UCD School of Economics

Family-structure and children's educational attainment

Zaur Tahirovich Gouliev

Student ID: 18718545

An undergraduate thesis submitted in part fulfilment of the degree of

BSc. (Hons.) in Economics

Supervisor: Dr. Kevin Denny

9th June 2021
JEL classification: J12, I20, I30. Key words: Family structure, education.

Abstract

This paper aims to investigate the relationship between family structure and educational attainment. Traditional and classical economic thought surrounding parent structure makes the claim that having both parents in the household is fundamental to a child's development and that there is strong support for this claim. This paper aims to isolate a component of development, which is education and focus on seeing if there is any difference in the educational attainment levels of a child when looking at the differences in their family structure, i.e., two-parent households vs one-parent households. The current body of literature indicates a positive relationship, suggesting that those who come from nuclear households have better life outcomes, including education. This does not go without debate, as there is also confounding research that finds little to no links. This unsettled question warrants the interest and nature of this paper. Using comprehensive data from the Growing Up in Ireland (GUI) dataset, We classify family structures and compare them against reading and mathematics test scores (educational outcomes) alongside the parent's emotional well being (socio-emotional outcomes). We run OLS regression analysis in attempt to draw empirical evidence. The results show statistical significance and a positive relationship in family structure on educational attainment of 2.9% on math test scores and 2.6% on literacy test scores of children and a negative relationship of 0.8% on socio-emotional wellbeing of the parents. The limitations of this paper are also discussed, alongside further considerations.

Acknowledgement

I would like to thank Dr. Christopher Jepsen [1] and Dr. Kevin Denny [2] for their continuous support and guidance in writing this thesis. I would like to also thank the UCD School of Economics for their high quality teaching, they have equipped me with the right tools to think like an economist. Their BSc Economics program is of the highest quality of academic teaching.

[1] Christopher Jepsen is a lecturer at the School of Economics at University College Dublin (UCD). His primary research areas are in the economics of education and labor economics, with a particular focus lately on the labor-market returns to vocational postsecondary schooling.

[2] Kevin Denny is a Senior Lecturer in the School of Economics, University College Dublin. His current interests are various but broadly speaking are empirical and microeconomic with an interest in behavioural and life sciences. They include educational disadvantage, the costs of labour in maternity hospitals, inequality and educational attainment, social returns to education and attitudes towards immigration.

Table of Contents

Chapter 1: **Introduction**

The number of children that are being raised outside of two parent households has steadily been increasing in the last few decades around the world. In the **1970s**, there were *3,000,000* single parenting mothers and *393,000* single parenting fathers in the United States, this rose to *10,000,000* single parenting mothers and *2,300,000* million single parenting fathers by **2006** (Garfield, 2009). In Ireland, the population which will be studied in this paper, we observe data showing one parent families with children increased by 1.5 per cent to *189,112* between **2011** to **2016** (CSO, 2016), and data analysed by the Pew Research Center (2017) have found that about one-in-five children are living with a solo mother in the United States. In similarity, from data analysed by the Central Statistics Office (2016) we find that 1 in 5 people in Ireland live in a one-parent family and 1 in 4 families with children in Ireland are one-parent families. In this paper, we will not go in depth as to why the probable causes of this might be, but it is noteworthy to mention that these rising figures can be directly attributed to increasing divorce rates, unintended pregnancies and the death of a spouse – all which can result in a single parent family.

A rise of single parent families poses multiple concerns for economists and policy makers as a whole, one obvious concern is the inequality of income that can arise. If we look at single mothers, we see that they face much higher financial difficulty than two-parent households and this largely boils down to the possible income sources that a single mother has access to; the labour market, the absent parent and/or the welfare state (Hakovirta, 2011). Public policy has been aimed over the last 30 years in Ireland to attempt to address these concerns. Services and entitlements have been set up by the Irish Government such as One-Parent Family Payment (OFP), Jobseeker's Transitional Payment (JTP) and a special tax credit called the Single Person Child Carer Credit (SPCCC). Other entitlements are also available such as medical card and child benefit which all, in large part aim to ease the costs associated with raising a child alone. These benefits have been shown to be effective, and cross-national studies regarding the effectiveness of social welfare programmes and their ability to reduce poverty has held up (Kenworthy, 1999) and shown to have strong support for relieving poverty. Studies by Gregg and Harkness, (2003) in the United Kingdom found that welfare reform and employment initiatives show substantial benefits for lone parents. They found that changes in UK government policy in the **1990s** that addressed how to better go about enabling lone parents to return to employment was successful in increasing the lone parents employment from 40 per cent in **1993** to 53 percent in **2002**, and these changes were more pronounced when the New Deal for Lone Parents (NDLP) and Working Families Tax Credit (WFTC) came into effect indicating that policy reform can be a major contribution to tackling income inequality between family structure differences.

This paper makes an attempt to try to understand how the composition of a household, i.e., one-parent vs. two-parent households affect a child's development. It employs the Growing Up in Ireland (GUI) dataset, which we will go into in further detail in Chapter 3: Data and Methodologies. This data contains information on socio-economic outcomes and educational outcomes, both which we will be running ordinary least squares regression (OLS) to determine what impact family structure has on these outcome variables. This investigation has interdisciplinary interest, since the results can be useful for a variety of reasons such as modelling the correct welfare policy, introducing schemes and education policy aimed at helping those from socio-economically disadvantaged in education and helping the government create better incentives to help vulnerable lone parents.

Chapter 2: **Literature Review**

In essence, the main motivation behind studying this topic is to try answer an important question; do children who have 1 parent do worse in education and have worse-off emotional outcomes than children who have 2 parents in the household? If this is the case then to what measurable extent is it, and what are the economic consequences. In our data set we have access to the a Strength and Difficulties Questionnaires (Goodman; 1997; 2001), which is a relatively rigid psychometric and this allows us to have a close look at the assessment results of emotional well being of the children and determine if there is any statistical significance that would be used as an indicator that family structure has a negative effect on children's emotional outcomes. We also have access to the Drumcondra Tests of Early Numeracy (DTEN) and Drumcondra Tests of Early Literacy (DTEL) test scores, which will give us a metric to measure children's educational outcomes. It is important however to understand that this type of measurement is not the be-all and end-all, as this subject matter is extensive, and so before doing our analysis, a literature overview is crucial in understanding where our knowledge on this topic currently stands.

Family background studies have long been an area of research where economists have had mixed agreement on what extent the structure of a family has on a child's outcome, for instance, research by (Boggess, 1998; Krein & Beller 1988) found that factors such as family structure, parent's education and parent's work patterns have had some role to play in children's achievements but has little or no effect on educational attainment once we control for economic status. Research by (Milne, Myers, Rosenthal & Ginsburg 1986) found that students under perform from single-parent families and have consistently had lower attainments than students from both two biological parents but that "parental investment", i.e., the amount of time parents invest with their children, may be a key role in explaining this, and studies by (Aakvik, Vaage & Salvanes 2005) looked at educational attainment and family background to consider what effects educational attainment had on a children and how credit constraints can affect the attainment levels.

These papers primarily relied on national longitudinal studies. In the case of (Milne et al. 1986) they looked specifically at single parents, and focused on single mothers. They ran ordinary least squares (OLS) regression on data from The High School and Beyond (HSB) study which contained data on *58,270* high school students. They looked at intervening variables such as mothers education, mothers employment status (broken into average hours worked) and household income. These were measured to see how they related to the student's Comprehensive Test of Basic Skills (CTBS) which is an almost identical metric to the one used in this study, Drumcondra Test Scores, as a benchmark for educational attainment, similarly it measured the students reading and math scores. The results were statistically significant at the $.05$ level and 0.01 level and found a positive association between single parent families and educational attainment. It investigated further and found that the effect of number of parents balances out when family income is added, and it stands out as the most important of the intervening variables. It concluded that other studies had failed to control for income, and understated how important this variable is when analyzing studies regarding one-parent families. The paper also put a large emphasis on time investment, i.e., parental investment and saw positive changes to student's educational attainment when time spent with the student was increased.

The study done by (Aakvik et al. 2005) also used OLS regression to estimate the effects of family background variables on the educational attainment of the child in Norway, this study measured family income at different periods of a child's life in an attempt to create a distinction between short-term effects vs long-term effects. It used data from the national statistics registry in Norway and had a huge pool of *N=155,752*. The basic econometric model they used was as follows:

$$E_i = \beta_0 + I'_f\beta_1 + E'_f\beta_2 + W'_m\beta_3 + S'_f\beta_4 + \gamma_C + \delta_M + \varepsilon_i$$

I_f is the vector of family income variables which consists of both the father and mother's income. E_f is the family education variables which consists of if either the mother or father has higher education. W_m is an interesting variable which counts whether or not the mother was employed in the child's life from ages 4-6 (pre-school) and age 8-10. S_f is the vector that considers the amount of siblings the child has. The variables δ_M and ε_i looked at the age cohort and the country where the family grew up. The paper results indicated that early years, such as 0-6 years of a child are most crucial for educational attainment than later years. Factors such as permanent family income and parental education were key variables in determining the development of the child.

Boggess (1998) used data from the Panel Study of Income Dynamics (PSID) which was gathered by the University of Michigan's Survey Research Center (SRC). This data was also a longitudinal survey. He attempted to examine the effect of family structure on high school graduation when categorized into race and gender. It looked at approximately 1985 young adults who turned 17 and the sample looked at mother-only, mother-stepfather or first families at age 17. The dependant variable was high school completion which was equal to 1 if the respondent self-reported school graduation, or 0 otherwise. The independent variables included various measures of family structure, this ranged from father absenteeism, mothers or father's education, the region the family lived in, household income and employment experience. These measures were mostly broken down into discrete variables, for instance $0 - 18$ indicating the number of years of completed education of parents. In this study a logistic regression model was used with a total of 9 model specifications, of which the first 2 included only independent variables for family structure. The results found in this paper were statistically significant (< 0.05) and showed a number of things, for instance growing up in a stepfather family had a significant negative effect of a white male and female's likelihood of high school completion as opposed to a two-parent household by up to 1 percentage point. The results were also the same for growing up in a single mother household for white females, lowering the probability of high school completion by almost 1 percentage point. The same also applied for black males and females, showing that the absence effect on a parent during childhood reduces the probability of graduation by up to 15 percentage points, however, in further model specifications when the researchers begin to control for the family's economic status, they find that the traditional family structure has much less of an effect associated with high school graduation. This would indicate that economic circumstances that are typically prevalent in single parenthood may be an explanation for the results, rather than the family structure itself. The researchers conclude the significant positive effect of income on educational attainment, however the paper concludes that it generally supports the result from numerous other studies that growing up outside the traditional two-parent household has a negatives effect on educational attainment, but that this seemingly lessens when controlling for economic circumstances.

The final paper that is researched is by Krein et al. (1988), they examined the effect of living in a single-parent family on educational attainment by gender and race. The basis of the study was a focus on the household production theory, a theory proposed and developed popularly by economist Gary Becker (1965; 1981). The general idea surrounds human capital investment, and so if a reduction in parental resources occurs in children living in a single-parent family, naturally, we should find that educational attainment will also lower. The data utilized by the researchers in this paper are from three data sets which look at two generations; mothers and their sons or daughters taken from the **1967-1979** Surveys of Mature Women (mothers), the **1966-1980** Surveys of Young Men (boys), and the **1968-1980** Surveys of Young Women (girls). This data are from national longitudinal surveys. The empirical model used here was as follows:

$$S = \ f(T_{1,...,n}, X_{1,...,n}, E_{1,...,n}),$$

Here S is the number of years of formal schooling completed by age 26. $T_{1,...,}$ is a vector of

time inputs which consists of T_1 which is one of the two measures of living in a single-parent family, T_2 which is whether the mother had employment outside of the home, T_3 is the number of siblings present in the family. The $X_1, ...,_n$ is a vector of income (goods) inputs which consists of X_1 which is total family income during the high school years, X_2 is the availability of reading materials. $E_1, ...,_n$ is the vector of the ability and human capital of the parents and other control variables, this consists of E_1 the education level of the mother, E_2 the education level of the father and E_4 is the region of residence. The paper used ordinary least squares regression analysis with different specifications of living in a single-parent family, one equation looked at the number of years an individual spent living in a one parent family, while the another equation looked at the number of years that were spent living in a one parent family during each period of childhood which was split into preschool, elementary and high school. The empirical findings were consistent with the idea that living in a single-parent family has an overall worse off effect on the educational attainment of a child, but this paper added a more diverse angle, and looked at period and length of exposure to living in a single-parent household and noticed that it was primarily during the preschool years that single parenthood has the most biggest negative effect. Interestingly, this paper found that the negative effect is larger for boys than for girls. The findings by this paper are consistent with the household production theory by Becker (1965; 1981) which isn't too surprising since their is less investment in human capital for the children living in a one-parent family as opposed to living in a two-parent family.

If we take a critical look at the literature and try to summarize what can be learned, it seems obvious to suggest that a child has better outcomes when living with both parents. The findings from studies mentioned above, and many others research papers in the body of economic work give strong empirical evidence for this. The findings give us an impression that the severity of the effect can be balanced out when variables such as family income, parent education and region of residence are added to the models. In our paper, we will include these 3 independent variables also, to see what explanatory power it may or may not have.

Chapter 3: Data and Methodologies

3.1 Data Description

The data utilized for this research project is taken from the first wave of the Growing Up in Ireland (GUI) study. The Growing Up in Ireland is a national longitudinal study of children in Ireland (ISSDA). It is representative on a national level of children who live in Ireland. It tracks the development of a two cohorts of children born between **1998** and **2008**. The GUI was funded by the Irish government and it was jointly carried out by the ERSI and Trinity College Dublin. It is currently managed by the Department of Children, Equality, Disability, Integration and Youth. It is a unique study, and is a one of a kind study in Ireland. The data is a particularly well known source and has been used for many well cited papers on a range of different subjects due to the comprehensive nature.

The GUI follows two cohorts: an infant cohort and a child cohort. The child cohort began in **1998**, who are now 22 years old and the infant cohort began in 2008 who are now 12 years old. In this paper we will focus on the child cohort. We are using panel data for a specific cohort for the specific year, we will not be utilizing multiple observation over time for the same individuals.

The primary aim for the study was to inform government policy in relation to young people and families. This would allow for statistical analysis that could help drive well informed policy to help tackle issues facing young people in Ireland. The data was collected for the child cohort in **2008**. It consisted of *8,500* children who were aged 9. Information was collected from parents, teachers, principals and the children themselves. There is a host of different topics explored in the study, these range from health & wellbeing, emotional & cognitive behavior, economic wellbeing and more. Using this data we can gain many insights, and explore different factors that effect children's lives in Ireland, for this reason we chose this data source to examine family status in relation to educational attainment. This data was requested and accessed through the Irish Social Science Data Archive (ISSDA). In this next section we go into more details regarding the variables I investigated.

3.2 Variables

Due to the nature, i.e., sheer size of the GUI dataset, it contained more variables than required for this project so I had to do some cleaning in order to isolate it into exactly what we needed.

The variables that I chose to use for my regressions are in the table below, I give a detailed explanation in the description and labels section of the table, to give an idea of what housework needed to be done in order to make the regression run correctly and interpret easier.

Table 3.1: Description of Variables used in this Study

Code	Definition and Labels
fam_struc	This variable indicates if there is a partner in the household, it is a binary value, where 0 represents if no partner is present and 1 represents if a partner is present.
h_income	This variable represents the equivalised annual household income. It ranges from a minimum of €503.7783 and a maximum of €223115.6. The average income is €21258.82. Equivalised income is the measure of household income that takes into account differences in household sizes and composition. It is calculated using the modified OECD equivalence scale. For this regression, we used and created a log of this income.
urb_rur	This is a binary variable for region, indicating if the family in question lives in either 0 which represents an urban region or 1 which represents a rural region. In this data, N=3,890 live in urban regions and N=4,659 live in rural regions.
read_score	This variable indicates the child's Drumcondra literacy score, it contains the number of correct answers on the test for each individual child. The questions themselves range from a minimum of 0 correct answers to a maximum of 40 correct answers, the variable is stored as percentages and is represented as 0% - 100%. The average score in the sample is 27 answers correct.
math_score	This variable indicates the child's Drumcondra mathematics score, it contains the number of correct answers on the test for each individual child. The questions themselves range from a minimum of 0 correct answers to a maximum of 30 correct answers, the variable is stored as percentages and is represented as 0% - 100%. The average score in the sample is 15 answers correct.
distance	This variable contains the distance of school from the child's home. It has 4 criteria which range from less than 1km, in between 1-2km, in between 2-8km and over 8km. The average distance a child travelled to school in this sample was 3-4km.
gender	This variable contains the gender of the study child, it is a binary variable which contains 0 for male (N=4,024) and 1 for female (N=4,242).
birth_wgt	This variable contains the birth weight of the child, it is measured in kilograms (kg), the minimum is 1.7kg and the maximum is 6.1kg. The average weight of the child is 3.51kg.
household_tot	This variable contains the total number of people in the household, the minimum is 2 and the maximum is 7 or more. The average number of people in the household is 5.
teach_exp	This variable contains the number of years the primary school teacher has been teaching at primary school level. It ranges from 1 year or less to over 40 years of teaching experience. In this sample, the average teaching experience was 13 years.
pock_mon	This variable indicates if the child in question receives pocket money per week. The lower bound is €0 and the higher bound is €25 or more. The average amount is €2.6 per week.
smk_preg	This variable indicates if the mother smoked while pregnant, the variable contains 3 answers, 0 for "yes daily" N=6,402, 1 for "occasionally" N=787 and 2 for "never" N=6,402.

Code	Definition
sdq_score	This variable contains the total score on the Strengths and Difficulties Questionnaire. This is an aggregated score and consists of 5 sub scales such as emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and pro-social behaviour. These sub scales are scored individually and an overall score is given. The score ranges from 0 as the lowest to 37 as the highest.
trav_tosch	This variable tells us how the child travels to school. It contains options ranging from walking, taking public transport (N=66), by bicycle (N=109), by car or a school bus (N=1,025). The majority go to school by car (N=5,278) or walk (N=2,024).
maternal_educ	This variable indicates the highest level of maternal education completed to date. It ranges from none, primary education to postgraduate education.
trav_backsch	This variable tells us how the child travels back to school. It contains options ranging from walking, taking public transport (N=75), by bicycle (N=102), by car or a school bus (N=1,183). The majority come home from school by car (N=4,869) or walk (N=2,268).

The table in Appendix, Table 7.1 contains a more detailed description of the variables, i.e., summary statistics.

It is noteworthy to mention that most of our independent variables are not continuous, but ordinal. The survey participants were not given exact choices, this is to say that the number of choices were quite high so it would be considered **censored data**, because the last category in a lot of the independent variables is e.g., 8km or over, €25 or over, or 40 years teaching experience or over. One possible solution, would have been to model them as ordinal variables, however we do not do this but it is worth mentioning that this would be the ideal way to measure the true effect of our key independent variables. In subsection 3.5, we acknowledge this problem and offer a solution to it.

Below you will find a list which contains a diagnostic test on our variables to test for collinearity before our regression is run, this is to check for confounding variables.

Table 3.2: Collinearity Diagnostics

Variable	VIF	SQRT VIF	Tolerance	R-Squared
fam_struc	1.40	1.18	0.7158	0.2842
gender	1.01	1.01	0.9853	0.0147
birth_wgt	1.04	1.02	0.9570	0.0430
smk_preg	1.14	1.07	0.8735	0.1265
distance	1.32	1.15	0.7576	0.2424
trav_tosch	2.50	1.58	0.4004	0.5996
trav_backsch	2.54	1.59	0.3934	0.6066
pock_mon	1.07	1.03	0.9356	0.0644
maternal_educ	1.28	1.13	0.7835	0.2165
urb_rur	1.09	1.04	0.9164	0.0836
teach_exp	1.01	1.01	0.9860	0.0140
h_income	1.41	1.19	0.7111	0.2889
household_tot	1.36	1.17	0.7363	0.2637

Mean VIF: 1.40

3.3 Methodology

In this paper we use ordinary least squares regression (OLS) for my analysis with robust standard errors. We will examine the coefficients on the OLS regressions, this will give us an interpretation of how many units does our dependent variable change when we increase our independent variable for 1 unit. In a hypothetical scenario, if we take family structure (0 vs. 1), a 1 unit change indicates having a single parent vs. having a family and we will look at what that change lead to. In our dependent variable, for example, a 1 unit change would mean a percentage as our dependent variable and since it is measured from 0 to 100, this means a unit change will equal to a 1% change.

We have 3 model specifications which are as follows:

Model Specification **(1.1)**

$$\begin{aligned}\text{Math.Score}_t = \alpha &+ \beta_1 \text{Family.Structure}_t \\ &+ \beta_2 \text{Household.Income}_t \\ &+ \beta_3 \text{Mother.Education}_t \\ &+ \beta_4 \text{Urban.Rural}_t \\ &+ \beta_5 \text{Distance}_t \\ &+ \beta_6 \text{Travel.ToSchool}_t \\ &+ \beta_7 \text{Travel.FromSchool}_t \\ &+ \beta_8 \text{Gender}_t \\ &+ \beta_9 \text{Pocket.Money}_t \\ &+ \beta_{10} \text{Birth.Weight}_t \\ &+ \beta_{11} \text{Smoking.Pregnancy}_t \\ &+ \beta_{12} \text{Teaching.Experience}_t \\ &+ \beta_{13} \text{Household.Total}_t + \epsilon\end{aligned}$$

Model Specification **(1.2)**

$$\begin{aligned}\text{Reading.Score}_t = \alpha &+ \beta_1 \text{Family.Structure}_t \\ &+ \beta_2 \text{Household.Income}_t \\ &+ \beta_3 \text{Mother.Education}_t \\ &+ \beta_4 \text{Urban.Rural}_t \\ &+ \beta_5 \text{Distance}_t \\ &+ \beta_6 \text{Travel.ToSchool}_t \\ &+ \beta_7 \text{Travel.FromSchool}_t \\ &+ \beta_8 \text{Gender}_t \\ &+ \beta_9 \text{Pocket.Money}_t \\ &+ \beta_{10} \text{Birth.Weight}_t \\ &+ \beta_{11} \text{Smoking.Pregnancy}_t \\ &+ \beta_{12} \text{Teaching.Experience}_t \\ &+ \beta_{13} \text{Household.Total}_t + \epsilon\end{aligned}$$

Model Specification **(1.3)**

$$\begin{aligned}\text{SDQ.Score}_t = \alpha &+ \beta_1 \text{Family.Structure}_t \\ &+ \beta_2 \text{Household.Income}_t \\ &+ \beta_3 \text{Mother.Education}_t \\ &+ \beta_4 \text{Urban.Rural}_t \\ &+ \beta_5 \text{Distance}_t \\ &+ \beta_6 \text{Travel.ToSchool}_t \\ &+ \beta_7 \text{Travel.FromSchool}_t \\ &+ \beta_8 \text{Gender}_t \\ &+ \beta_9 \text{Pocket.Money}_t \\ &+ \beta_{10} \text{Birth.Weight}_t \\ &+ \beta_{11} \text{Smoking.Pregnancy}_t \\ &+ \beta_{12} \text{Teaching.Experience}_t \\ &+ \beta_{13} \text{Household.Total}_t + \epsilon\end{aligned}$$

In model **(1.1)** our focus is the mathematical achievement of the child, we will use this as our dependent variable. Our independent variables will consist of variables that look at the household characteristics such as income, total number of people in household, gender of child, and of course the structure of the family (1 parent vs 2 parent). We also look at the quality of the teacher, as I believe this may be a factor underlying educational attainment, one of the key independent variables to predict teacher "quality" will be the years of experience has. in teaching. Other independent variables that I believe may affect educational attainment have also been included, such as distance from school, the method of travel to school, the method of travel back from school and weekly pocket money the child receives. Some final variables were included purely out of curiosity to determine if they had a relationship, these include the birth weight of the child and if the mother smoked during pregnancy. In all 3 of our model specifications, the same explanatory variables were used and only out dependent variables changed.

3.4 Acknowledgment of Problems and Robustness

In this paper there certainly is a possibility that some problems may occur with the model, and it would be disingenuous to not mention some of these beforehand, as it later forms part of the limitations & further considerations section, found in (Limitations & Further Considerations). Some of the variables used in the model such as pocket money *(pock_mon)* is most likely not so important and it could very well be endogenous. This could make our model less accurate. The variables such as travel to school (*trav_tosch*) and back from school (*trav_backsch*) could both be highly correlated, it is not obvious from looking at the data as to why this would effect the study child's emotional traits. In theory, if you have a child who is introverted and shy or has possible emotional problems, this might result in the fact that they may not take public transport and so the causality goes the other way around. The same logic might apply to distance. These characteristics could cause problems and cause some of our results to be less robust.

In this model and paper, I am using a dataset which contains thousands of variables and I am only choosing a small number of them. In terms of model specification, it is important to note that a mis-specified model can be bias, and if we are omitting variables that are correlated with the variables that I am putting in my equation, then my results will change in unpredictable ways.

To control for this, I will provide a correlation matrix for all my independent variables to see if there is a potential multicollinearity problem. I also run a formal test using VIF, which can be found in our appendix in Table 7.2. This test looks at variables that might be confounding one another. I have manually looked at each of the independent variables to see how the key coefficients in question behave to ensure that they are not just statistical artifacts, rather than causal or correlational relationships. This can occur if our R2 is low and we have statistically significant coefficients.

One other problem is the fact that the ideal manner to estimate our model would have been to ensure the data is ordinal as some of the variables were continuous, the solution here would be to create a series of dummy variables and do predicted value analysis with that and fix our variables at any levels, the model would then give us a constant plus an effect of our independent variable which we can then use to create a situational estimation, e.g., comparing a child with a single parent, living in specific distance from the school is expected to have X amount effect in their math or reading score.

Chapter 4: **Results**

Table 4.1:

	(1.1. DV: Math Score)	(1.2. DV: Reading Score)	(1.3. DV: SDQ Score)
Family Structure	2.930^{***}	2.643^{***}	-0.874^{***}
	(0.890)	(0.876)	(0.215)
Household Income	4.744^{***}	5.463^{***}	-0.828^{***}
	(0.534)	(0.526)	(0.130)
Maternal Education	3.006^{***}	3.346^{***}	-0.450^{***}
	(0.212)	(0.208)	(0.051)
Urban or Rural Region	0.882^{*}	2.129^{***}	-0.017
	(0.497)	(0.490)	(0.120)
Distance	-0.292	-0.381	0.143^{**}
	(0.286)	(0.281)	(0.069)
Travel to School	-0.092	-0.019	0.018
	(0.292)	(0.288)	(0.071)
Travel from School	0.505^{*}	0.322	-0.068
	(0.285)	(0.281)	(0.070)
Gender	-2.559^{***}	0.311	-0.641^{***}
	(0.478)	(0.470)	(0.116)
Weekly Pocket Money	-0.325^{***}	-0.502^{***}	0.045^{**}
	(0.076)	(0.075)	(0.018)
Birth Weight	1.058^{***}	1.376^{***}	-0.244^{**}
	(0.396)	(0.390)	(0.096)
Smoking while Pregnant	-2.123^{***}	-1.651^{***}	0.816^{***}
	(0.366)	(0.360)	(0.089)
Teaching Experience	0.075^{***}	0.013	0.004
	(0.021)	(0.021)	(0.005)
No. People in Household	0.232	-0.846^{***}	-0.388^{***}
	(0.246)	(0.243)	(0.060)
Constant	37.690^{***}	48.945^{***}	12.959^{***}
	(2.419)	(2.376)	(0.585)
Observations	7,011	6,939	7,064
R^2	0.108	0.125	0.079
Adjusted R^2	0.106	0.123	0.077
Residual Std. Error	19.834 (df = 6997)	19.417 (df = 6925)	4.824 (df = 7050)
F Statistic	64.935^{***} (df = 13; 6997)	76.175^{***} (df = 13; 6925)	46.501^{***} (df = 13; 7050)

Note: *p<0.1; **p<0.05; ***p<0.01

If we take a look at our first regression results below (Table 4.2), we find that the mathematics score of the children is positively correlated with family structure, we see that our coefficient is positive of +2.930. This indicates that on average a child who has two parents in the household will score **2.9%** higher in their mathematics test score. We find that household income plays the largest role in mathematics test scores, showing an increase of +4.744. This suggests a one point increase in household income causes an increase of **4.7%** in test scores for the child. Other explanatory variables such as maternal education have a positive association with the child's mathematical test score. We see a +3.006 coefficient on maternal education, suggesting that on average, children who have more educated mothers will see higher returns in their education in terms of attainment levels of up to **3%** when looking purely at the test score as a benchmark. These 3 explanatory variables are statistically significant with a p value $<$**0.01**. We also notice that whether or not you live in a rural or urban region possibly plays some role in math score achievement, a positive coefficient of +0.882 suggests to us that there is a positive association with living in an urban region, however this is only slightly positive of **0.9%**.

Some other factors such as gender play a large role, and we see a negative relationship (coefficient of +2.559) which indicates that females do better off than males on average, meaning a male will score **2.6%** less on average than females in math score. This is statistically significant also with a p value $<$**0.01**.

There are many of our independent variables in this model that are statistically significant and indicate a relationship between math score, one that is very interesting is smoking while pregnant which shows a coefficient of -2.123, here, on average, mothers who smoked while pregnant will see the math score of their children negatively affected by approximately -**2.1%**.

Table 4.2: Regression Results for Model 1.1

	Dependent variable:
	Math Score
Family Structure	2.930*** (1.466, 4.395)
Household Income	4.744*** (3.865, 5.624)
Maternal Education	3.006*** (2.658, 3.353)
Urban or Rural	0.882* (0.065, 1.700)
Distance	−0.292 (−0.761, 0.178)
Travel to School	−0.092 (−0.572, 0.389)
Travel back School	0.505* (0.036, 0.973)
Gender	−2.559*** (−3.345, −1.773)
Pocket Money	−0.325*** (−0.450, −0.201)
Birth Weight	1.058*** (0.407, 1.710)
Smoking while Pregnant	−2.123*** (−2.725, −1.521)
Teaching Experience	0.075*** (0.040, 0.110)
Household Total	0.232 (−0.173, 0.637)
Constant	37.690*** (33.712, 41.668)
Observations	7,011
R^2	0.108
Adjusted R^2	0.106
Residual Std. Error	19.834 (df = 6997)
F Statistic	64.935*** (df = 13; 6997)
Note:	*p<0.1; **p<0.05; ***p<0.01

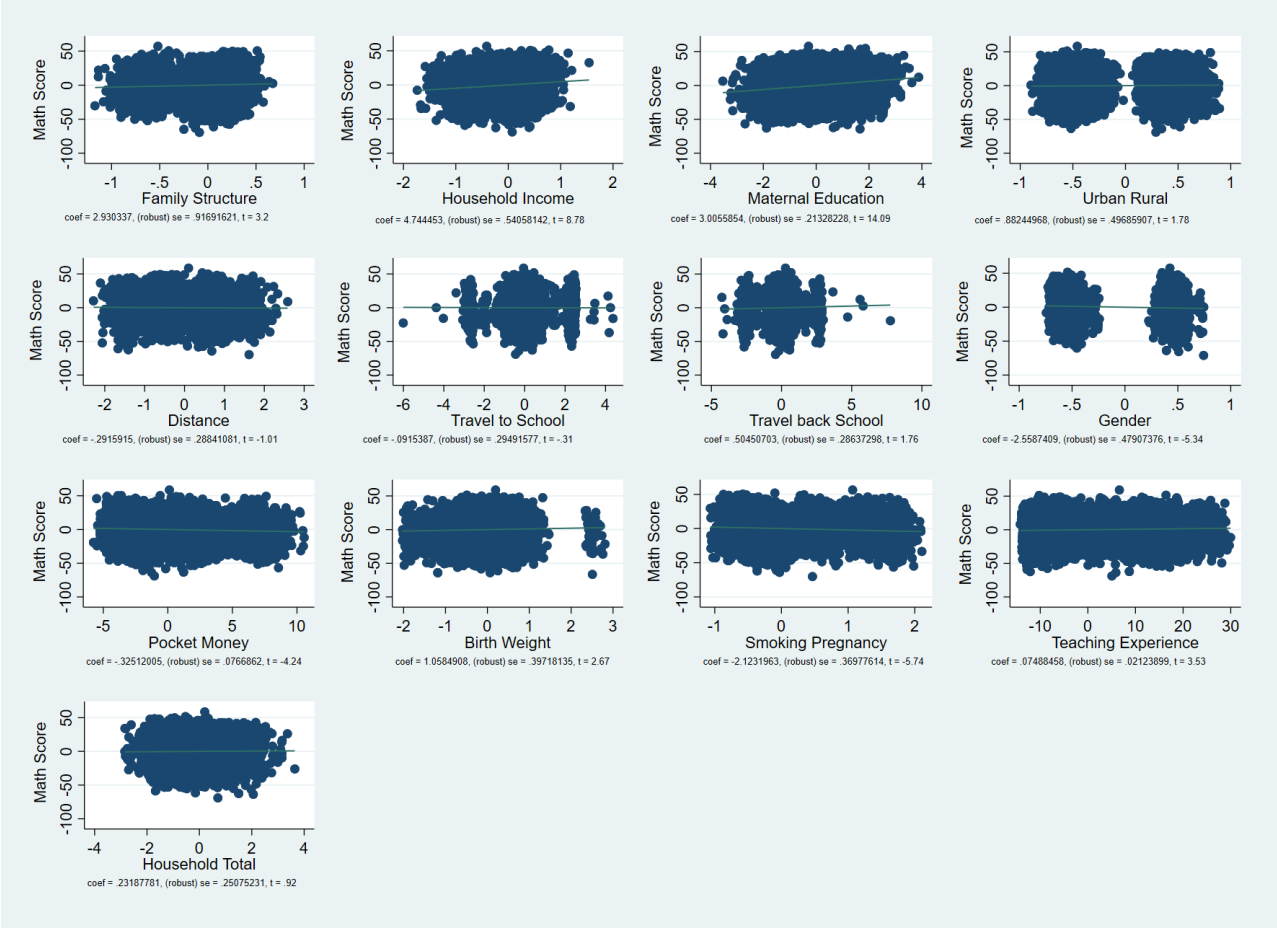

Figure 4.1: AV Plots of Maths Score on Y variables

Our second model 1.2, shows a similar structure, we find a positive correlation between reading/literacy score and family structure which shows a **2.6%** increase associated with the type of family. This indicates that on average a child with a single parent family, does **2.6%** less on their test score than a child with both parents in the household. This score seems to be less affected than our first model, suggesting that mathematics may be a harder subject and its learning is more affected by household structure. We find that income has a much higher weight in this model, with a coefficient of 5.463, meaning that on average a child scores **5.5%** higher than a child when the income is higher in the household. Similarly to our first model, the region the child grew up in plays a large role also, suggesting a **2.1%** increase in test score when the individual lives in a urban region. Maternal education seems a little more important in literacy test scores, as we find an average of **3.3%** increase in reading test scores when the mother has higher education. All of these variables are significant to the p value $<$**0.01**.

Table 4.3: Regression Results for Model 1.2

	Dependent variable:
	Reading Score
Family Structure	2.643*** (1.202, 4.083)
Household Income	5.463*** (4.597, 6.328)
Maternal Education	3.346*** (3.004, 3.688)
Urban or Rural	2.129*** (1.323, 2.934)
Distance	−0.381 (−0.844, 0.081)
Travel to School	−0.019 (−0.492, 0.454)
Travel back School	0.322 (−0.140, 0.783)
Gender	0.311 (−0.462, 1.084)
Pocket Money	−0.502*** (−0.624, −0.379)
Birth Weight	1.376*** (0.735, 2.017)
Smoking while Pregnant	−1.651*** (−2.243, −1.059)
Teaching Experience	0.013 (−0.021, 0.047)
Household Total	−0.846*** (−1.245, −0.447)
Constant	48.945*** (45.037, 52.853)
Observations	6,939
R^2	0.125
Adjusted R^2	0.123
Residual Std. Error	19.417 (df = 6925)
F Statistic	76.175*** (df = 13; 6925)
Note:	*p<0.1; **p<0.05; ***p<0.01

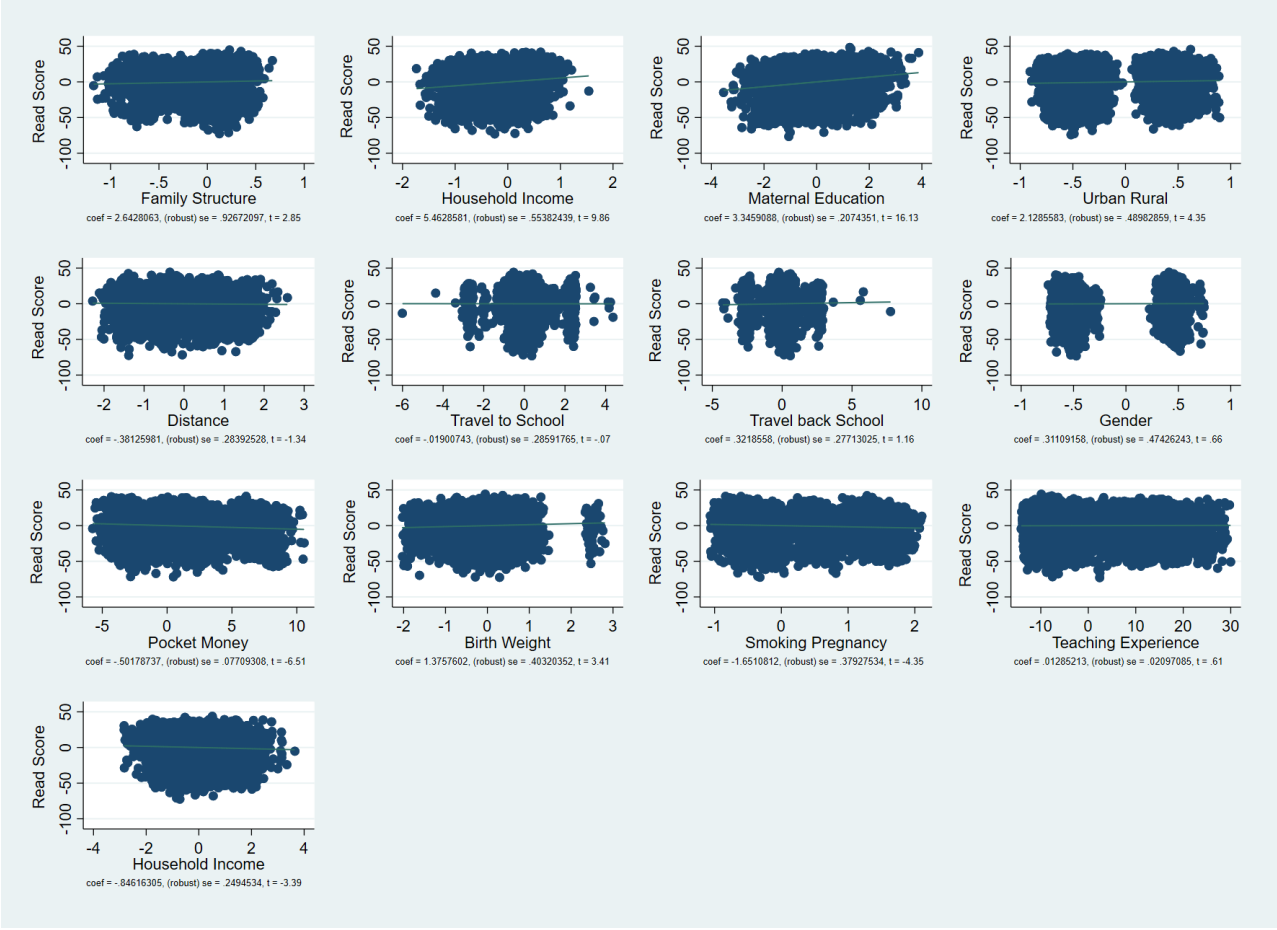

Figure 4.2: AV Plots of Read Score on Y variables

In our third and final model, we see that family structure has a very small effect on the socio-emotional score of a parent, but only by **0.9%**, similarly household income only effects this by **0.8%**. We see that maternal education has a very minor effect on socio-emotional status of **0.4%**. This negligble amount may be difficult to quantify in terms of real world effects.

Table 4.4: Regression Results for Model 1.3

	Dependent variable:
	SDQ Score
Family Structure	-0.874^{***} $(-1.228, -0.520)$
Household Income	-0.828^{***} $(-1.041, -0.615)$
Maternal Education	-0.450^{***} $(-0.534, -0.366)$
Urban or Rural	-0.017 $(-0.215, 0.181)$
Distance	0.143^{**} $(0.029, 0.256)$
Travel to School	0.018 $(-0.099, 0.135)$
Travel back School	-0.068 $(-0.182, 0.047)$
Gender	-0.641^{***} $(-0.831, -0.451)$
Pocket Money	0.045^{**} $(0.015, 0.076)$
Birth Weight	-0.244^{**} $(-0.402, -0.086)$
Smoking while Pregnant	0.816^{***} $(0.670, 0.962)$
Teaching Experience	0.004 $(-0.004, 0.012)$
Household Total	-0.388^{***} $(-0.486, -0.289)$
Constant	12.959^{***} $(11.996, 13.921)$
Observations	7,064
R^2	0.079
Adjusted R^2	0.077
Residual Std. Error	4.824 (df = 7050)
F Statistic	46.501^{***} (df = 13; 7050)
Note:	*p<0.1; **p<0.05; ***p<0.01

Chapter 5: **Limitations & Further Considerations**

In this paper there are limitations that can definitely make the results and regression less valid. One of which is always of concern when using OLS regression and that is omitted variable bias. In my regression I used variables which I believe affected the educational attainment of a child in the long-run, my focus was the structure of the family in terms of coming form a nuclear household or a single parent household, however, there may be something that is missed, for instance, it is hard to measure short-term changes that occur in day to day life. Behavioural changes can occur that cause an individual to take different paths, this could include picking up bad habits that do not delay gratification, which is a trait that is often highly associated with educational success and attainment. This is not something we have access to in a precise manner, as the factors that determine delayed gratification are multifaceted. Some other possible variables that we did not investigate and that may have a direct relationship to our dependent variables are the number of siblings the individual has, birth order of the child or even available family support and resources for the child such as a grandparent, uncle, auntie, etc. These variables were not introduced in our regression. The question also implies a further conversation of "parental" quality vs education quality. Are these complimentary or substitute? What role does the quality of education play? In our paper we did not have a large enough vector to control for educational quality, some possible variables could have been the type of degree the teacher has (education based or non-education based), the style of teaching (authoritative or laizze-faire), the class size, the ratio of boys to girls, and the type of homework/assessments given. These things may play a huge role in determining the attainment levels. On the other side, measuring "parental" quality is also not clear cut, we assumed that it was purely the fact that having 1 parent or 2 parents that are causing some sort of negative relationship but we did not calculate the parental investment made by the parent to the child, this can mean anything from the amount of time the parent spends with the child per day, the level of affection given to the child, the type of shared activities and hobbies with the child. The latter may be important, as it is theoretically plausible to imagine that parents who (whether single parent or not) spend time reading every evening or giving them games that are intellectually and cognitively challenging such as crosswords, chess or checkers create a better outcome for their child's future education. We can naturally think of this as simply coaching or mentoring your child with the right tools, or developing a framework of good habits that will guide the child to making better decisions in the future. One example could be grades on a college assignment, if you have been trained as a child to ensure you do your tasks in a timely manner, say for example cleaning your room daily or doing your homework before dinner time, this may translate in the real world in college when you do an assignment for a module well before the deadline giving you extra free time later to prepare and work on other assignments in other modules. This type of methodology may aid in getting better grades, which in return will aid in achieving more in education and opening more possibilities for further study.

Further studies could be done in this paper, instrumental variable regression (IV) could be used to try to find an instrument that will be able to account for the possible unexpected behaviors between variables, for example, health might play a role in educational achievement which in turn plays a role in household income as it is not cheap to support a health problem e.g, Diabetes or Heart Disease, especially in countries like the United States. These correlated things all impact our life and trying to validate a model purely based on single parenthood vs both parents is challenging in this regard because we may make an error on pinning attainment entirely on family structure. Health plays a role in terms of quality of life, and in some cases may make a partner widowed if

their loved one passes away. IV regression is a useful tool that could be implemented further down the line.

If we look at model **(1.1)** and **(1.2)**, our dependent variable is educational attainment (math and reading scores) and our main independent variable is family structure. Here we would find an instrument that has no effect on education of the student, but the structure of the family, what we would be worried about here, is that if instrumental variable estimation is needed is that there are factors we do not observe that could affect both the family structure and the educational attainment, for example socio-economic status (SES). One example is social class may influence the probability of a family being divorced, or a single mother raising a child alone which in return also affects education of the student. The point here would be that we aim to theorize a correct instrument. In this specific case, we could use a factor that would affect the probability of having a single parent family or two parent family and has no affect on math scores or reading scores, which could be the neighbourhood where the parents grew up. This might affect the probability of becoming a single parent but no direct affect on the child, especially if they are living in a different place currently. Albeit this might be a weak instrument, but nonetheless it would be intriguing to use IV regression estimation to determine the validity.

One final issue that was noted as a limitation of our paper was that in our dataset, we originally expected and had 8,568 observations, and when we ran our models we encountered issues with missing data, this problem was caused by some variables that contained missing values, one of these was teaching experience (*teach_exp*) which was missing at random, this may have caused some skewness in our results.

Chapter 6: **Conclusion**

This paper set out to investigate the relationship between a child, his/her parents and what that translates to in terms of his/her education attainment. Our model formulated was set up to control for other factors and look at score levels as some sort of indicator, this would give us a sense of future educational attainment. Our models proved our alternative hypothesis, and showed that family structure can negatively impact the child, and it holds more truth for mathematics test scores than reading test scores. These are somewhat reconciled when we focus on household income, as that plays the largest role in a child's achievement levels. This is consistent with previous research that shows that income is a much more stronger indicator. Our final model showed that single parenthood negatively affects the socio-emotional outcome of a parent, but only my a small amount.

It is not easy to see how these results play out in the real world, the effect isn't so clear when trying to map a picture out of how this plays out in the real world, for instance, we see that a negative effect occurs on socio-emotional outcomes via the SDQ score in our third model, however, but this isn't clear as to why what effect this really has in the general population in terms of mental health and general mental wellbeing. A similar argument can be made for our first and second model, even if we find a negative score on math and reading rates for single parents, it isn't obvious as to what this might translate to, educational benefits for those socio-economically and educationally disadvantaged can reconcile imbalances, or at least attempt to ease any vast differences and taking this into consideration, it makes this question a little more difficult to answer.

Chapter 7: **Appendix**

Table 7.1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
fam_struc.	8,568	.8843371	.319839	0	1
gender.	8,568	1.514006	.499833	0	1
birth_wgt (kg).	8,450	3.516012	.6180295	1.7	6.1
smk_preg.	8,234	1.349405	.6936895	1	3
distance (km).	8,568	2.193161	.9618204	1	9
trav_tosch.	8,568	3.184407	1.286479	1	6
trav_backsch.	8,568	3.080532	1.329485	1	6
pock_mon (€).	8,568	2.644725	3.255914	0	12
maternal_educ.	8,568	3.658613	1.275262	1	6
sdq_score.	8,518	7.393167	5.038099	0	37
urb_rur.	8,549	1.544976	.4980022	0	1
teach_exp (yrs).	7,992	12.73148	11.27952	1	40
read_score (%).	8,356	70.66251	20.81477	0	100
math_score (%).	8,449	56.42972	20.99349	0	100
h_income1 (log).	7,942	1.084192	.5262983	0	1.609438
household_tot.	8,568	4.734711	1.12642	2	7

Table 7.2: VIF table

	VIF	1/VIF
trav_backsch	2.55	0.392753
trav_tosch	2.50	0.399801
h_income	1.40	0.712712
fam_struc	1.39	0.717098
household_tot	1.36	0.736975
distance	1.32	0.755902
maternal_educ	1.27	0.785390
smk_preg	1.14	0.875127
urb_rur	1.10	0.913012
pock_mon	1.07	0.935899
birth_wgt	1.05	0.956408
gender	1.02	0.984680
teach_exp	1.01	0.985701
Mean VIF:	1.40	

Table 7.3: VIF table

	Eigenval	Cond Index
fam_struc 1	11.8019	1.0000
h_income 2	0.6360	4.3077
maternal_educ 3	0.4189	5.3077
urb_rur 4	0.2523	6.8393
distance 5	0.2443	6.9504
trav_tosch 6	0.1520	8.8104
trav_backsch 7	0.1075	10.4798
gender 8	10.1011	10.8023
pock_mon 9	0.0864	11.6887
birth_wgt 10	0.0805	12.1077
smk_preg 11	0.0473	15.8002
teach_exp 12	0.0342	18.5853
household_tot 13	0.0296	19.9792
Condition Number:	38.3542	
Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)		
Det(correlation matrix):	0.1371	

Table 7.4: Cameron & Trivedi's decomposition of IM-test for Model 1.1

	chi2	df	p
Heteroskedasticity	136.27	101	0.0111
Skewness	137.28	13	0.0000
Kurtosis	124.83	1	0.0000
Total	398.39	115	0.0000

Table 7.5: Cameron & Trivedi's decomposition of IM-test for Model 1.2

	chi2	df	p
Heteroskedasticity	237.83	101	0.0000
Skewness	650.12	13	0.0000
Kurtosis	0.23	1	0.6292
Total	888.18	115	0.0000

Table 7.6: Cameron & Trivedi's decomposition of IM-test for Model 1.3

	chi2	df	p
Heteroskedasticity	251.19	101	0.0000
Skewness	295.59	13	0.0000
Kurtosis	37.28	1	0.0000
Total	584.07	115	0.0000

Chapter 8: References & Bibliography

Craig Garfield. **Variations in Family Composition** [*Developmental Behavioral Pediatrics (Fourth Edition)] 2009, Pages 94-102*.
https://doi.org/10.1016/B978-1-4160-3370-7.00009-2

CSO 2016. **Central Statistics Office** [*Census of Population 2016 - Profile 4 Households and Families*.]
https://www.cso.ie/en/releasesandpublications/ep/p-cp4hf/cp4hf/fmls/

Gretchen Livingston. **About one-third of U.S. children are living with an unmarried parent** [*Pew Research Center, April 27, 2018*.]
https://pewrsr.ch/2Fm1qO1

Mia Hakovirta **The income sources of single parents: A comparative analysis** [*LIS Working Paper Series, No. 282] Luxembourg Income Study (LIS), 2001*.
https://www.econstor.eu/handle/10419/160954

Paul Gregg, Susan Harkness. **Welfare Reform and the Employment of Lone Parents** [*Dickens R., Gregg P., Wadsworth J. (eds) The Labour Market Under New Labour, pp. 98-115] Palgrave Macmillan, London*.
https://doi.org/10.1057/9780230598454_8

Lane Kenworthy. **Do Social-Welfare Policies Reduce Poverty? A Cross-National Assessment** [*Vol. 77, No. 3 (March 1999), pp. 1119-1139. Published By: Oxford University Press*.
https://doi.org/10.2307/3005973

Robert Goodman. **The Strengths and Difficulties Questionnaire: A Research Note** [*J. Child Psychol. Psychiat. Vol, 38. No, 5, pp. 581-586, 1997 Cambridge University Press] Association for Child Psychology and Psychiatry*.
https://acamh.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-7610.1997.tb01545.x

Robert Goodman. **Psychometric Properties of the Strengths and Difficulties Questionnaire** [*Journal of the American Academy of Child Adolescent Psychiatry, Volume 40, Issue 11, November 2001, Pages 1337-1345] The American Academy of Child and Adolescent Psychiatry.*.
https://doi.org/10.1097/00004583-200111000-00015

Drumcondra Tests. **Drumcondra Tests of Early Numeracy (DTEN) & Drumcondra Tests of Early Literacy** [*DCU St Patrick's College Campus]*.
https://www.tests.erc.ie/primary-tests-menu

Scott Boggess. **Family structure, economic status, and educational attainment** [*Journal of Population Economics volume 11, pages 205–222 (1998)]*.
https://doi.org/10.1007/s001480050065

Sheila Fitzgerald Krein & Andrea H. Beller. **Educational attainment of children from single-parent families: Differences by exposure, gender, and race** [*Demography of Children and Youth. Demography 25, pages 221–234 (1988)*.
https://link.springer.com/article/10.2307/2061290

Ann M. Milne, David E. Myers, Alvin S. Rosenthal and Alan Ginsburg. **Single Parents, Working Mothers, and the Educational Achievement of School Children** [*Sociology of Education Vol. 59, No. 3 (Jul., 1986), pp. 125-139 (15 pages)*].
https://www.jstor.org/stable/2112335?seq=1

Arild Aakvik , Kjell Vaage and Kjell G. Salvanes. **Educational Attainment and Family Background** [*German Economic Review 6(3): 377–394*].
https://doi.org/10.1111/j.1468-0475.2005.00138.x

Gary S. Becker. **A Theory of the Allocation of Time** [*The Economic Journal Vol. 75, No. 299 (Sep., 1965), pp. 493-517 (25 pages) Published By: Oxford University Press*].
https://doi.org/10.2307/2228949

Gary S. Becker. **A Treatise on the Family** [*ISBN 9780674906990 - Publication Year: 1993*].
https://www.hup.harvard.edu/catalog.php?isbn=9780674906990

Economic and Social Research Institute (ESRI) and Trinity College, Dublin (TCD). **Growing up in Ireland (GUI): National Longitudinal Study of Children** [*Irish Social Science Data Archive (ISSDA)*].
https://www.ucd.ie/issda/data/growingupinirelandgui/

Marek Hlavac. **stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. 2018** [*Central European Labour Studies Institute (CELSI)*].
https://cran.r-project.org/web/packages/stargazer/citation.html/