found in the relation of childlessness and fertility regarding individual specific covariates.
2.3 **Research questions**

The aim of the study is to answer the questions:

- Is there any variation in mean between women in different educational fields regarding being childless? Does the effect of any individual specific covariate on being childless vary across educational fields?
- Is there any variation in mean between women in different educational fields regarding having one child? Does the effect of any individual specific covariate on having one child vary across educational fields?
- Is there any variation in mean between women in different educational fields regarding having two children? Does the effect of any individual specific covariate on having two children vary across educational fields?
- Does the expected annual income after graduating an educational field explain some of the variations in the coefficients across educational fields?

3 **Data**

This chapter gives an overview of the data used in this paper. The variables that are included are introduced and discussed. To get a deeper understanding of the data, two figures are included to show the differences between the educational fields regarding childlessness and the mean number of births.

3.1 **Sample**

The data for this study is from the administrative agency Statistics Sweden. The variables included in this study were collected from the registers: Registret över totalbefolkningen; Total Population Register (SCB, 2011b), Flergenerationsregistret; Multi-Generation Register (SCB, 2013) and STATIV; The register of data for integration studies (SCB, 2015a). The population study in this report is Swedish women born in 1970-74. Women who are born in Sweden and lives in Sweden the year they turn 40 is included in the paper. Implementing these restrictions results in a study population of 247 252 women. The study is limited to women that have
a known educational field which decrease the study population to 243 512 women. In addition, women that have registered parents with known educational levels were included, which reduces the study population to 214 992 women.

Due to technical limitations of the computers at the administrative agency Statistics Sweden the full sample could not be analysed. The analysis was based on a random sample of 125 000 women. Educational specification fields with more than 50 women in each specification field are included in the random sample and for the study. This results in a decrease from 330 to 241 educational specification fields. This is done to avoid estimation bias which is discussed in the chapter of methods. In Table 1, the sample is summarized by the total number of births a woman had at age 40.

Table 1: The number of births

<table>
<thead>
<tr>
<th>Births</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequencies</td>
<td>17006</td>
<td>18596</td>
<td>61393</td>
<td>22446</td>
<td>4250</td>
<td>1309</td>
</tr>
<tr>
<td>Percent</td>
<td>13.6</td>
<td>14.8</td>
<td>49.1</td>
<td>17.9</td>
<td>3.4</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2 Variables

The dependent variable children is the total number of births a woman had until the age of 40. Twins, triplets and so forth is treated as one birth. The explanatory variables in this study are centred to zero. The importance of centering variables is discussed in the method chapter. A complete description and motivation of the explanatory variables in this paper is as follows:

A dummy variable indicating if a woman ever has been married up to the age of 40 is included in the analysis. It takes on the value 1 if ever married and it takes on the value 0 if the woman never has been married. The variable married is included since it could be expected to vary between educational fields. As mentioned in the previous chapter, women that strive for jobs with social contact tend to have more children and might have a higher probability of getting married than others.

A continuous variable called municipality measuring the percentage of the total number of women in the ages 21-30 in relation to men at ages 23-32, who lived in a specific municipality when women in the study were 27 years old, was included. This variable is centred by the grand mean and is included in the analysis to see if the gender distribution of the municipality
where the women lived at the mean age of first birth, explains some of the childlessness.

Women’s age at first birth, called $MAgeMother$, will be included as an explanatory variable in the analysis. This variable is centred by the grand mean.

A dummy variable, called $FatherEdu$, which indicates if the father of the respondent has a low or high educational level is included in the analysis. This variable takes on the value 1 if the father is highly educated. High education level is defined as studies at the University level and low educational level is defined as studies at High School or less. The variable was collected the year the respondent turned 27. A similar variable was included for the mothers education, called $MotherEdu$.

The expected mean annual income for women in each educational field, called $MIncField$, is included measuring the mean annual labour income in hundreds in SEK of each educational field in the Swedish educational system SUN2000 (SCB, 2000). This variable is centred by the grand mean.

The municipalities in Sweden are divided into nine groups according to the number of residents and commuting patterns. Previous studies have shown that women who live in large cities tend to have fewer children than women who live in smaller towns (Alm-Stenflo, 2002). To control for this nine dummy variables was included indicating where the women lived at age 40. The classification of the variables is specified by the Swedish Association of Local Authorities and Regions. The nine dummies are: If the woman lived in a Large city ($LargeCity$), which is a municipality with at least 200 000 residents. If the woman lived in a commuting municipality near a large city ($ComMunLargeC$). That is; municipalities with at least 40 % of its residents which commute to work in a large city. Medium-sized towns ($MediumTowns$) which is a municipality with at least 50 000 residents. Commuting municipality near medium-sized towns ($ComMunMediumT$) indicates if the woman lived in a municipality with at least 40 % of the residents which commute to work in a medium-sized town. Commuting municipality with a low commuting rate near medium-sized towns ($LowComMunMediumT$) indicates if a woman lived in a municipality with less than 40 % of the residents that commute to a medium-sized town. Small towns ($SmallCity$) is municipalities with at least 15 000 residents. Commuting municipalities near small towns ($ComMunSmallC$) is included which indicates if a women lived in a municipality with at least 30 % of its residents which commute to work in a small town. Rural municipality ($RuralComMun$) indicates if a women lived in a municipality with less than 15 000 residents and less than 30 % of its residents which commute to work. The
last dummy variable is a rural municipality with a visitor industry \((RuralComMunT)\) which indicates if a woman lived in a municipality with a visitor industry.

### 3.3 Childlessness and the mean number of births by educational field

To further describe the dataset, Figure 2 displays the percentage of childlessness among Swedish women born in 1970-74 for the main nine educational fields in Sweden. The lowest percentage of childlessness can be seen in the Education and Teaching, and in the Health and Welfare educational fields. The highest percentage of childlessness can be seen in the Humanities and Art, General programmes and in the Agriculture and Veterinary educational fields.

![Percent childless by education fields](image)

**Figure 2: Childlessness by educational fields among Swedish women born in 1970-74**

In Figure 3, the mean number of births is displayed for the main nine educational fields. The childlessness in Figure 2 are somewhat reflected in the mean number of births. The Humanities and Art educational field has the highest childlessness and has the lowest mean number of births in Figure 3. The Education and Teaching, and the Health and Welfare educational fields have the highest mean number of births.
4 Methods

As mentioned in the previous chapters, the aim of this paper is to further examine the variation between the educational fields regarding childlessness, having one child and having two children. The number of educational fields is quite large (241), which would make a regular multiple regression model hard to grasp. That is, it would not be reasonable to include and interpret 241 dummy variables for each educational field. The hierarchical modelling approach solves that problem by allowing for random effects in the model. That is, the approach allows the intercept and the slope coefficients in the model to vary between the educational fields. The data for this paper includes a variable which is measured for each educational field and not for each individual. It exist variables on two different levels; the individual-level which is the lowest-level and the educational fields-level which is the second/highest-level in the model. This chapter aspires to explain the multilevel modeling approach and discuss the consequences of ignoring a hierarchical data structure in the analysis.

Observations that are nested within some groups, for example students within schools or individuals within geographical areas, tend to have a high correlation within each group. If this intraclass correlation (ICC) is ignored and a regular multiple regression or analysis of variance is applied, it can cause several problems. First of all, the additional information that can be gained from the decomposition from the variation of the within and between variance, will be
lost. Secondly, and perhaps more importantly, the estimates may be more or less uninformative. That is, the fixed effects model assume that all clusters have the same effects, if that is not true than the estimated effect is an average that does not really describe anyone in the sample well. This ultimately may lead to faulty interpretations (Hox, 2002, p. 2-4).

4.1 Two-level linear model

When using two-level models it is assumed that the data of interest has a hierarchical structure and that the dependent variable is measured on the lowest-level. The analysis of this paper is restricted to two-level models; hence higher level models will not be discussed.

For an easier understanding of a two-level model, a classic example of students nested within schools (Hox, 2002, p. 11-35) will be used throughout the explanation of the model. In Equation 1 the dependent variable $Y_{ij}$ can denote a test score of an national math test. The $i = 1, \ldots, n$ stands for the students that have taken the test and $j = 1, \ldots, m$ denotes the school which the student attends.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}.$$  (1)

Let the explanatory variable $X$ be equal to zero if the student is a girl and equal to one if it is a boy. The slope coefficient $\beta_{1j}$ is allowed to vary between all the $m$ schools. Hence, each school has a different slope coefficient. The intercept $\beta_{0j}$ in Equation 1 is also allowed to vary across all the $m$ schools. This means that each school is allowed to have a school specific mean score of girls $\beta_{0j}$ as well as a school specific mean difference between the girls and boys $\beta_{1j}$. These varying effects are called random effects and are referred to as random coefficients. The $\epsilon_{ij}$ is the residual error and is assumed to have mean zero and some positive finite variance (Hox, 2002, p. 11-21). The intercept of Equation 1 $\beta_{0j}$ is regressed on a school specific covariate $Z$, where the intercept and slope of this regression are the same for all schools.

$$\beta_{0j} = \alpha_0 + \alpha_1 Z_j + e_{0j}$$  (2)

Implying that the difference in the intercept of the schools in Equation 1 is at least partly explained by the variation across schools on the covariate $Z$. In this example it could be the number of hours scheduled for math lessons in each school. The residual $e_{0j}$ is the usual regression residual and is assumed to be independent and identically distributed across the
schools. \( \epsilon_{0j} \) is commonly assumed to be normally distributed and independent of the lowest level residual \( \epsilon_{ij} \).

The slope coefficient of Equation 1 is given by: the intercept \( \alpha_{10} \) and the slope coefficient \( \alpha_{11} \) which is the same through all cluster categories schools.

\[
\beta_{1j} = \alpha_{10} + \alpha_{11}Z_j + \epsilon_{1j} \tag{3}
\]

Note that the variable \( Z_j \) can be understood as a moderator variable for the relationship between the variable \( X \) and the dependent variable \( Y \). Hence, the number of hours scheduled for math lessons in a specific school may explain some of the variation between the schools regarding gender differences in math. \( Z \) is used to model the differences between clusters in relation of \( X \) and \( Y \). The residual \( \epsilon_{1j} \) is the usual regression residual and is assumed to be independent and identically distributed across the schools. Equations 2, 3 and 4 can be summarized as a hierarchical model:

\[
Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij}
\]

\[
\beta_{0j} = \alpha_{00} + \alpha_{01}Z_j + \epsilon_{0j} \tag{4}
\]

\[
\beta_{1j} = \alpha_{10} + \alpha_{11}Z_j + \epsilon_{1j}
\]

Sometimes in the literature the reduced form of the interaction model is displayed as:

\[
Y_{ij} = \alpha_{00} + \alpha_{01}Z_j + \epsilon_{0j} + \alpha_{10}X_{ij} + \alpha_{11}Z_jX_{ij} + \epsilon_{1j}X_{ij} + \epsilon_{ij} \tag{5}
\]

The level of variation between the clusters is of great interest in a two-level model. By computing the intercept-only model an intraclass correlation can be calculated. The intercept-only model, displayed in Equation 6, is a two-level model without any explanatory variables. Hence, the variance is decomposed into level-1 and level-2 variation, or within and between school variations, to follow the school analogy.

\[
Y_{ij} = \beta_{0j} + \epsilon_{ij}
\]

\[
\beta_{0j} = \alpha_{00} + \epsilon_{0j} \tag{6}
\]

The lowest-level variance of \( \epsilon_{ij} \) is denoted as \( \sigma^2_\epsilon \) and the second-level variance of \( \epsilon_{0j} \) is denoted as \( \sigma^2_{\epsilon_0} \). The intraclass correlation is then calculated by the following equation:
\[ \rho = \frac{\sigma^2_{e0}}{\sigma^2 + \sigma^2_{\epsilon}} \]  

(7)

where \( \rho \) is the proportion of variance explained by the cluster categories. If there is none or very little variation between the clusters it implies that there is no clustering/hierarchy of the data. In applied studies without repeated measurements, the intraclass correlation proportion often lays around 0.1 (Hox, 2002, p. 11-21).

### 4.2 Two-level logistic model

The choice of analysing the data in three separate models led to dichotomous outcomes. This means that the linear model in the previous section does not fit the data and therefore the logistic link is used. The assumptions of normally distributed variables and homoscedastic error terms for a basic two-level model are violated for dichotomous data. Instead a two-level logistic regression will be applied which mainly consists of three parts: a link function, a random part and a systematic component (Wang et al., 2011, p. 113-118). The random part of the model defines the distribution of the error term. In a linear regression modelling one often makes the assumption that the data is normally distributed but in the logistic model it can be specified to be any member of the exponential family. In the present study the data comes from a binomial distribution. The systematic component is the linear predictor \( \eta \) of a multiple regression:

\[ \eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \]  

(8)

The link function can be denoted as \( \eta = f(\mu) \) where it links the expected values of the dependent variable to the predicted values \( \eta \). When the outcome variable is binary the link function is commonly specified with the logit function \( \eta = logit(\mu) \). The logistic regression can be written as:

\[ y = logit(\beta_0 + \beta_1 X_1 + \beta_2 X_2) \]  

(9)

Note that the Equation 9 does not have an error term. This is because the variance of the binomial distribution is a function of the mean (Hox, 2002, p. 103-122).

A two-level logistic model can be written as in Equation 10, where \( i \) stands for the lowest-level i.e. the individual level and \( j \) stands for the second-level i.e. the clustering level. The dependent variable \( y \) can be only take on a binary outcome \( \pi \) which has a binomial error
distribution with the expected value $\mu$. The expected value of $\mu$ is then predicted by the logistic regression. Note that there is no explanatory variable included in the model.

$$y_{ij} = \pi_{ij} \quad \text{where} \quad \pi \sim Binomial(\mu)$$

$$\pi_{ij} = \logit(\eta_{ij})$$

$$\eta_{ij} = \beta_{0j}$$

$$\beta_{0j} = \gamma_{00} + e_{0j}$$

(10)

The intercept-only model, or empty model, is of interest since it gives the information needed to calculate the ICC. It is therefore common practise to first estimate the empty model and examine the ICC before moving on to more intricate models. The ICC for a two-level logistic model cannot be calculated as for a linear two-level model as mentioned before. The error variance for the individual level in the model is part of the specification of the error distribution. In the case of modelling dichotomous data a scale factor for the variance is commonly set to one. This implies the assumption that the observed error in the model follows the binomial distribution exactly.

$$Var(\pi_{ij}) = \sigma^2 \pi_{ij}(1 - \pi_{ij})$$

(11)

Equation 11 shows how the lowest level variance is modelled and where the scale factor is $\sigma^2$ which commonly is set to one. If the data does not follow the binomial distribution exactly there can be a problem with under- or over dispersion (Hox, 2002, p. 103-122). The intraclass correlation can be calculated with the lowest level variance as $\pi^2 / 3 = 3.29$ according to the findings in Evans et al. (1993). The intraclass correlation can then be calculated as in Equation 12, note that $\sigma^2_{e0}$ is the second-level variance of the intercept-only model.

$$\rho = \frac{\sigma^2_{e0}}{(\sigma^2_{e0} + 3.29)}$$

(12)

If the ICC is reasonably high the model can be extended to explain the variation between the clusters on the higher level. Equation 13 shows a two-level model with explanatory variables included.
\[ y_{ij} = \pi_{ij} \quad \text{where} \quad \pi \sim \text{Binomial}(\mu) \]
\[ \pi_{ij} = \text{logit}(\eta_{ij}) \]
\[ \eta_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \beta_{2j}W_{ij} \] (13)

Equation 13 describes a dichotomous dependent variable which has a binomial distribution with mean=\( \mu \). The logit link function describes that the \( \mu \) of the distribution is modelled by a logistic regression. Let \( X \) and \( W \) be level-1 covariates, where \( i = 1, ..., n \) stands for the individual level/the lowest level and the \( j = 1, ..., m \) is the cluster level/the second level. The \( \beta \):s can be random or fixed in the model. For example the \( \beta \):s can be defined as the following:

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}Z_{0j} + e_{0j} \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + e_{1j} \]
\[ \beta_{2j} = \gamma_{20} \] (14)

The intercept \( \beta_{0j} \) is considered as a random intercept which means that the intercept varies between the clustering categories. The intercept \( \gamma_{00} \) and the slope coefficient \( \gamma_{01} \) is the same across all the clustering categories. The cluster specific covariate \( Z_{0j} \) is the predictor variable on the second-level which varies across all the clustering categories. The random slope coefficient \( \beta_{1j} \) is specified to vary across the clustering categories. The intercept \( \gamma_{10} \) and the slope coefficient \( \gamma_{11} \) is the same across all the clustering categories. The cluster specific covariate \( Z_{1j} \) varies across all the clustering categories. The residual \( e_{1j} \) also varies between the clustering categories. The last slope coefficient \( \beta_{2j} \) in the equation is fixed across all the clustering categories and does not have any random term. This means that the level one slope is the usual regression coefficient. If some slopes are random and some are fixed in a model, it is commonly referred to as a mixed or mixed effects-model.

The interpretation of the coefficients of a two-level logistic model is defined by the underlying logit transformation. To interpret the coefficients in terms of the dependent variable in proportions the inverse of the logit transform have to be applied.

\[ f(x) = \frac{e^x}{1 + e^x} \] (15)
By inserting the estimated coefficient in Equation 15 it can be interpreted as the expected rate of the dependent variable (Hox, 2002, p. 103-122). If returning to the math score example, but in this case the dependent variable $Y$ is dichotomous and describes if a student passed or not passed the math test. Remember, $X$ was taking the value 1 if there was a boy and 0 if there was a girl taking the test. Lets say that the estimated slope coefficient is $\beta_{1j} = 0.6$ and the intercept in the model is $\beta_{0j} = -1.96$. To get the estimated pass rate on averaged for a boy, one would first add the intercept and the coefficient together $(-1.96 + 0.6) = -1.36$. Then by transforming -1.36 by the inverse logit transformation the estimated pass rate on averaged for a boy is 20%. Also, let’s say that a school specific variable $Z$ is included on the second level for $\beta_{1j}$. $Z$ specifies the number of scheduled hours for math on a school with a slope coefficient $\alpha_{11} = 0.3$. If the scheduled hours for maths were to increase 1 unit for this school we would expect the pass rate to be $(-1.96 + 0.6 + 0.3) = -1.06$. After the inverse logit transformation the estimated pass rate on averaged for a boy has increased to 25%.

4.3 Estimation

There are several estimation methods for estimating two-level logistic models. For example there are a number of Pseudo-likelihood estimations that could be used: Residual log Pseudo-Likelihood, Restricted Maximum Pseudo-likelihood, Restricted Minimum Pseudo-Likelihood. Pseudo-likelihood estimations are commonly applied when the data arise from multistage sampling. The estimation allows for weights accounting for different sampling probabilities (SAS, 2017b). The estimation used in this paper is a maximum-likelihood based estimation, called the Laplace approximation, this choice is motivated by several reasons. The Laplace approximation has been criticized, mainly for being biased in small sample analysis. However, in studies for large datasets with a large number of clustering categories and large number of observations in each cluster, the Laplace approximation has good properties (Capanu et al., 2013). The approximation works computationally fast which is crucial for this particular analysis. Another advantage of the approximation is that likelihood ratio tests can be computed to compare the model fit (Wang et al., 2011, p. 126-128).

The marginal distribution based on the Laplace approximation can be derived as:

$$p(y) = \int e^{\exp (c_l f(y, \beta, \theta; \gamma))} d\gamma,$$

16
where $\beta$ denotes the vector of fixed-effects parameters, $\gamma$ denotes the vector of random-effects in the model and $\theta$ denotes the vector of covariance parameters, $\phi$ denotes the scale parameter and $c_l$ denotes a constant. Then, if $c_l$ is large the integral of Laplace is:

$$L(\beta, \theta; \hat{\gamma}, y) = \left(\frac{2\pi}{c_l}\right)^{\frac{m_a}{2}} \left| - f''(y, \beta, \theta; \hat{\gamma}) \right|^{-1/2} e^{c_l f(y, \beta, \theta; \hat{\gamma})},$$

(17)

where $n_\gamma$ denotes the elements in $\gamma$. If the first derivative satisfies the condition $\frac{\partial f(y, \beta, \theta; \gamma)}{\partial \gamma} = 0$ then the second derivative is the following:

$$f''(y, \beta, \theta; \hat{\gamma}) = \frac{\partial^2 f(y, \beta, \theta; \gamma)}{\partial \gamma \partial \gamma'} \Big|_{\hat{\gamma}}.$$

(18)

The objective function for the Laplace coefficient estimation in SAS (SAS, 2017a) is

$$-2 \log(L(\beta, \theta; \hat{\gamma}, y))$$

and since the random effects $\hat{\gamma}$ depends on $\hat{\theta}$ and $\hat{\beta}$ the estimation method solves the sub-optimization problem for given values of $\hat{\theta}$ and $\hat{\beta}$, so the solution maximizes $f(y, \beta, \theta; \hat{\gamma})$.

When dealing with clustered data the marginal distribution can be derived as the following:

$$p(y) = \prod_{i=1}^{m} \int \exp\left(n_i f(y, \beta, \theta; \gamma_i)\right) d\gamma_i$$

(19)

where

$$n_i f(y, \beta, \theta; \gamma_i) = \sum_{j=1}^{n_i} \log\left(p(y_{ij} | \gamma_i) + n_i \log p(\gamma_i)\right)$$

(20)

Assuming that the observations within each cluster category $n_i$ is large, then the Laplace estimation of the $i$th individual’s marginal probability density function is:

$$p(y_i | \beta, \theta) = \int \exp\left(n_i f(y, \beta, \theta; \gamma_i)\right) d\gamma_i = \frac{2\pi^{n_i/2}}{\left|-n_i f''(y_i, \beta, \theta; \hat{\gamma})\right|^{-1/2}} \exp\left(n_i f(y, \beta, \theta; \hat{\gamma})\right)$$

(21)

Here is $n_\gamma$ denoted as the dimension of the random effects $\gamma_i$. Then the Laplace approximates to the marginal likelihood as in Equation 22 (SAS, 2017a).

$$\log(L(\beta, \theta; \hat{\gamma}, y)) = \sum_{i=1}^{m}(n_i f(y, \beta, \theta; \hat{\gamma}_i)) + \frac{n_\gamma}{2} \log(2\pi) - \frac{1}{2} \log\left|-n_i f''(\beta, \theta; \hat{\gamma}_i)\right|$$

(22)
A more detailed explanation of the estimation technique can be found at SAS User’s Guide (2017).

4.4 Covariance specification

When estimating a two-level model a covariance specification has to be chosen. Hence, when the estimated covariance matrix is not positive definite during the function evaluation of the estimator, an algorithm has to remake it. The algorithm evaluates and returns the covariance matrix positive definite again.

Since we do not want to make any constrains on the covariance structure, the chosen covariance specification for this study is the Cholesky decomposition. Each covariance and variance of the random effects is estimated from the data. This structure is often used when estimating a small number of random effects within a large sample (Kiernan et al., 2012).

4.5 Sample size recommendations

Two-level models are sensitive for small sample sizes for both the individual level and the group level. Convergence problems may arise when the cluster categories are small or when the number of observations in each category are small. The random and fixed coefficients for a two-level logistic model tend to be overestimated when the cluster categories are small. Two-level logistic models with five observations in each cluster category tend to have severe biases for both fixed and random effects and for both small and large sample sizes of cluster categories. The sample size recommendations for two-level logistic models are at least 50 observations in each cluster category for a sample size of 100 groups or more (Moineddin et al., 2007).

4.6 Centering of variables

The aim of centering variables is to make the interpretation of estimates easier especially when dealing with interactions effects and for convergence problems. If we want to model a multiple regression a linear transformation of the explanatory variables would not change the nature of the estimates of the model. The residual variance would still be the same for the model. If we would do the same linear transformation for a regression model with random effects, that is a multilevel model where the slopes of the estimates can vary between groups, the linear transformation fails (Hox, 2002, p. 54-58). To handle this problem for multilevel modelling
the explanatory variables can be centred by subtracting the grand means. The interpretation of the intercept in a multilevel model is the expected value of the dependent variable when all explanatory variables are zero. That can be a problem if zero does not have meaningful value of the explanatory variables in the model (Enders and Tofighi, 2007). For example if a variable measures the age of women when they had their first child. The intercept would correspond to a woman having her first child at zero, which of course makes no sense. This makes the interpretation meaningless, and the parameter is in a way wasted. If the grand mean instead is subtracted from the variable it would have a meaningful zero. The interpretation of the intercept would then instead be the mean age of a woman when she had her first child.

It can also be useful to center and even standardize the explanatory variables to avoid numerical convergence problems when estimating the models. For example, if we are dealing with dummies in the model that only has the observed values 0 and 1, and we are also dealing with a variable that explains the income expressed in hundreds, it can cause convergence problems (Hox, 2002, p. 54-58).

4.7 Analysis strategy

When approaching a hierarchal data set from an exploratory point of view, it is essential to have a strategy to avoid arbitrary choices. A natural starting point is to analyse the intercept-only model. As mentioned earlier in this chapter the intercept-only model is the model without any explanatory variables included. The intraclass correlation declares the variance explained by the clustering categories. If the ICC is too small, two-level analysis is not appropriate. If there is substantial between-level variation, indicated by a reasonably high ICC, the next step is to model that variation by including between-level covariates. The significance of the variables can be investigated together with tests of model fit. If the estimation method is for example the Laplace approximation, a log likelihood ratio test can be computed between the intercept-only model and the model for the lowest-level variables. If the later one has the best model fit, the third step of the analysis can be conducted.

In the third step the second level explanatory variables are included in the model. This model tells us if the second level covariates explain any of the variations between the cluster categories. The model fit can again be tested with a log likelihood ratio test of the previous model versus this model.

The fourth step is to investigate if any of the explanatory variables on level-1 has a random
effect. In other words, if any of the explanatory variables on the first level have random slopes which would imply cluster variation in the slope coefficients. A straightforward approach is to include the random effects variable-by-variable. The explanatory variables that had a significant random effect will be included in the model and a test of model fit can be conducted again.

The fifth step is to add the highest-level variables to model the lowest-level explanatory variables that had a significant random effect. All together this gives a model where all variables on the first level, including the intercept is investigated for cluster variation. In the cases of significant variation, the between-level variables are used to try to explain that variation. Again, model fit can be tested by a log likelihood ratio test (Hox, 2002, p. 49-53).

5 Results

This chapter presents the results of the analysis. The estimation was done in SAS 9.4 with the PROC GLIMMIX procedure. The analysis strategy, explained in the previous chapter, was followed when conducting the analysis. The process of being childless will be presented and interpreted first. Then, the process of having one child in relation to having more than one child will presented. Last, the process of having two children in relation of having more than two children will be presented.

5.1 A two-level logistic model of childlessness

The first process modelled is childlessness. A two-level logistic regression model with educational fields as clustering variable was fitted. Note that, as explained in the previous chapter, only educational fields with more than 50 women in the population were included in the analysis which generated a total of 241 educational fields. The number of childless women in the study were 17 006 and women that had one or more children were 107 994.

In Table 2 the estimates for the two-level logistic model for childlessness are presented. The first column in Table 2 displays the intercept-only model, the second column displays the model with only first level covariates and the third column displays the final model with one slope coefficient as a random effect. The intraclass correlation for the intercept-only model is presented in Equation 23. The intraclass correlation 0.11 can be interpreted as the proportion of the total variance accounted for by the between-level variation among the educational fields.
\[ ICC = \frac{0.4177}{0.4177 + 3.29} = 0.11 \] (23)

The random intercept in the intercept-only model in Table 2 can be interpreted, after an inverse logit transformation, as the estimated average childlessness rate over all educational fields. The inverse of the logit transformation is \( \frac{e^{-1.7659}}{1 + e^{-1.7659}} = 0.146 \), 14.6\% is the estimated average childlessness rate over all the educational fields.

The Level-1 model in Table 2 shows the estimates of variables on the individual level. The error variance of the model is smaller than that of the intercept-only model. The model fit indicated by the -2 log likelihood is also smaller for this model. To evaluate the model of fit of the intercept-only model and the level-1 model a log likelihood ratio test is performed, where the test statistic is: \( \chi^2 = -2LL_{Model1} - (-2LL_{Model2}) \) and when implemented: 97510 - 83896 = 13614 where \( df = 12 \). Based on the results of the likelihood ratio test, the level-1 model is preferred.

The Level-2 model in Table 2 includes a random slope coefficient and a level-2 covariate. The error variance for the model is 0.3551 and the model fit for the -2 log likelihood is 79728. Again a likelihood ratio test is performed and this time between the level-1 model and the level-2 model. The test statistic for the likelihood ratio test is \( \chi^2 = 83896 - 79728 = 4168 \) with \( df = 13 \). Based on the results of the likelihood ratio test, the level-2 model is preferred.

The final model is thus the two-level logistic model with the random slope coefficient for the variable which indicates if a woman ever been married. The random intercept in the model has an average estimate over all the educational fields of -0.9348. When interpreting the intercept all the covariates in the model is held at zero. After the inverse logit transformation, the estimated childlessness rate is 28.5 \%; that is for a woman that has: never been married, lived in a large city at age 40, lived in a municipality with an average gender distribution at age 27, parents with low educational level and attended an educational field that has an annual average income over all educational fields after graduation.

If the mean gender ratio in a municipality where a woman lived at age 27 increases by one unit the estimated childlessness rate increases on average to 28.7 \% when all the other covariates in the model are zero. If the father is highly educated the estimated childlessness rate on average increases to 29.3 \% holding every other variable in the model at zero. The
Table 2: Childlessness in relation of having one or more children

<table>
<thead>
<tr>
<th>Variables</th>
<th>Intercept model</th>
<th>Level-1 model</th>
<th>Level-2 model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept -1.7659* (0.029)</td>
<td>-0.9197* (0.033)</td>
<td>-0.9348* (0.034)</td>
</tr>
<tr>
<td></td>
<td>Married -1.9041* (0.029)</td>
<td>-1.9638* (0.030)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Municipality 0.0213* (0.006)</td>
<td>0.0224* (0.006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FatherEdu 0.0491* (0.023)</td>
<td>0.0585* (0.024)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MotherEdu -0.0207 (0.021)</td>
<td>-0.0087 (0.024)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LargeCity -</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MediumTowns 0.5835* (0.026)</td>
<td>0.5804* (0.027)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SmallCity -0.1045* (0.031)</td>
<td>-0.0973* (0.032)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ComMunLargeC -0.2467* (0.038)</td>
<td>-0.2452* (0.038)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ComMunMediumT -0.2649* (0.028)</td>
<td>-0.2526* (0.029)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ComMunSmallC -0.2472* (0.045)</td>
<td>-0.2538* (0.046)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LowComMunMediumT -0.2739* (0.043)</td>
<td>-0.2755* (0.044)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RuralComMun -0.3232* (0.051)</td>
<td>-0.3129* (0.051)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RuralComMunT -0.1653* (0.078)</td>
<td>-0.1721* (0.083)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIncField -0.0243* (0.003)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Error Variance**

<table>
<thead>
<tr>
<th></th>
<th>Intercept 0.4116* (0.025)</th>
<th>0.3563* (0.024)</th>
<th>0.3551* (0.025)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Civil</td>
<td></td>
<td>0.1503* (0.035)</td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit**

|                | -2 Log Likelihood 97510 | 83896 | 79728 |

*Estimates standard errors shown in parentheses. *p<0.05.*
estimate of the mother's educational level was not significant. The variable controlling for where the woman lived at age 40, when holding all the other variables in the model at zero, had the estimated childlessness rate on average when living in a: Medium-sized town 41 %, small town 26 %, commuting municipality near a large city 23.5 %, commuting municipality near medium-sized towns 23.3 %, commuting municipality near small towns 23.4 %, commuting municipality with a low commuting rate near medium-sized towns 22.9 %, rural municipality 22.3 % and rural municipality with a visitor industry 24.8 %.

The covariate that indicates if a woman has ever been married, has a random slope coefficient in this model. This implies that the slope coefficient varies across the educational fields. The intercept in the regression of the random slope coefficient of being married is estimated to -1.9638. After transformation the average childlessness rate for women that has ever been married is 5.2 % when holding every other variable in the model at zero. The level-2 variable, the expected mean annual income after graduating a specific educational field, can be understood as a moderator variable for the relationship between childlessness and if a woman been married. The expected mean annual income explains some of the variation across the educational fields. The relationship between childlessness and a woman that is or been married varies according to the value of the moderator variable. The slope coefficient for the expected mean annual income is -0.0243. The educational field specific variable, the expected mean annual income, varies between the fields. If the level-2 variable increases by one unit, the estimated average childlessness rate drops to 4.1 %, when holding every other variable in the model at zero.

5.2 A two-level logistic model of having one child vs. more than one

The second process of interest is having one child in relation of having more than one child. A two-level logistic regression model with educational fields as clustering variable was fitted. The number of women with one child were 18 596 and women that had more than one child were 89 398, which is a total of 107 994 women.

In Table 3 the estimates for the two-level logistic model of having one child is presented. The ratio of women having one child rather than more than one child is called one-child rate for short. The first column in Table 3 displays the intercept-only model, the second column displays the model with only first level covariates and the third column displays the final model. The intraclass correlation for the intercept-only model is presented in Equation 24. The intraclass
correlation 0.104 can be interpreted as the proportion of the total variance accounted for by the between-level variation among the educational fields.

\[ ICC = \frac{0.3825}{0.3825 + 3.29} = 0.104 \] (24)

The random intercept in the intercept-only model in Table 3 can be interpreted as that the estimated average one-child rate is 17.9% across all the educational fields.

The model fit of the intercept-only model for the -2LL is 98 488, then by including the level-1 variables the model fit of the -2LL decreases to 81 882. A log likelihood ratio test is conducted to see which model that has the best fit of the data. The test statistic for the test is \( \chi^2 = 98488 - 81882 = 16606 \) with \( df = 13 \). Based on the results of the likelihood ratio test, the level-1 model is chosen. To determine the best model fit of the level-1 model and the level-2 model a log likelihood ratio test is conducted. The test statistic for the test is \( \chi^2 = 81882 - 78102 = 3780 \) with \( df = 14 \). Based on the results of the likelihood ratio test, the level-2 model is chosen.

The final model in Table 3 did not have any significant random slope coefficients. The level-2 covariate, the expected mean annual income, is included to explain the variation in the random intercept of the model. The intercept of the level-2 regression of the random intercept is -1.3803, which after transformation has an estimated average one-child rate of 20% when holding all the other variables in the model at zero. That is for a woman that: never been married, had the first child at the mean age of first birth, lived in a large city at age 40, lived in a municipality with an average gender distribution at age 27, has parents with low educational level and attended an educational field that has an annual average income over all educational fields after graduation.

The educational field specific covariate, the expected mean annual income, has a slope coefficient of -0.0295. The estimated average one-child rate decreases to 19.6% when the expected mean annual income increases by one unit and holding every other variable in the model at zero.

If the variable for the gender ratio in a municipality where a woman lived at age 27 increases by one unit the estimated one-child rate on average is 19.6%. If the woman have ever been married the estimated one-child rate on average decreases to 10% holding every other
Table 3: Having one child in relation to having more than one child

<table>
<thead>
<tr>
<th>Variables</th>
<th>Intercept model</th>
<th>Level-1 model</th>
<th>Level-2 model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.5180* (0.023)</td>
<td>-1.3250 (0.053)</td>
<td>-1.3803* (0.033)</td>
</tr>
<tr>
<td>Married</td>
<td>-0.8221* (0.018)</td>
<td>-0.8124* (0.018)</td>
<td></td>
</tr>
<tr>
<td>Municipality</td>
<td>-0.0319* (0.006)</td>
<td>-0.0389* (0.007)</td>
<td></td>
</tr>
<tr>
<td>MAgeMother</td>
<td>0.2037* (0.002)</td>
<td>0.2046* (0.002)</td>
<td></td>
</tr>
<tr>
<td>FatherEdu</td>
<td>-0.1906* (0.024)</td>
<td>-0.1746* (0.025)</td>
<td></td>
</tr>
<tr>
<td>MotherEdu</td>
<td>-0.1537* (0.023)</td>
<td>-0.1472 (0.023)</td>
<td></td>
</tr>
<tr>
<td>LargeCity</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>MediumTowns</td>
<td>0.1274* (0.029)</td>
<td>0.1353* (0.029)</td>
<td></td>
</tr>
<tr>
<td>SmallCity</td>
<td>-0.1238* (0.032)</td>
<td>-0.1248* (0.032)</td>
<td></td>
</tr>
<tr>
<td>ComMunLargeC</td>
<td>-0.1098* (0.036)</td>
<td>-0.1135* (0.037)</td>
<td></td>
</tr>
<tr>
<td>ComMunMediumT</td>
<td>-0.2162* (0.027)</td>
<td>-0.2104* (0.027)</td>
<td></td>
</tr>
<tr>
<td>ComMunSmallT</td>
<td>-0.1888* (0.045)</td>
<td>-0.1949* (0.046)</td>
<td></td>
</tr>
<tr>
<td>LowComMunMediumC</td>
<td>-0.0778* (0.042)</td>
<td>-0.0822 (0.043)</td>
<td></td>
</tr>
<tr>
<td>RuralComMun</td>
<td>-0.0746* (0.048)</td>
<td>-0.0632 (0.049)</td>
<td></td>
</tr>
<tr>
<td>RuralComMunT</td>
<td>-0.1779* (0.079)</td>
<td>-0.1856* (0.081)</td>
<td></td>
</tr>
<tr>
<td><strong>Level-2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIncField</td>
<td>-0.0295* (0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Error Variance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.3825* (0.025)</td>
<td>0.2956* (0.021)</td>
<td>0.2947* (0.023)</td>
</tr>
<tr>
<td><strong>Model Fit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>98488</td>
<td>81882</td>
<td>78102</td>
</tr>
</tbody>
</table>

*Estimate standard errors shown in parentheses. *p<0.05.
variable in the model at zero. If the mean age of first birth of the mother increases by one year the estimated one-child rate on average increases to 23.6 %, when all the variables in the model is equal at zero. If the father has a high education level the estimated one-child rate on average decreases to 17.5 % holding everything else in the model to zero. The estimate of the mothers educational level was not significant. The variable controlling for where the women lived at age 40, when holding all the other variables in the model to zero, has the estimated one-child rate on average when living in: Medium-sized town 22.3 %, small town 18.2 %, commuting municipality near a large city 18.3 %, commuting municipality near medium-sized towns 16.9 %, commuting municipality near small towns 17.2 %, commuting municipality with a low commuting rate near medium-sized towns 18.8 %, rural municipality 19.1 % and rural municipality with a visitor industry 17.3 %.

5.3 A two-level logistic model of having two children vs. more than two

The third process of interest is having two children in relation of having more than two children. A two-level logistic regression model with educational fields as clustering variable was fitted. The number of women with two children were 61 393 and women that had more than two children were 28 086, which is a total of 89 479 women.

In Table 4 the estimates for the two-level logistic model of having two children is presented. The ratio of women having two children rather than more than two children is called two-child rate for short. The first column in Table 4 is the intercept-only model, the second column displays the model with only first level covariates and the third column shows the final model. The intraclass correlation for the intercept-only model is presented in Equation 25. The intraclass correlation 0.08 can be interpret as the proportion of the total variance accounted for by the between-level variation among the educational fields.

$$ICC = \frac{0.2968}{0.2968 + 3.29} = 0.08$$

The random intercept in the intercept-only model in Table 4 can be interpreted as the estimated two-children rate on average is 70 % across all the educational fields.

The model fit of the intercept-only model for the -2LL is 109 454 then by including the level-1 variables the model fit of the -2LL decreases to 98 843. A log likelihood ratio test is
Table 4: Having two children in relation to having more than two children

<table>
<thead>
<tr>
<th>Variables</th>
<th>Intercept model</th>
<th>Level-1 model</th>
<th>Level-2 model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.8510* (0.022)</td>
<td>1.4807 (0.031)</td>
<td>1.4676* (0.032)</td>
</tr>
<tr>
<td>Married</td>
<td>-0.5191* (0.018)</td>
<td>-0.5111* (0.018)</td>
<td></td>
</tr>
<tr>
<td>Municipality</td>
<td>-0.0306* (0.005)</td>
<td>-0.0283* (0.005)</td>
<td></td>
</tr>
<tr>
<td>MAgeMother</td>
<td>0.1757* (0.002)</td>
<td>0.1743* (0.002)</td>
<td></td>
</tr>
<tr>
<td>FatherEdu</td>
<td>-0.2319* (0.021)</td>
<td>-0.2253* (0.021)</td>
<td></td>
</tr>
<tr>
<td>MotherEdu</td>
<td>-0.2198* (0.020)</td>
<td>-0.2173* (0.023)</td>
<td></td>
</tr>
<tr>
<td>LargeCity</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>MediumTowns</td>
<td>0.0580* (0.029)</td>
<td>0.0649* (0.030)</td>
<td></td>
</tr>
<tr>
<td>SmallCity</td>
<td>-0.1328* (0.026)</td>
<td>-0.1347* (0.027)</td>
<td></td>
</tr>
<tr>
<td>ComMunLargeC</td>
<td>-0.0698* (0.030)</td>
<td>-0.0741* (0.031)</td>
<td></td>
</tr>
<tr>
<td>ComMunMediumT</td>
<td>-0.0811* (0.023)</td>
<td>-0.0733* (0.024)</td>
<td></td>
</tr>
<tr>
<td>ComMunSmallC</td>
<td>-0.1119* (0.035)</td>
<td>-0.1026* (0.036)</td>
<td></td>
</tr>
<tr>
<td>LowComMunMediumT</td>
<td>-0.0957* (0.034)</td>
<td>-0.0893* (0.035)</td>
<td></td>
</tr>
<tr>
<td>RuralComMun</td>
<td>-0.2709* (0.040)</td>
<td>-0.2547* (0.040)</td>
<td></td>
</tr>
<tr>
<td>RuralComMunT</td>
<td>-0.2006* (0.066)</td>
<td>-0.1811* (0.068)</td>
<td></td>
</tr>
<tr>
<td><strong>Level-2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIncField</td>
<td>-0.0086* (0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Error Variance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.2968* (0.019)</td>
<td>0.2890* (0.021)</td>
<td>0.2828* (0.020)</td>
</tr>
<tr>
<td>MAgeMother</td>
<td></td>
<td>0.0101* (0.003)</td>
<td></td>
</tr>
<tr>
<td><strong>Model Fit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>109454</td>
<td>98843</td>
<td>95101</td>
</tr>
</tbody>
</table>

*Estimate standard errors shown in parentheses. *p<0.05.
conducted to see which model that has the best fit for the data. The test statistic for the test is $\chi^2 = 109454 - 98943 = 10511$ with $df = 13$. Based on the results of the likelihood ratio test, the level-1 model is chosen. In the level-2 model the variable for the mean age at first birth has a random slope coefficient and the level-2 variable, the expected annual income after graduating in an educational field, is included as a predictor on the second level. To determine the best model fit of the level-1 model and the level-2 model a log likelihood ratio test is conducted. The test statistic for the test is $\chi^2 = 98843 - 95101 = 3742$ with $df = 14$. Based on the results of the likelihood ratio test, the level-2 model is preferred.

The intercept can be interpreted as the estimated two-child rate on average is 81%, when holding all variables in the model at zero. That is for a woman that: never been married, had her first child at the mean age of first birth, lived in a large city at age 40, lived in a municipality with an average gender ratio at age 27, has parents with low educational level and attended an educational field that has an annual average income over all educational fields after graduation. The estimated two-child rate on average decreases to 32.2% when the women are or has been married at age 40, holding all the variables in the model at zero. When increasing the variable for the gender ratio in a municipality where a woman lived at age 27 by one unit the estimated two-child rate on average is 80%, holding all the variables in the model to zero. When educational level of the father is high the estimated two-child rate on average decreases to 77.5% and if the mother also has a high educational level the estimated average two-children rate is 77.7% when holding all the other variables in the model at zero. The variable controlling for where the women lives at age 40, when holding all the other variables in the model to zero, has the estimated two-children rate on average when living in: Medium-sized town 82%, small town 79%, commuting municipality near a large city 80%, commuting municipality near medium-sized towns 79.9%, commuting municipality near small towns 79.6%, commuting municipality with a low commuting rate near medium-sized towns 79.7%, rural municipality 74.1% and rural municipality with a visitor industry 78.2%.

For this process the variable for the age at first birth has a random slope coefficient. The random slope coefficient has a level-2 intercept of 0.1743. After transformation, the estimated two-child rate on average is 83.7% for a year unit increase in the mean age at first birth, when holding every other variable in the model at zero. The level-2 covariate, the expected average annual income of an educational field, can be understood as a moderator variable for the relationship between the mean age at first birth and the two-child rate. The expected average
annual income of an educational field explains some of the variation between the educational fields. The relationship between the mean age at first birth and the dependent variable varies according to the value of the moderator variable. The slope coefficient for the moderator variable is -0.0086. For a one unit increase in the level-2 variable, the estimated two-child rate on average drops to 83.6%.

6 Conclusions

The results in the previous chapter demonstrate some disparity between the intraclass correlations between the three different data generating processes. The process of being childless had the highest intraclass correlation of 0.11, the process of having one child versus having more than one child had a ICC of 0.104 and the process of having two children versus having more than two children had the lowest ICC of 0.08. The variation between the intraclass correlations validates three distinct data generating processes. It can also be concluded that the variations regarding number of births between the educational fields becomes lower when women have more children. The choice of going from childless to having one child is presumably the greatest cost. When a woman already has a child it is likely to expect that the additional cost of having one more child is lower than going from zero to one. The additional cost of having two children compared to three children and so forth should be even lower and the ICC between the educational fields can be concluded to be lower for any additional child.

In the results of attempting to describe the variation in childlessness between educational fields, the variables that indicate if a woman is or has been married varies substantially across the educational fields. It is not surprising that if a woman that has been married or is married, has great negative impact of the risk. That is, if a woman is married the risk of being childless is smaller. However, a more surprising thing is that the expected childlessness rate of being married varies between the educational fields. One reason could be that women who attend educational fields that has a low childlessness level such as Nursing and Teaching surely have a higher probability of getting married compared to women that attends other educational fields. For example, since the highest level of childlessness is in the fields of Humanities and Art, it is likely to expect that women who attend that field also have a lower probability of getting married.

The second-level variable, the expected average annual income, models the differences
between the educational fields in the relation of ever being married and childlessness, and has a negative impact. So, if a woman is married or ever has been married and attended a field of education that results in job with an expected average annual higher income, then the estimated childlessness rate becomes even smaller. The results is quite reasonably since raising a child is not cheap and when attending an educational field that leads to a stable job with a good income, married women should feel more secure to have a child.

Another interesting result is that the educational level of the father of a woman was positively associated with childlessness. Women with highly educated fathers might be more encouraged than others to attend educational fields that are male-dominated. Since the male-dominated fields lead to male-dominated jobs, which are somewhat less encouraging of childbearing, it could have some effect in the choice of being childless (Hoem et al., 2006b).

The second model attempts to declare the variation in having one child versus having more than one child between the educational fields. The fact that there was no random coefficient, in comparison with the childlessness model, gives support to the notion that these are different processes. The variation might depend on a social factor that cannot be measured or a factor that is not included in this study.

The result for the second process also declares a negative relationship between just having one child and ever being married. So a woman that is married/been married has a lower estimated one-child rate. A reason of that could be that married women might feel more secure with their partner than those who are not married, which leads to more than just one child. The estimated one-child rate increases the older the women are when they have their first child. Which is reasonable since it gets harder for women to get pregnant the older they become. The varying mean through all the educational fields has a negative relationship with a higher average annual income. The educational fields that results in a job where the annual income is higher than average, has a lower estimated one-child rate.

The third model attempts to declare the variation between the educational fields of having two children versus having more than two children. The mean age at first births varies between the educational fields. This variation might be due to that some educational fields are female and male-dominated, and that the job prospects look different across fields. For example, women that study Humanities and Art, which has the lowest mean number of children, might struggle with getting a stable job after graduating and therefore postponing childbearing. If the first child is born when the mother is of relative old age, the willingness of having more
than two children might decrease. Women that work through the government in fields where there exists staff shortages and it is easy to get a job, in for example nursing or teaching, might feel more secure in their work situation and therefore might have their first child in an early age. An early age at the first birth might make women more willing to have more than two children. The income variable moderates the negative relationship between age at first birth and childlessness. If the average annual income is higher than the average income, then the tendency of just having two children decreases.

7 Discussion

There are some limitations in this study that will be addressed below. The population size was approximately 250,000 women which a random sample of 125,000 women were drawn from. It can be questioned if some essential part of the population was neglected from the sample drawn. Since the sample was randomly drawn and the sample size is large, the sample should not neglect any essential group of the population and should not pose any problems in the analysis. Another factor that should be discussed is the significance of the estimates in the paper. Since the sample size in the study is large even the smallest difference will be significant. The focus should instead be on the effect size of the estimates. Also, the three models analysed have different sample sizes which results in different power in the estimation of each model. The smallest sample size is 89,479 women and the largest is 125,000 women. The sample could instead have been drawn so there were equally many subjects in each model. However, since the smallest sample size in this study is large anyway, this poses no problems. Also, the 89 educational fields that were excluded in this study due to the few number of observations in each field, could be examined further. The excluded fields could for example be examined graphically for those who are interested in the differences between them regarding childlessness and the number of children.

The educational fields explained 0.08-0.11 of the total variation regarding childlessness and childbirths. There are obviously a lot of other important factors explaining the childbirth processes of women, which are not included in this study. For further studies it might be interesting to look at a different clustering variable, for example different workplaces.

Since this paper investigates the processes of being childless, having one and having two children, it is important to discuss how the number of children is measured. All the births a
woman has are accounted for even if for example a woman delivers a stillborn. The woman will then be placed to the group of women that has one child, even though she is in fact childless. Also, those women who have twins, triplets and so forth will for example be placed to the group that has one child, even though they have two or more children. This paper does not account for these special cases. However, this should not be an issue in this study since there is approximately 3% of all births that are twins, triplets et cetera and the percentages of stillborn is far below 1% (Tollebrant, 2011).

Some alternative explanatory variables that would be interesting to examine that was not included in this paper, would for example be a variable that indicates if the choice of being childless is due to medical difficulties or not. It could be interesting to see if the number of siblings of a woman has some effect on the number of children she has. Also, the level-2 variable, the expected average annual income, has a small effect in all three models. There might be even better choices for second-level covariates. It would have been interesting to have a gender ratio variable for every educational field. It could also be interesting to have a variable to measure the job prospects after graduating from each field.

The aim of this study was to further investigate Hoem et al. (2006a,b) findings of variability between the Swedish educational fields regarding childlessness and the number of children a woman chooses to have. The results in this paper agree with the findings in previous research regarding the variability between the fields. One of the main findings are that the relative variability between the educational fields is highest for childlessness and is decreasingly smaller for having one and having two children.

The process of having one child did not vary across educational fields for the covariates considered. The effect of the level-2 covariate, the expected annual income, had a small negative effect on the intercept/mean of the process of having one child. Women who attended an educational field that lead to a job with an average higher annual income than other fields, had a higher expectancy rate of having more than one child.

The effect of the age of the first birth on two-child rate varied across the educational fields. If a woman was above the average age when she had her first child the estimated two-child rate increased. Furthermore, women that were above the average age when having her first child and attended an educational field that resulted in a job with a higher expected average annual income than other fields, had a lower two-child rate.

These findings support the results in Hoem et al. (2006a,b). There seems to be differences
in educational fields that are important to childlessness and the childbirth rate of women. It is likely that educational fields are a better proxy for income and employment stability than the educational level.

**Acknowledgement**

I would like to thank the administrative agency Statistics Sweden for providing the data in this paper. I would also like to thank Vitor Miranda for guidance and useful feedback.
References


