Influence of Parental Earning on High School Dropout Probability during Covid-19 Pandemic

Master Thesis Presented to the

Department of Economics at the

Rheinische Friedrich-Wilhelms-Universität Bonn

In Partial Fulfillment of the Requirements for the Degree of

Master of Science (M.Sc.)

Supervisor: Dominik Liebl

Submitted in 09/2021 by:

Ying-Xuan Wu

Matriculation Number: 3302951

Contents

Abstract

This paper uses 2020's Current Population Survey data from the US Census Bureau to inspect whether the low to none effect of parental earnings on high school dropout decisions of children found in previous studies alter after the global pandemic has taken effect. The result from both the GLM probit model and IV probit model suggests the relationship between parental earning remains unchanged during the early stage of pandemic. The irrelevancy also applies to racial minority groups and it stays true even if replacing earning by total parental income.

1 Introduction

The eruption of Covid-19 wreaked havoc on students' studies from several aspects. In Japan, an application for major private universities exams in April 2020 dropped by 12 percentage points compared to 2019, and exceeding 1300 students dropped out of higher education institutions from April 2020 to December 2020 with the reason having difficulty paying tuition fees, based on a leading education institute and ministry of education (Kakuchi, 2021). A small sample study (Vovk Viktoria, 2020) in Japan also reveals a decrease in mental state including efficiency and productivity after on-site training was replaced by digital learning. According to "UNESCO Press Release No.2020-73" 2020, 24 million students worldwide are at risk of dropping out from pre-primary to tertiary education after the Covid-19-induced closure in 2020. US National Dropout Prevention Center predicted student number of those who are at risk of fall behind would grow 2 to 3 times, and the dropout rate would spike in 2020 (Hollingsworth, 2021).

Mayer, 1998 suggests there are two possible ways that parental income could affect children's education. Good parent theory is the indirect effect which states that higher income would render parents less pressure than being in poverty, therefore, have the luxury to do better at parenting. Investment theory is the direct effect that suggests that parents would invest time and earnings onto the children. These two effects are hard to be distinguished in practice.

In past researches, parental income usually suggests little to no effect on the educational attainment of children, while factors i.g. parental education that is highly related to parental income, have much more significant effects. Blau, 1999 finds that the effect of current income is small, the effect of permanent income which means income average over time is relatively larger, however, family background characteristics are far more influential. Shea, 2000 uses 2SLS estimation finding that parental income had a negligible effect on children's human capital, except for the case that the father has low educational attainment. Jenkins/Schluter, 2002 uses sample in Germany and found a minor impact of parental income during late childhood, though trivial comparing to parental education or residence state impact. Chevalier et al., 2013 controls the endogeneity of both education and income effect, they find while father's education affects daughter's academic attainment, the proxy of permanent income was insignificant.

Based on J.Causey et al., 2020, the overall dropout rate doesn't change much for students in tertiary education of the United States, but the peak time of intra-term change shifts from March to April, and there is a large reduction in the new enrollment rate compared to April last year. For African American and Hispanic students, there is a rapid rise in taking leaves of absence. Moreover, according to the new data published on their website, the enrollment rate fell to new lows in the 2021 spring.

However, none of the current studies yet re-examine how the secondary education attainment decision is affected by Covid-19 and whether there is a difference in the impact on children from wealthy or humble family backgrounds. Oreopoulos, 2003 shows drop out of high school has a serious consequence as a lifetime wealth and health depletion using exogenous change in minimum schooling age in the US, Canada, and the UK. The global pandemic and lockdown policy causing a severe impact on income and educational attainment. Plus, finishing secondary education has a predominating importance. It is meaningful to re-examine whether the previous study results still applied in this difficult time. Identifying the existence of potential harms to economical minority students is particularly crucial for policymakers to prevent the children from subjugating life-long social-economical depletion.

In this paper, the connection between the high school dropout probability of students around 16 years old to 18 years old and earning of their parents is estimated by generalized linear model (GLM). The data used is the newest annual report from the Current Population Survey (CPS). The link between parental income and dropout probability is also examined. Another estimation between parental earning and dropout choice is done restricted to racial minorities to investigate the effect pandemic might pose on racial minority groups. To rule out the potential endogeneity of parental earnings and children's behavior, instrumented probit regressions are also executed.

The rest of the paper is structured as follows: chapter 2 introduces the properties of models that are used. Chapter 3 explains how CPS data contents are utilized and presents descriptive statistics. Chapter 4 display the estimated coefficients, and tests. Lastly, in chapter 5 conclusion is drawn.

2 Methodology

2.1 Generalized Linear Model use Probit as Link Function

As the model used in Ermisch/Francesconi, 2001 to evaluate educational attainment and parental investment, in this paper educational attainment is assumed to be a choice that is decided by human capital accumulated by parental investment in children and affected by a series of other related factors. The further assumption has been made to assume there exists a value α, α represents the threshold to keep attaining the school. If human capital accumulation is smaller than α, the student would not perceive keep studying as the most beneficial choice, therefore, he would choose to suspend and drop out of high school. The specification is written as 2.1.

$$
\begin{aligned}
H &= Earning\beta_1 + x\beta_2 + e \\
& \qquad\qquad\qquad e \sim N(e)
\end{aligned}
\tag{2.1}
$$

$$
Dropout = 1\{H < \alpha\} = \begin{cases} 1, & \text{if } H < \alpha \\ 0, & \text{otherwise} \end{cases}
\tag{2.2}
$$

The outcome variable, $Dropout$ is a binary choice variable. H represents human capital. $Earning$ represents paternal and maternal earnings. x are other observed variables, this varies by the regression specification, therefore, they are symbolized by x in the deduction process. In the fullest specification, regression (4), in Table 4.1, x includes gender, health status, health insurance coverage, citizenship, and race of the student, family type, family poverty status, residence state, parental job type, parental education, parental health status, parental disability, parental health insurance coverage, and intercept. Capitalized X represents the overall independent variables used, includes $Earning$. To estimate the binary model, the generalized linear model is a common choice, since this could limit the estimated outcome to binomial. Two common choices for the link functions are logit and probit. The cumulative distribution function curve for Logistic distribution and normal distribution have very similar shapes, except at the tail part (Garrett M. Fitzmaurice, 2001). In the following estimation, probit is chosen for discipline tradition and by assuming H is normally distributed. Furthermore, for the later section of instrumental variables (IV), IV probit model allows a simple factorization of the joint distribution into the conditional distribution of the dependent variable given the endogenous variable and the marginal distribution of the endogenous variable, while it would be much more complicated for a logit function (Hansen, 2021).

To simplify deduction, the indicator function is transformed.

It is assumed that:

$$H^* \begin{cases} = \text{-}(H - \alpha) \\ = \text{-}Earning\beta_1 - x\beta_2 - e + \alpha \end{cases} \quad \text{then}$$

$$Dropout = 1\{H^* > 0\} = \begin{cases} 1, & \text{if } H^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

Then the probability of dropping out could be written as follow.

$$\begin{aligned} E[Dropout = 1|X] &= \mathbb{P}[Dropout = 1|X] \\ &= \mathbb{P}[H^* > 0|X] \\ &= \mathbb{P}[\text{-}Earning\beta_1 - x\beta_2 - e + \alpha > 0|Earning, X] \\ &= \mathbb{P}[e < \text{-}Earning\beta_1 - x\beta_2 + \alpha|Earning, X] \\ &= \mathbb{P}[e > Earning\beta_1 + x\beta_2 - \alpha|Earning, X] \\ &= 1 - \Phi(Earning\beta_1 + x\beta_2 - \alpha) \\ &= \Phi(-Earning\beta_1 - x\beta_2 + \alpha) \end{aligned} \tag{2.4}$$

So as shown in Chapter 25.4 of Hansen, 2021, the response probability could be estimated by an index model and normal link function. Here α could be viewed as the intercept of the estimation. Notice that this implies the threshold for dropout decision is considered to be homogeneous.

In the above case, the parameters are explained as the shifts in standard cumulative normal which is not an easily understandable interpretation. Therefore, the marginal effect is also reported in the result part to explain the magnitude of estimated parameters. The average marginal effect(AME) is computed by averaging the derivative for the chosen variable across observations with a slight change each time. The AME for $Earning$ could be written as Equation 2.5. The other common way is to use marginal effect at mean, the estimations from these two methods only differ in minor degrees (Perraillon, 2019). The average marginal effect is chosen considering there are dummy variables in the data set and the interpretation of average marginal effect is more straightforward and meaningful in this case (Perraillon, 2019).

$$\begin{aligned} A\hat{M}E &= \frac{1}{N}\sum_{i=1}^{n}\hat{\delta}(Earning_i) \\ &= \frac{1}{N}\sum_{i=1}^{n}\frac{\partial}{\partial Earning}\mathbb{P}[Dropout = 1|X] \\ &= \hat{\beta}_1\frac{1}{N}\sum_{i=1}^{n}\varphi(-Earning\beta_1 - x\beta_2 + \alpha) \end{aligned} \tag{2.5}$$

2.2 Instrumental Variable Probit Model

Based on the deduction in Hansen, 2021 the probit estimation will be consistent, if the 4 conditions are satisfied:

1. Y and X are i.i.d..
2. $E[X]^2$ exists.
3. Full rank condition holds.
4. Estimated parameters are bounded.

With further assumption that $E[X]^4$ exists, the estimation would be asymptotically normal. However, parental earning is also largely affected by parental education and other unobservables that also affect dropout decisions such as the personality and intelligence of parents. For example, comparing to careless parents, detailed-oriented parents are more likely to obtain career success and take better care of their children. Since the parental earning has an endogenous problem that might violate the i.i.d. assumption, instrumental variables need to be introduced.

$$\begin{aligned} H &= Earning\beta_1 + x\beta_2 + e_1 \\ Earning &= z\gamma_1 + x\gamma_2 + e_2 \\ Dropout &= 1\{H < \alpha\} \end{aligned} \tag{2.6}$$

The new specification could be written as Equation 2.6 above in which z represents the exogenous variable that affects only earning. Then based on chapter 25.12 of Hansen, 2021, it could be estimated by maximum likelihood as long as e_1 and e_2 are jointly normal distributed conditioned on both x and z.

Shea, 2000 assumes earning is affected by two factors, one is the human capital accumulation, the other is luck. To estimate the impact of luck, he adopted 3 instrumental variables: belonging to a union or not, his job industry, and whether he is currently unemployed. In the later estimation, industry and joining the union or not is used as instrumental variables. Considering there are lots of job losses during the lockdown, compounding with unemployment compensation and subsidy that varies by industry and states and there is no explicit specification in data set further complicates the measrument of the effect of job loss on earning. Therefore, this instrument is left out. The relevance and exclusion condition would be further examined in the data section. Also, validness tests would be presented in the result section.

3 Data Description

The data used is the Current Population Survey from United States Census Bureau in the year 2020. The data set used in the estimation is the annual record instead of the monthly tracking survey. It covers the necessary variables that are missing in the monthly file, but the downside is that it is not possible to isolate the date that is before and after the Covid impact. According to the supplement from U.S. Census Bureau, 2020, the file covered 280 counties and 40 central cities. The number of the total persons recorded in 2020 is 157959. To measure the variance of the full sample, 160 replicate weights are also provided to independently create 160 replicate estimates. The final weight that is applied should adjust the sample population to the known distribution of the entire population. All the sample sizes reported in the following results are unweighted while all the regression estimation has used the replicate weights. Sample size reports in regression estimation are also unweighted.

The dependent variable is the dropout rate of teenagers between the age of 16 to 18. Dropout is defined by a teenager without a high school diploma that is not enrolled in a high school, college, or university in the last week. Based on statistics from the Education Commission of the States (Mikulecky, 2013), 21 states in the US set the compulsory schooling year to 16 years old, 10 states set at 17, 20 states at 18[1]. Therefore, in some states, the dropout of high school is allowed by the law. For teens younger than 16, CPS would not show their highest educational attainment. For young adults that are older than 18, it is possible that they have already left their parents and heads to college. In that case, CPS will not be able to trace their parents, since it only records people who live in the same household.

3.1 Family Structure of Sample Students

Table 3.1 report the unweighted mean, standard error, and sample size for each group of teenagers. The group is the combination of dummy variable dropout and age, 1 means the students in that group are dropout students. Sample size, n, increases considerably as the age increases. The variables step/adopt, single, live alone, and both parents are all dummy variables, so the group means could be interpreted as a percentage within the age and dropout group.

The variable step/adopt is a dummy variable that suggests whether one of the parents is a step-parent or whether one of the parent relationships is created by adoption. For the younger age group, the difference in the step/adopt ratio between dropout and non-dropout is large. 36 percent of 16 years old dropout students come from a step or adoptive family while only 5 percent of 16 years old non-dropout students come from a step or adoptive family. The difference is smaller for age 17 years old but it's still 5 percentage points higher for the dropout group. However, it's hard to say whether there is any strong correlation between having a step-parent or an adoptive parent and dropping out before 18 years old, since the sample size for dropout students at ages 16 and 17 is too small.

The same goes for single which is a dummy variable that suggests whether the child only lives with one parent, regardless of whether the type of parent is a biological, step, or adoptive. The difference is large for younger age but roughly the same between dropout or not for the 18-year-olds.

Live alone suggests the teenager does not live with any of his or her parents regardless of the parent type. Percentages of teenagers who live alone are consistently much higher for the dropout

[1]Kentucky raises compulsory schooling age to 18 at 2017-18 school year

groups across ages. Table 3.2 tests whether the mean for the live-alone students in dropout rate is the same for those who live with their parents. The p-value is small enough to reject the null hypothesis that the mean of the 2 groups is the same. It suggests a distinctive difference in the behavior of these two groups. For teenagers who don't live with their parents, it is not possible to trace their parents' records in CPS data. However, the main topic of this thesis is discussing the connection between parental income and dropout. For teenagers who move out to live alone before age of majority, the financial dynamics with their parents could have been very different from those who still live together. Therefore, leaving them out from the below estimation helps to get a clearer result. Further study could be conducted to discuss the connection in the future. Both parent is an indicator variable suggesting the student is neither from a single-family nor living alone. These are the teenagers that are mainly used in the regression estimations. Single-family has only data for one parent, while in most regressions both parents' variables are estimated at the same time. In GLM estimating, the model uses only one parent also reported for comparison. There are 27 same-sex parents families in the data set that is left after selection, considering previous studies have found impact from father and mother has a difference Chevalier et al., 2013, and the sample size is too small to make any statistically significant estimation, these subjects are also excluded.

Table 3.1: Family Type by Dropout and Age

dropout	age	step/adopt	single	live alone	both parent	SE1	SE2	SE3	SE4	N
0	16	0.05	0.29	0.05	0.67	0.01	0.01	0.00	0.01	12
1	16	0.36	0.56	0.21	0.23	0.18	0.17	0.15	0.13	2194
0	17	0.07	0.28	0.06	0.66	0.01	0.01	0.01	0.01	49
1	17	0.12	0.33	0.12	0.56	0.05	0.08	0.06	0.09	2224
0	18	0.05	0.28	0.09	0.63	0.01	0.01	0.01	0.01	279
1	18	0.05	0.27	0.21	0.51	0.01	0.03	0.03	0.03	1878

SE1 is the standard error of step/adopt, SE2 is for single and so on
Means and standard errors are weighted
N is unweighted sample size

Table 3.2: T-Test for Dropout Difference in Live Alone and Live with Parent Students

	t	df	p-value	difference in mean	CI
H0: mean difference = 0	5.0492	158	0.000001	0.099	[0.06, 0.14]

The T-test has taken survey replication weight into account.

3.2 Influence of Race on Dropout Rate

Bauman/Cranney, 2020, the most recent CPS report regarding school enrollment status, is based on American Community Survey (ACS) which surveys 3.5 million households per year and detailedly record social and economical factors about household characteristics. The report suggests that there is a long-exist gap in the dropout rate based on race. J.Causey et al., 2020 also claims there exists a racial difference in higher education enrollment status. To further examine how race affects education and earning, Table 3.3 is presented.

Table 3.3 shows that there is indeed a difference in the dropout rate. It is higher for African Americans and Hispanics, Lower for Asian or mixed-race students. The schooling year is close for both parents and among different racial groups. Except for Hispanics has distinctively shorter study time. On the other hand, the Log earning is largely affected by gender and race.

Table 3.3: Dropout Rate, Parental Education and Earning by Race

Race	dropout	LEarning(m)	Schooling(m)	LEarning(f)	Schooling(f)	N
Mean						
White	0.04	8.31	11.91	10.47	11.89	3402
Black	0.06	8.53	11.81	9.52	11.76	588
American native	0.04	8.49	11.98	9.20	11.95	84
Asian	0.01	7.27	11.65	10.26	11.61	336
Pacific islander	0.25	8.68	11.35	6.68	11.85	24
Hispanic	0.06	6.51	10.68	9.96	10.47	1341
Mixed	0.01	8.76	11.88	10.20	11.97	205
SE						
White	0.00	0.13	0.02	0.06	0.02	
Black	0.02	0.32	0.08	0.26	0.08	
American native	0.03	0.85	0.01	0.75	0.03	
Asian	0.01	0.34	0.09	0.23	0.11	
Pacific islander	0.13	1.41	0.64	2.00	0.11	
Hispanic	0.01	0.19	0.11	0.11	0.11	
Mixed	0.01	0.43	0.11	0.32	0.03	

LEarning refers to log earning, Schooling here refers to Schooling year until high school, (m) refers to mother, (f) refers to father

Means and standard errors are weighted

N is unweighted sample size

3.3 Descriptive Statistic Comparison for non-Topcoded and non-white Sample

Table 3.4 reports the means and standard error of the used variables for two different samples. To protect the privacy of high-income individuals, the CPS data set adoptes a technique called "rank proximity swapping" to topcode 26 income relative variables such as annuities amount, capital gains value, etc.. For subjects that exceed any of the 26 thresholds, their income value in the specific category would be exchange with the other subjects that have approximate values. Although the swapping technique preserves the distribution for the whole sample, it will disturb the estimation result. As the result, these samples are also excluded. If no extra specification is provided in the following results, the estimation uses non-topcode sample. The other sample presented here follows the race gap in Table 3.3, single out non-white students from the non-topcode sample.

The federal poverty level is an estimation of the income line below which a family would be considered to be poverty. It is calculated by US Census Bureau annually, the value would vary by price level in the living state, members in the family, etc. Percentage multiples of the poverty level are commonly used to decide eligibility for the welfare program. The dummy variable poverty would be 1 if the student belongs to a family that is below 150 percent of the poverty level, otherwise, it would be 0. Cutting off at 150 percent is because this is the furthest CPS data set recorded. Past research (Ermisch/Francesconi, 2001) suggests that parents in the lower-income family would not invest in their children to optimal values due to the restricted resources. Hence, income has a larger effect on the educational attainment of the children.

The original educational attainment record in CPS data is not recorded as a continuous variable. CPS only reports the highest educational attainment. For study after secondary school, CPS only reports whether the subject has accomplished the study. For study up to high school, CPS would report 9th to 12th grade separately and whether the high school diploma or equivalent is acquired. 7th and 8th, 5th and 6th, and 1th to 4th grades are grouped in the CPS data. Attempts for preserving as much information in basic parental education attainment are made, since educational attainment of parents shows significant importance on children's welfare in past researches, especially the fundamental education (Oreopoulos/Page/Stevens, 2006, Carneiro/Meghir/Parey, 2013, Govindasamy/Ramesh, 1997). The education attainment is separated into two parts, one is the schooling years until secondary school. If the subject has a grouped record, then it will be replaced by the mean of the group. For example, if a person belongs to the 7th to 8th group, then he will be coded as having 7.5 schooling years until high school. For everyone that has any record above attaining 12th grade, he will be coded as having 12 schooling years. For tertiary education, dummy indicates acquiring the degree, such as college, master, and doctoral. There is also a dummy indicate finishing high school. The dummy is made by assuming that the subject has to obtain the preliminary study before pursuing higher educational achievement. For example, if a subject is reported to have a master's degree by CPS, then he'll be assumed to also have a college and high school degree. By separating the education attainment into 2 parts, the estimation tries to capture the full effect of educational attainment, specifically when marginal utility of parental education is large.

Health Status is the self-reported health condition, using a scale of five, with 1 represents excellent health and 5 represents poor health. Health insurance coverage refers to whether the person had a health care insurance last year. Type of the disability includes having difficulty dressing or bathing, running errands such as seeing the doctor, walking or climbing stairs, focusing, or having serious difficulty in hearing or seeing. CPS only report this data for adult civilian, so the disability status of children are not included.

Earning includes incomes from the following three sources:

1. earn from this employer.

2. farm self-employment earnings.

3. own business self-employment earnings.

All the other revenues are considered as non-earning income, such as retirement income, annuities, capital gains, rents, interests are all counted in this category. For most of the above-listed income types, there exists a threshold for topcoding as mentioned previously. Those who exceed the threshold have been eliminated from the non-topcode sample.
Collar types are indicator variables constructed from the major occupation type recorded in CPS data. If the subject works in:

1. Management, business, and financial occupations.
2. Professional and related occupations.
3. Office and administrative support occupations.

Then he'll be labeled as a white-collar worker. If the subject works in:

1. Farming, fishing, and forestry occupations
2. Construction and extraction occupations.
3. Installation, maintenance, and repair occupations.
4. Production occupations.
5. Transportation and material moving occupations.
6. Armed Forces.

Then he'll be labeled as a blue-collar worker. The rest of the occupation categories are more related to customer interaction, such as entertainment, sales, or other service-oriented work which is called a pink-collar job. To avoid Collinearity, a dummy variable indicates pink-collar isn't included.

Comparing group means from non-white to the full sample, besides the decrease in both earning and non-earning income, it also profiles the common factors that are linked to poverty and it stays true for all family members. In the non-white sample, there is a higher dropout rate, fewer full-time students, lower proportion of US citizens, lower education attainment, poorer health condition, smaller proportion covered by health insurance, more blue-collar workers, and fewer white-collar workers.

Table 3.4: Variables by non-Topcoded and non-White

	non-Topcode		**non-White**	
	mean	SE	mean	SE

Female	0.4928	0.0059	0.4960	0.0093
Dropout	0.0450	0.0039	0.0505	0.0068
Full time student	0.8821	0.0061	0.8624	0.0104
Health status	1.6671	0.0157	1.7565	0.0255
Health insurance coverage	0.9429	0.0052	0.9144	0.0081
Native US citizen	0.9372	0.0050	0.8758	0.0106
Race white	0.5653	0.0076		
Race black	0.0876	0.0058	0.2016	0.0121
Race American native	0.0086	0.0019	0.0198	0.0043
Race Asian	0.0644	0.0039	0.1482	0.0085
Race Pacific islander	0.0019	0.0007	0.0045	0.0015
Race Hispanic	0.2352	0.0061	0.5412	0.0120
Race mixed	0.0367	0.0035	0.0845	0.0081
step/adopt	0.0714	0.0050	0.0683	0.0075
Poverty	0.1213	0.0064	0.1884	0.0117
	Mother			
Native US citizen	0.7362	0.0086	0.4670	0.0159
Disability	0.0328	0.0043	0.0282	0.0065
Log earning	7.8579	0.0880	7.2679	0.1285
Log non-earning income	3.6017	0.0785	2.6055	0.1047
Schooling year until high school	11.5957	0.0307	11.1816	0.0672
High school graduate	0.8986	0.0058	0.8005	0.0124
College graduate	0.5301	0.0097	0.4112	0.0144
Master graduate	0.1423	0.0071	0.1127	0.0102
Doctoral graduate	0.0157	0.0030	0.0096	0.0032
Health status	2.1439	0.0189	2.2763	0.0272
Health insurance coverage	0.8985	0.0063	0.8293	0.0116
White-collar worker	0.5003	0.0096	0.3768	0.0147
Blue-collar worker	0.0654	0.0054	0.1025	0.0099
	Father			

Native US citizen	0.7351	0.0076	0.4676	0.0147
Disability	0.0502	0.0043	0.0537	0.0070
Log earning	10.2250	0.0562	9.9067	0.0997
Log non-earning income	4.5095	0.0787	3.5220	0.1140
Schooling year until high school	11.5270	0.0301	11.0601	0.0653
High school graduate	0.8727	0.0065	0.7689	0.0136
College graduate	0.4895	0.0100	0.3783	0.0145
Master graduate	0.1508	0.0070	0.1117	0.0097
Doctoral graduate	0.0325	0.0042	0.0302	0.0069
Health status	2.1501	0.0190	2.2645	0.0279
Health insurance coverage	0.9001	0.0062	0.8312	0.0115
White-collar worker	0.4127	0.0110	0.3144	0.0144
Blue-collar worker	0.3317	0.0101	0.3998	0.0143
N	3609		1482	

Means and standard errors are weighted

N is unweighted sample size

3.4 Relation between Parental Income and Dropout

Table 3.5 presents the relationship of Log Earning of both parents by dropout and age group. For comparison purposes, schooling years until high school is also presented for both parents. For Log parent's earning, dropout groups are higher than non-dropout group except for Mother in 16-year-old group. However, as presented in Table 3.1, the dropout group of 16-years-old has only 12 people. For schooling years until high school, in 16-year-old group both parents has an average of 12 years. Again it shows the impact of the extremely small sample. For the dropout, it seems mother's secondary schooling years has a limited effect since for the 16 and 17-year-old group the schooling years are longer and it's almost equal in 18-year-old group. While fathers of the non-dropout students have longer average secondary education in both 17 and 18-year-old groups.

Nevertheless, this is only a superficial look at the 2 variable. The result after controlling all other factors such as parental education, and the result from using instrumental variables would be presented in the next chapter.

Table 3.5: Parental Income and Schooling Year until High School by Child's Age and Dropout

dropout	Age	LEarning(m)	Schooling(m)	LEarning(f)	Schooling(f)	SE1	SE2	SE3	SE4
0	16	7.77	11.54	10.26	11.54	0.14	0.06	0.09	0.05
1	16	7.88	12.00	5.21	12.00	2.23	0.00	3.03	0.00
0	17	8.03	11.63	10.24	11.56	0.15	0.05	0.09	0.05
1	17	7.05	11.82	9.24	11.35	1.60	0.17	1.11	0.39
0	18	7.84	11.61	10.26	11.51	0.15	0.05	0.09	0.06
1	18	7.32	11.60	9.81	11.23	0.53	0.12	0.31	0.20

SE1 is the standard error of Learning(m), SE2 is for Schooling(m) and so on

LEarning refers to log earning, Schooling here refers to Schooling year until high school, (m) refers to mother, (f) refers to father

Means and standard errors are weighted

N is the same as Table 3.1

3.5 Relevancy and Exogeneity of Instrumental Variables

Figure 3.1 shows log parental earning by union status. 0 means the parent is not a union member, 1 means otherwise. The dashed line is the mean of the distribution. For mothers who are not in a union, there is a cluster around zero to no income and an accumulation around 10 but the distribution is more forward for those are inside the union which suggests an overall lower pay. For fathers, the difference is not as clear as for mothers, but the distribution line for fathers who are not in the union is still more forward than those who are in, besides it also peaks earlier for those who are in the union. Accordingly, Figure 3.1 suggest that for both parents join in the union yield a higher earning. Gregg Lewis, 1986 shows that the substantial wage difference in union and non-union workers does not stem from skill difference. Chevalier et al., 2013 explore an English cohort study data from 1970 and find that union and non-union fathers differ not much in parenting attitudes and behaviors. The two studies suggest there is little possibility union status would affect children's dropout status through anyway except for earning.

Figure 3.2 shows log parental earning by branch of industry. 1 to 14 represents 14 fields of industry that is recorded by CPS data set.[2] The graph shows that mean earning is substantially different by the industry for both parents. However, the exogeneity of this instrumental variable is not as valid as the union. For some specific industries, parental behavior might be different. Denny, 2005 analyses data from Great Britain and Ireland and find that children who have a father as a third-level teacher or mother as a second-level teacher tend to have a better grade at school and often receive help from the parent on school work. But this effect is very limited to a small specific group of parents. For most of the industry how to educate children is not part of the occupational training and therefore this instrument is still presented as an IV along with the union.

Notice Figure 3.1 and Figure 3.2 do not include survey weight due to the restriction of plotting packages.

4 Results

In the first part of this chapter, dropout regress on six different sets of variables using the non-topcode sample and the marginal effect of it would be reported at first. Then a similar process would be applied to the non-white sample to see whether the earning effect on dropout rate subjugated to belonging to racial minority. Lastly, regression using the IV probit model and the test for instrumental variables would be introduced.

4.1 GLM Probit non-Topcode Sample

Table 4.1 presents the non-Topcode sample with six sets of different variables using the GLM probit model. Regression (1) using only parental earning on the child's dropout rate. In this first regression, the father's earning is significant in reducing the dropout rate. Regression (2)

[2] 1 is Agriculture, forestry, fishing, and hunting, 2 is Mining, 3 is Construction, 4 is Manufacturing, 5 is Wholesale and retail trade, 6 is Transportation and utilities, 7 is Information, 8 is Financial activities, 9 is Professional and business services, 10 is Educational and health services, 11 is Leisure and hospitality, 12 is Other services, 13 is Public administration, 14 is Armed Forces

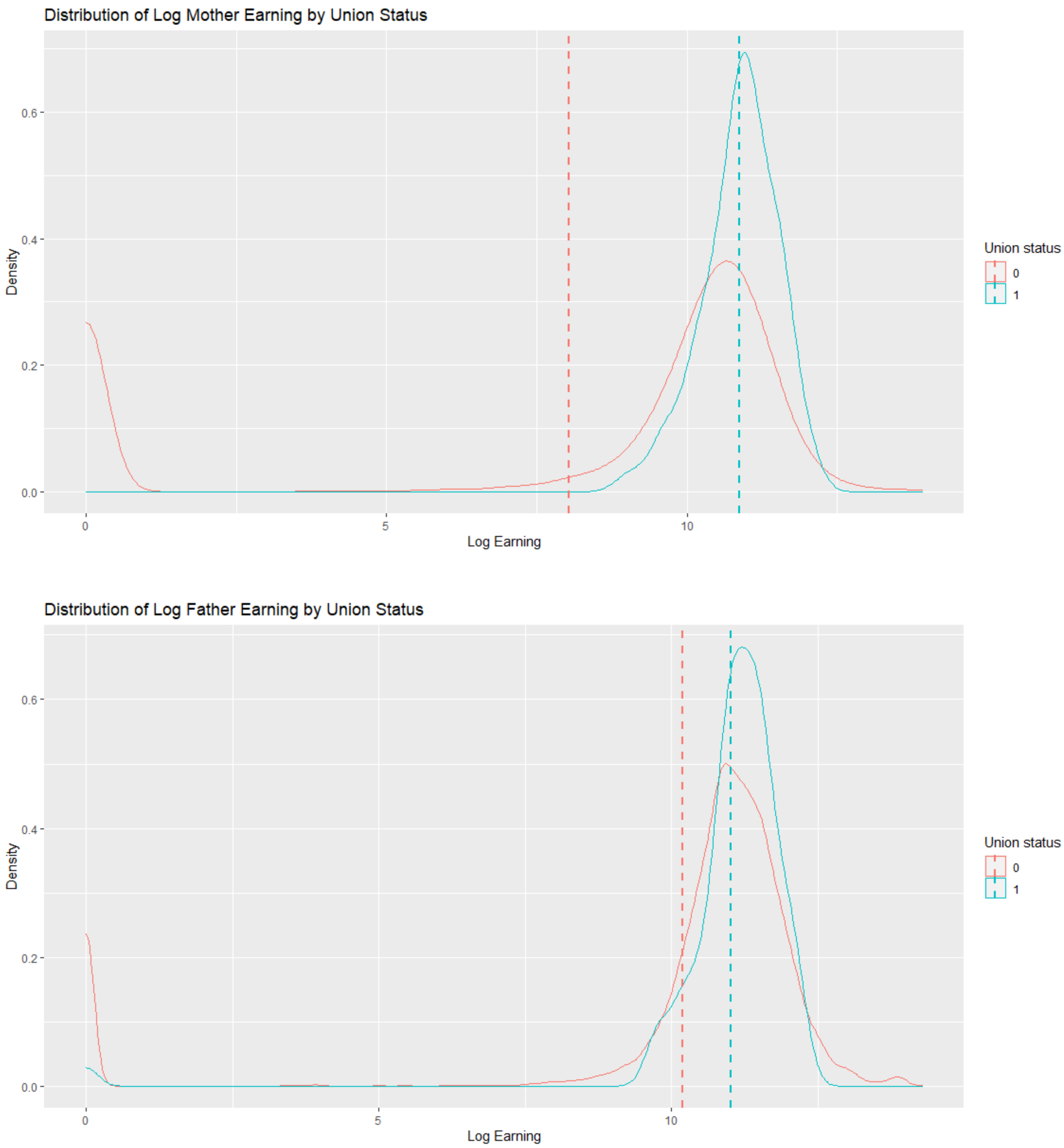

Figure 3.1: Distribution of Income by Union Status

adding in Personal traits of the student as control which includes biological gender, health condition, health insurance coverage, native US citizen dummy, and ethnicity. In this regression, the father's earning, though less significant, still reduces the dropout rate.

Regression (3) added in family type and parental trait. Family type indicates whether it's a step or an adopted family. Parental trait includes health condition, health insurance coverage, native US citizen dummy, blue or white-collar worker indicator, log non-income earning, schooling year until high school and a series of dummies regarding obtained degree. In this setting, we

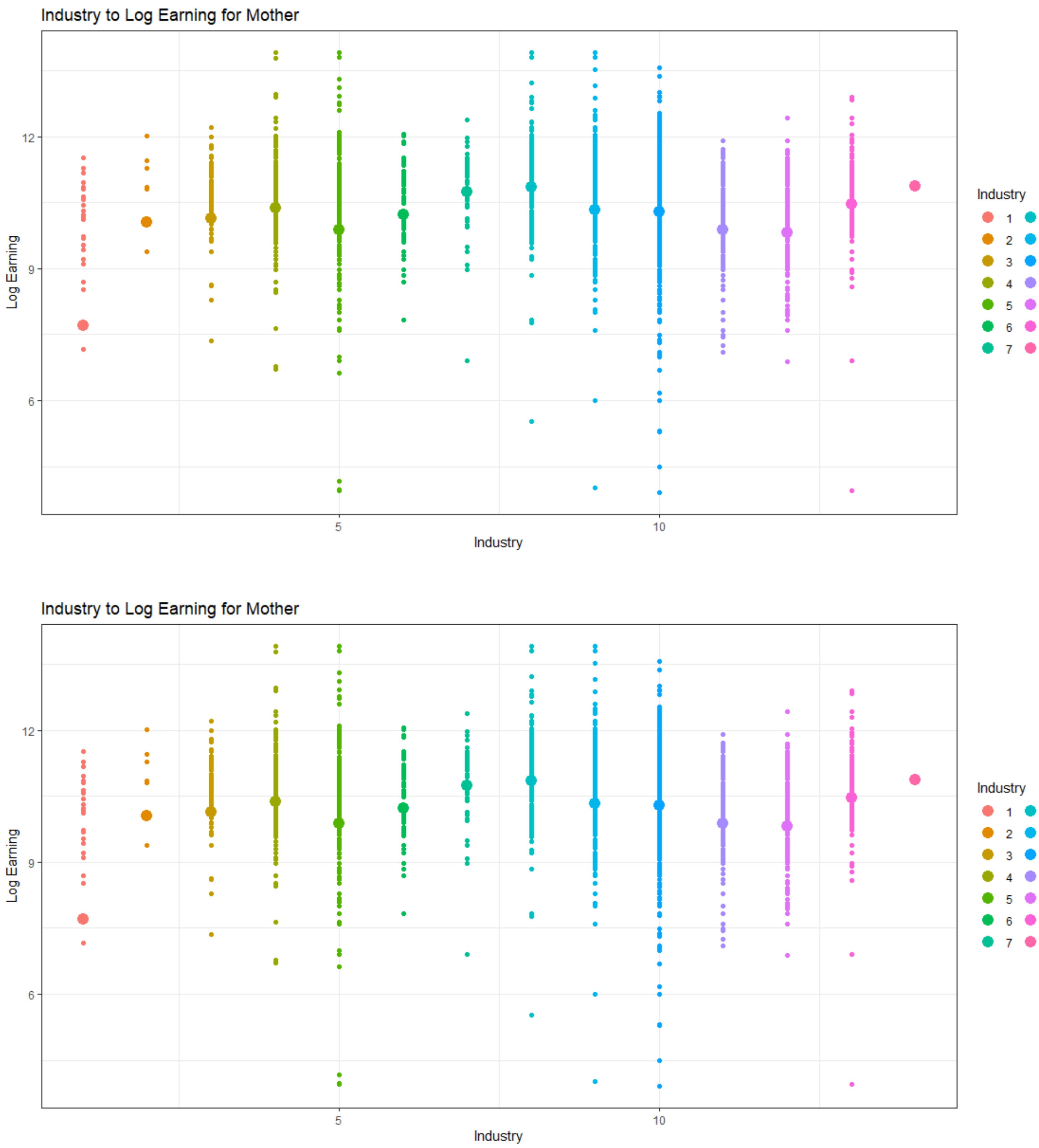

Figure 3.2: Income Mean by Different Industry

could see the effect of earning is being taken by the educational attainment. Table 4.2 reports AME using the same set of variables as regression (3). It suggests, though not significant, on average paternal earning does not affect dropout at all. The same goes for maternal earnings. Regression (4) includes states effect, it does not change the equivocal influence of parental income. Regression (4) has the lowest Log-likelihood of all which indicates it is the best-fitted model provided. Nevertheless, if the Akaike information criterion is also taken into account then regression (3) is 50 billion times as probable as the regression (4) to minimize the information loss [3] . If taken simplicity of the model into account then regression (3) is a clear win. Accordingly, regression (3) is the most repeated setting that is presented in the following section.

Regression (5) uses only maternal variables and regression (6) using only paternal ones. These two regressions are created to capture potential effects which some other studies find that only relate to one of the parents. These two regressions report a difference in sample sizes due to some of the children lack a cohabited father or mother, and these children would be excluded from regression (1) to (4) since they require information from both parents. Coefficients estimated for parental earnings are still insignificant in explaining the dropout rate. While tertiary educational attainment of both parent and job type of the mother has a prominent impact on reducing the dropout rate.

Some other salient variables that remain significant across regressions but are not shown in the tables are listed in this paragraph. Being a female reduced the rate of dropout in all regression that it is included except (5). According to AME estimated as Table 4.2, being a female reduced the probability of dropout by 0.01 percent points with a p-value below 5 percent. Reeves/Buckner/Smith, 2021 gathers disaggregated data from 2016 to 2019 across different states in the US and finds that girls are more likely to finish high school studies. Autor et al., 2020 use data in Florida and finds girls have advantages in childhood behavior so as academic outcomes. They also find boys are more likely to be subjected to low socioeconomic status than girls and deepen the gender gap at the lower tail of the socioeconomic distribution. Being Hispanic unequivocally increases the dropout rate in all regression if it is included, same as the finding in J.Causey et al., 2020. Although J.Causey et al., 2020 also sees a similar effect in African Americans, this is not the case with the result presented here. The last stably strong indicator is the health insurance coverage status in last year. For the child, AME calculated in Table 4.2 suggests that it will reduce the probability of dropout by 0.05 percent points with a p-value lower than 2 percent. Yeung, 2020 looks into this interesting link between dropout and health insurance coverage using a difference in difference method on states that adopted the Medicaid expansion and find a 0.658 percentage point reduction in dropout rates. The author argues that there might be an indirect positive impact from protecting family against financial crisis if any unexpected illness or accident happens. Therefore, families would be more willing to spend their income

[3] 50 billion times comes from exp((1431.845 - 1382.346)/2)=56049440261.5

on other sources. However, this explanation contradict with other estimation from Table 4.1 and Table 4.2. Since AME also suggests that if a father is covered by health insurance in the last year then the children's probability of dropout would increase by 0.05 percent points with a p-value below 5 percent. While the mother's health insurance status appears to be irrelevant in all regression and AME estimation. The above mentioning result is organized and put into the appendix A Table 1 and Table 2

Table 4.1: Estimation from GLM Probit Model Using nonTopcode Sample

	Dependent variable:					
	dropout					
	(1)	(2)	(3)	(4)	(5)	(6)
Poverty			−0.139	−0.094	0.028	−0.180
			(0.147)	(0.151)	(0.090)	(0.140)
Mother						
Log earning	−0.011	−0.008	0.005	0.005	0.001	
	(0.010)	(0.010)	(0.012)	(0.013)	(0.010)	
Log non-earning income			−0.014	−0.012	−0.016*	
			(0.012)	(0.012)	(0.009)	
White-collar			−0.253**	−0.264**	−0.239**	
			(0.122)	(0.121)	(0.097)	
Blue-collar			−0.145	−0.164	−0.154	
			(0.180)	(0.181)	(0.147)	
High school graduate			0.212	0.206	0.189	
			(0.240)	(0.244)	(0.158)	
College graduate			−0.118	−0.119	−0.216**	
			(0.113)	(0.111)	(0.088)	

Master graduate			-0.004	0.004	-0.009	
			(0.152)	(0.154)	(0.139)	
Doctoral graduate			-3.531^{**}	-3.720^{**}	-3.611^{**}	
			(1.684)	(1.750)	(1.404)	
Father						
Log earning	-0.026^{**}	-0.021^{*}	-0.020	-0.023		-0.017
	(0.012)	(0.013)	(0.017)	(0.017)		(0.016)
Log non-earning income			0.001	0.002		-0.002
			(0.014)	(0.014)		(0.013)
White-collar			0.032	0.013		0.021
			(0.123)	(0.126)		(0.113)
Blue-collar			0.048	0.069		0.073
			(0.104)	(0.112)		(0.103)
High school graduate			-0.144	-0.099		-0.164
			(0.241)	(0.224)		(0.193)
College graduate			-0.157	-0.161		-0.236^{**}
			(0.123)	(0.126)		(0.114)
Master graduate			-0.401^{**}	-0.439^{***}		-0.373^{**}
			(0.158)	(0.155)		(0.166)
Doctoral graduate			0.258	0.273		0.245
			(0.339)	(0.318)		(0.330)
Control for						
Personnel Trait	N	Y	Y	Y	Y	Y
Family type	N	N	Y	Y	Y	Y
Parental Trait	N	N	Y	Y	Y	Y
State	N	N	N	Y	N	N
Observations	4,103	4,103	4,103	4,103	5,806	4,626
Log Likelihood	-703.842	-298.054	-272.275	-258.981	-413.264	-319.210

Akaike Inf. Crit.	1,413.684	624.109	624.550	697.962	892.528	704.420

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

Estimation are weighted by survey replication weights

N is unweighted sample size

Table 4.2: AME for Regression (3) Using nonTopcode Sample

factor	AME	SE	z	p	lower	upper
Poverty	-0.01	0.01	-0.88	0.38	-0.04	0.01
Mother						
Log earning	0.00	0.00	0.45	0.66	-0.00	0.00
Log non-earning income	-0.00	0.00	-1.25	0.21	-0.00	0.00
White-collar	-0.02	0.01	-2.01	0.04**	-0.04	-0.00
Blue-collar	-0.01	0.02	-0.74	0.46	-0.04	0.02
High school graduate	0.02	0.02	0.74	0.46	-0.03	0.06
College graduate	-0.01	0.01	-1.29	0.20	-0.03	0.01
Master graduate	-0.00	0.01	-0.11	0.92	-0.03	0.03
Doctoral graduate	-0.32	0.14	-2.35	0.02**	-0.59	-0.05
Father						
Log earning	-0.00	0.00	-1.18	0.24	-0.00	0.00
Log non-earning income	-0.00	0.00	-0.02	0.99	-0.00	0.00
White-collar	0.00	0.01	0.03	0.98	-0.02	0.02
Blue-collar	0.00	0.01	0.32	0.75	-0.02	0.02
High school graduate	-0.01	0.02	-0.58	0.56	-0.06	0.03
College graduate	-0.01	0.01	-1.33	0.18	-0.04	0.01
Master graduate	-0.04	0.01	-2.49	0.01***	-0.07	-0.01
Doctoral graduate	0.02	0.03	0.75	0.45	-0.04	0.08

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

Estimation are weighted by survey replication weights

This paper follows Chevalier et al., 2013, using earning instead of the full income as the main regressor in most of the estimation. Choosing earning over income also ensures the direct link between the endogenous variable and instrumental variables. Besides, for majority of people, earning is the main source of income. In Table 4.1, log non-income earning is also a controlled variable to warrant the effect of non-income earning is also taken into accounts. However, when a family is planning financial spending, all the income regardless of source might be considered as a whole. Instead of looking at the earning and non-earning as 2 different variables, Table 4.3 using income as a whole. The regression setting is the same as Table 4.1, the other variables show highly similar estimation results, hence, they are skipped in Table 4.3. For the mother's income, it does not seem to matter starting from regression (1). On the other hand, for fathers, in regression (1) where there is only log income from both parents, it seems to have a slight effect in reducing dropout probability. But once other variables are taken into account, the significance disappears right away. In Table 4.4, AME uses income as regressor for regression model (3) in Table 4.3. The average marginal effect is 0 for both parents. In conclusion, the result implies the 2 different ways to frame income as regressor shows an identical result. In the later estimations, income would be split by its source and the focus would be on the earning.

Table 4.3: Estimation from GLM Probit Model Using Full Income as Regressor

	Dependent variable:					
	dropout					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Income Mother	−0.012	−0.007	0.001	0.004	0.005	
	(0.011)	(0.012)	(0.014)	(0.014)	(0.011)	
Log Income Father	−0.036**	−0.025	−0.020	−0.020		−0.014
	(0.017)	(0.020)	(0.023)	(0.023)		(0.022)
Control for						
Personnel Trait	N	Y	Y	Y	Y	Y
Family type	N	N	Y	Y	Y	Y
Parental Trait	N	N	Y	Y	Y	Y
State	N	N	N	Y	N	N
Observations	4,103	4,103	4,103	4,103	5,806	4,626
Log Likelihood	−703.842	−298.054	−272.275	−258.981	−413.264	−319.210
Akaike Inf. Crit.	1,413.684	624.109	624.550	697.962	892.528	704.420

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

Estimation are weighted by survey replication weights

N is unweighted sample size

Table 4.4: AME for Regression (3) Using Full Income as Regressor

factor	AME	SE	z	p	lower	upper
Log income mother	0.00	0.00	0.16	0.87	-0.00	0.00
Log income father	-0.00	0.00	-0.86	0.39	-0.01	0.00

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Estimation are weighted by survey replication weights

4.2 GLM Probit non-White Sample

Table 4.5 repeats the regression (1) to (3) in the previous chapter to see whether in racial minorities earnings of parents have a different impact on children's dropout choice. Including state has been overly complicated and might cause information loss, especially since the sample size is much smaller in this chapter. Therefore Regression (4) is excluded. Regression (5) and (6) are also excluded since they are not good fit judging by the low log-likelihood score in Table 4.1. Similar to what has been shown in the non-Topcode sample, only paternal earning in regression that controls for little factors suggests it is correlated with dropout decision. Once including more control variables, the effect is captured by the higher education factors. The limitation to racial minority does not give much difference in estimating parental income's effect, only the non-topcode sample would be used in the IV section.

Some interesting observed differences from this estimation is that for the racial minority group, paternal job type and identity are more important for racial minority students. Based on AME estimate, if the student's father is a blue-collar worker instead of the pink-collar worker then the student's probability of dropping out would decrease by 0.03 percent points. But if the father is a native US citizen, then the probability of dropping out would increase by 0.03 percent points. Health condition for racial minority students is also crucial. While for the whole sample health condition is insignificant, for racial minority students if health worsens by one scale, the dropout probability would increase by 0.02 percent points. The inexplicable synergy between health insurance coverage still proceeds, though the effect is milder. If a student from the racial minority group is covered by the health insurance in last year, then the chance of dropping out would decrease by 0.04 percent points, while if his father is covered by the health insurance the chance of dropout would increase by 0.03 percent points.

Table 4.5: Estimation from GLM Probit Model Using nonWhite Sample

	Dependent variable:		
	dropout		
	(1)	(2)	(3)
Poverty			-0.057
			(0.200)
Mother			
Log earning	0.003	0.005	0.008
	(0.014)	(0.015)	(0.018)
Log non-earning income			-0.007
			(0.020)
White-collar			-0.165
			(0.186)
Blue-collar			-0.108
			(0.206)
High school graduate			0.340
			(0.291)
College graduate			-0.334^{*}
			(0.174)
Master graduate			-0.036
			(0.277)
Doctoral graduate			-3.697^{**}
			(1.791)
Father			
Log earning	-0.027^{*}	-0.023	-0.016
	(0.016)	(0.017)	(0.023)

Log non-earning income			0.001
			(0.021)
White-collar			−0.106
			(0.190)
Blue-collar			−0.276*
			(0.161)
High school graduate			0.071
			(0.227)
College graduate			0.046
			(0.179)
Master graduate			−0.456
			(0.288)
Doctoral graduate			−3.633*
			(1.844)
Control for			
Personnel Trait	N	Y	Y
Family type	N	N	Y
Parental Trait	N	N	Y
State	N	N	N
Observations	1,656	1,656	1,656
Log Likelihood	−318.165	−134.890	−106.083
Akaike Inf. Crit.	642.331	295.779	290.165

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

Estimation are weighted by survey replication weights

N is unweighted sample size

Table 4.6: AME for Regression (3) Using nonWhite Sample

factor	AME	SE	z	p	lower	upper
Poverty	-0.00	0.02	-0.15	0.88	-0.04	0.03
Mother						
Log earning	0.00	0.00	0.59	0.56	-0.00	0.00
Log non-earning income	-0.00	0.00	-0.42	0.67	-0.00	0.00
Blue-collar	-0.01	0.02	-0.54	0.59	-0.05	0.03
White-collar	-0.02	0.02	-0.86	0.39	-0.05	0.02
High school graduate	0.03	0.03	1.08	0.28	-0.02	0.08
College graduate	-0.04	0.02	-2.39	0.02**	-0.07	-0.01
Master graduate	-0.01	0.03	-0.21	0.83	-0.06	0.05
Doctoral graduate	-0.35	0.16	-2.21	0.03**	-0.65	-0.04
Father						
Log earning	-0.00	0.00	-0.74	0.46	-0.01	0.00
Log non-earning income	-0.00	0.00	-0.14	0.89	-0.00	0.00
Blue-collar	-0.03	0.02	-1.73	0.08*	-0.06	0.00
White-collar	-0.01	0.02	-0.84	0.40	-0.05	0.02
High school graduate	0.01	0.02	0.38	0.70	-0.03	0.05
College graduate	0.00	0.02	0.12	0.91	-0.03	0.04
Master graduate	-0.05	0.03	-1.63	0.10*	-0.10	0.01
Doctoral graduate	-0.35	0.16	-2.19	0.03**	-0.65	-0.04

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

Estimation are weighted by survey replication weights

4.3 Result from IV Probit Estimation

Table 4.7 shows the result of using branches of industry and union status as IV to estimate the effect of parental earnings. The estimation is done separately for both parents. For estimating the mother's earning, mother's union status and industrial branch are used as IV while holding all the other paternal characteristics as controlled variables, estimation for father's earning is done in a likewise manner. After controlling for endogeneity, the estimates still suggest for both parents the impact on children's dropout decision is close to 0 which evince both parent's earnings are not related to children's dropout decision. The estimations suggest a similar result using the GLM model. Tertiary education is a more prominent indicator than money that is earned. Biological gender and race of the student, health coverage of student, and health coverage of father has the same effect as explained in section 4.1

If the instruments used have a weak link to the endogenous variable, it will cause the convergence speeds of the estimated coefficient's denominator and nominator to be different (Hansen,

Table 4.7: Instrumented Regression

	theta	SE	t_value	p_value
Instrumented Mother's Earning				
Log earning	-0.00	0.00	-0.01	0.99
Log non-earning income	-0.00	0.00	-0.97	0.33
White-collar	-0.02	0.01	-1.61	0.11
Blue-collar	-0.01	0.02	-0.62	0.54
High school graduate	0.03	0.03	0.92	0.36
College graduate	-0.01	0.01	-0.93	0.35
Master graduate	0.00	0.01	0.02	0.98
Doctoral graduate	-0.02	0.01	-2.01	0.04**
Instrumented Father's Earning				
Log earning	-0.00	0.00	-0.89	0.37
Log non-earning income	0.00	0.00	0.28	0.78
White-collarr	0.00	0.01	0.34	0.73
Blue-collar	0.00	0.01	0.20	0.84
High school graduate	-0.02	0.03	-0.59	0.56
College graduate	-0.02	0.01	-1.47	0.14
Master graduate	-0.02	0.01	-2.47	0.01***
Doctoral graduate	0.01	0.02	0.90	0.37
Note:	*p<0.1; **p<0.05; ***p<0.01			
Estimation are weighted by survey replication weights				

2021). The estimated coefficient will become inconsistent if a weak instrument is used. Therefore, it's crucial to conduct checks to ensure that the validness of the IV. Table 4.8 shows the test statistic for the mother's union status and mother's branch of industry, Table 4.9 shows father's. Both tests do not account for the replication weights of the survey.

Based on Colonescu, 2018 the specification test in statistical software R could be explained as follow. The null hypothesis of the weak instrument test is "All instruments are weak". In both tables, p-values are small enough to reject the null hypothesis at 0.01 level which suggests at least one instrument is strong. The parental earnings are estimated separately, in each regression, there is only one endogenous variable. As a result, it ensures the reliability of the weak instrument test. The null hypothesis of Wu-Hausman is that the covariance between earning and error term in GLM specification is zero. Do not reject null hypothesis means estimated coefficients from GLM and IV are equally consistent. In both tables, p-values are unable to reject the null hypothesis at all. This suggests the endogenous problem does not exist and both GLM and IV could yield a consistent result. Under the premise that both are consistent estimators, using

GLM would be much more efficient and preferable. Sargan test is the overidentifying test, it could only be performed when instrumented variables outnumber endogenous variables. The null hypothesis of Sargan is that the extra instrument is uncorrelated with the error term which ensures the exogeneity condition of the instrument. In Table 4.8 the null hypothesis is rejected. This means that the instruments which are used for mothers' earnings are valid. However, Sargon test for fathers in Table 4.9 is significant at a level that is close to 0.1. This suggests a mild correlation between instrument and error term. As the research Denny, 2005 quoted in chapter 3 suggests, if parental industry relates to children's education, it can cause a direct impact on children's educational performance. The result only speaks to the extra instrument, therefore, it is not possible to know whether joining a union is also an invalid estimator for dropout behavior. The wald statistic suggests the essentialness of the variables being tested. The null hypothesis is the subject coefficients are simultaneously equal to zero. In both Table 4.8 and Table 4.9, p-values are small enough to reject the null hypothesis at 0.01 which means excluding the variables from the estimation would not substantially harm the fit of the model. In conclusion, the test result indicates the estimation from the previous GLM chapter is consistent and trustworthy.

Table 4.8: Diagnostic tests for Instrument : Mother's Earning

	df1	df2	statistic	p-value
Weak instruments	14	4051	431.910	<2e-16***
Wu-Hausman	1	4063	0.079	0.778
Sargan	13	NA	11.718	0.551

Note: *p<0.1; **p<0.05; ***p<0.01
Residual standard error: 0.2045 on 4064 degrees of freedom
Multiple R-Squared: 0.02247, Adjusted R-squared: 0.01333
Wald test:2.453 on 38 and 4064 DF, p-value: 1.888e-06

5 Conclusion

This paper explored whether the parental earning is still generally independent of early school leaving under corona impact. The estimation that has been done on the full sample and racial

Table 4.9: Diagnostic tests for Instrument : Father's Earning

	df1	df2	statistic	p-value
Weak instruments	15	4050	231.91	<2e-16***
Wu-Hausman	1	4063	0.53	0.4665
Sargan	14	NA	21.40	0.0919 *

Note: *p<0.1; **p<0.05; ***p<0.01
Residual standard error: 0.2046 on 4064 degrees of freedom
Multiple R-Squared: 0.02233, Adjusted R-squared: 0.01319
Wald test:2.458 on 38 and 4064 DF, p-value: 1.773e-06

minority sample both suggests the independence remains true after the global pandemic takes effect. Regressing the dropout on full income instead of earning manifests a similar irrelevancy. Instrumented parental earning using union status and industry also suggests parental earning does not affect dropout decisions between 16 to 18 years old children. The weak instrument test of IV regression suggests that at least one of the instruments used is strong. The overidentifying test suggests that the extra instrument does not satisfy exogeneity restriction. However, the Wu-Hausman test suggests that there is no endogenous problem of parental earning, the result from GLM estimation is more efficient than IV regression and is already consistent.

Several restrictions of this paper are stemming from the data used. For example, can not access data for students who live alone, topcode mechanism in CPS data set, limited disclosure of parental educational years, and only shows educational attainment for students above 15 years old. Despite different variable coping methods is adopted, a full disclosure of data would undeniably yield a clearer picture of the true population.

One thing worthy of notice is that the spread of Covid-19 also affects the data collecting process. According to the supplement from U.S. Census Bureau, 2020, all in-person interview were suspended after March 15, 2020, and replaced by phone interviewe. In addition, there is a 10 percentage points drop in response rate to 73 percent comparing to the previous month or same month last year. The impact of Covid-19 might make the year 2020 incomparable to previous years. If the property of the 10 percent people who miss from the data this year have an underlying relationship with the dropping out decision, then it might cause a bias in the estimation. However, the reasons for decreasing the responding rate aren't explicitly listed in the supplement, it would be assertive to assume not participating is correlated with dropping out.

Further research could be done in the future based on this paper including research the link on a more detailed and unrestricted data set. It is also an option instead of dropout possibility questing whether parental wealth affects the student educational performance. Considering most of the schools resort to online teaching during the lockdown, being capable of affording electronic devices, having stable internet connection, suitable study environment at home, and supporting study resources would be critical to facilitate study and all of these factors are based on parental financial status. Most of the past researches suggest a causal link between parental education and student educational performance and dropout choice, exploring whether the link intensifies by the study at home restriction is another important research question. Understanding the link would be very helpful to relevant policy-making and minimizing the impact on students who have a financially curbed family background. Furthermore, as time goes by, the policies of the government and the reaction of parents and students also shift. The difference of parental incomes' impact on dropout decisions and the most recent one to see how their behavior migrate is also worthwhile of discovering. Lastly, in the paper, a mysterious synergy between health insurance of the student, health insurance of father, and the dropout rate is discovered and it subverts the assumed cause brought up in Yeung, 2020 that there might be an indirect positive impact from protecting family against financial crisis. Finding a plausible explanation for these dynamics is another direction for future studies.

In this paper, it is shown that at the beginning of the pandemic, the common finding that parental earning does not affect high school dropout decision remains the same as pre-pandemic time under several different angles of examination .

References

Autor, D./D. N. Figlio, et al. (2020): *Males at the Tails: How Socioeconomic Status Shapes the Gender Gap*. Working Paper 27196. National Bureau of Economic Research. DOI: 10.3386/w27196. URL: http://www.nber.org/papers/w27196.

Bauman, K./S. Cranney (2020): "School Enrollment in the United States: 2018". In: URL: https://www.census.gov/content/dam/Census/library/publications/2020/demo/p20-584.pdf.

Blau, D. (1999): "The Effect Of Income On Child Development". In: *The Review of Economics and Statistics* 81.2, pp. 261–276. URL: https://EconPapers.repec.org/RePEc:tpr:restat:v:81:y:1999:i:2:p:261-276.

Carneiro, P./C. Meghir/M. Parey (Jan. 2013): "Maternal Education, Home Environments, and the Development of Children and Adolescents". In: *Journal of the European Economic Association* 11.suppl$_1$, pp. 123–160. ISSN: 1542-4766. DOI: 10.1111/j.1542-4774.2012.01096.x. eprint: https://academic.oup.com/jeea/article-pdf/11/suppl_1/123/10315047/jeea0123.pdf. URL: https://doi.org/10.1111/j.1542-4774.2012.01096.x.

Chevalier, A./C. Harmon, et al. (2013): "The impact of parental income and education on the schooling of their children". In: *IZA Journal of Labor Economics* 2. ISSN: 2193-8997. URL: https://doi.org/10.1186/2193-8997-2-8.

Colonescu, C. (2018): *Using R for Principles of Econometrics*. lulu.com. ISBN: 1387473611.

Denny, K. (Apr. 2005): "Do teachers make better parents? -the differential performance of teachers' children at school". In: *Regional and Sectoral Economic Studies*.

Ermisch, J./M. Francesconi (2001): "Family structure and children's achievements". In: *Journal of Population Economics* 14.2, pp. 249–270. URL: https://EconPapers.repec.org/RePEc:spr:jopoec:v:14:y:2001:i:2:p:249-270.

Garrett M. Fitzmaurice Nan M. Laird, J. H. W. (2001): *Applied Longitudinal Analysis, Second Edition*. John Wiley & Sons, Inc. ISBN: 9781119513469. DOI: 10.1002/9781119513469.

Govindasamy, P./B. Ramesh (Jan. 1997): "Maternal Education and Utilization of Maternal and Child Health Services in India". In: *National Family Health Survey Subject Reports*.

Gregg Lewis, H. (1986): "Chapter 20 Union relative wage effects". In: vol. 2. Handbook of Labor Economics. Elsevier, pp. 1139–1181. DOI: https://doi.org/10.1016/S1573-4463(86)02010-2. URL: https://www.sciencedirect.com/science/article/pii/S1573446386020102.

Hansen, B. E. (2021): *Econometrics*. University of Wisconsin, Department of Economics.

Hollingsworth, H. (May 11, 2021): "US Schools Fight to Keep Students Amid Fear of Dropout Surge". In: *AP Morning Wire*. URL: https://apnews.com/article/coronavirus-pandemic-health-education-1b0eb720342a1978383c703e5a001a90 (visited on 05/11/2021).

J.Causey/Q.Liu, et al. (June 2020): "A COVID-19 Supplement to Spring 2020 Current Term Enrollment Estimates". In: URL: https://nscresearchcenter.org/current-term-enrollment-estimates/.

Jenkins, S./C. Schluter (2002): *The Effect of Family Income During Childhood on Later-Life Attainment: Evidence from Germany*. IZA Discussion Papers 604. Institute of Labor Economics (IZA). URL: https://EconPapers.repec.org/RePEc:iza:izadps:dp604.

Kakuchi, S. (Mar. 10, 2021): "Student dropout rate on the rise due to pandemic impact". In: *University World News*. URL: https://www.asahi.com/ajw/articles/13988451 (visited on 03/10/2021).

Mayer, S. E. (1998): *What Money Can't Buy Family Income and Children's Life Chances*. Harvard University Press. ISBN: 978-0674587342.

Mikulecky, M. (Apr. 2013): "Compulsory School Age Requirements". In: *Education Commission of the States*. URL: https://www.ecs.org/clearinghouse/01/07/03/10703.pdf.

Oreopoulos, P. (2003): *Do Dropouts Drop Out Too Soon? International Evidence From Changes in School-Leaving Laws*. Working Paper 10155. National Bureau of Economic Research. DOI: 10.3386/w10155. URL: http://www.nber.org/papers/w10155.

Oreopoulos, P./M. E. Page/A. H. Stevens (2006): "The Intergenerational Effects of Compulsory Schooling". In: *Journal of Labor Economics* 24.4, pp. 729–760. ISSN: 0734306X, 15375307. URL: http://www.jstor.org/stable/10.1086/506484.

Perraillon, M. C. (2019): *Interpreting Model Estimates: Marginal Effects*. URL: https://clas.ucdenver.edu/marcelo-perraillon/code-and-topics/marginal-effects.

Reeves, R. V./E. Buckner/E. Smith (Jan. 2021): “The unreported gender gap in high school graduation rates”. In: URL: https://www.brookings.edu/blog/up-front/2021/01/12/the-unreported-gender-gap-in-high-school-graduation-rates/.

Shea, J. (2000): “Does parents’ money matter?” In: *Journal of Public Economics* 77.2, pp. 155–184. URL: https://EconPapers.repec.org/RePEc:eee:pubeco:v:77:y:2000:i:2:p:155-184.

“UNESCO Press Release No.2020-73” (2020). In: *Press release UN Secretary-General warns of education catastrophe, pointing to UNESCO estimate of 24 million learners at risk of dropping out*. URL: https://en.unesco.org/news/secretary-general-warns-education-catastrophe-pointing-unesco-estimate-24-million-learners-risk.

U.S. Census Bureau (2020): “Current Population Survey 2020 Annual Social and Economic (ASEC) Supplement”. In: URL: https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar20.pdf.

Vovk Viktoria, M. A. (2020): “Comparative Analysis on the Impact of Distance Learning Between Russian and Japanese University Students, During the Pandemic of COVID-19”. In: *Education Quarterly Reviews* 3.4, pp. 438–446. ISSN: 2621-5799.

Yeung, R. (2020): “The Effect of the Medicaid Expansion on Dropout Rates”. In: 745–753. URL: https://doi.org/10.1111/josh.12937.

Appendix

Appendix A - Other Salient Variables in GLM Probit Model using non-Topcode Sample

Table 1: Estimation from GLM Probit Model using nonTopcode Sample

	Dependent variable:					
	dropout					
	(1)	(2)	(3)	(4)	(5)	(6)
Female		-0.181^{**}	-0.165^{**}	-0.178^{**}	-0.065	-0.183^{**}
		(0.076)	(0.080)	(0.080)	(0.066)	(0.076)
Race Hispanic		0.927^{**}	1.035^{**}	0.798^{*}	0.659^{*}	1.025^{**}
		(0.453)	(0.436)	(0.438)	(0.397)	(0.459)
Health insurance coverage		-0.250^{*}	-0.540^{**}	-0.562^{**}	-0.324^{*}	-0.539^{***}
		(0.143)	(0.222)	(0.221)	(0.169)	(0.161)
Health insurance coverage(m)			-0.092	-0.130	0.139	
			(0.286)	(0.289)	(0.147)	
Health insurance coverage(f)			0.600^{**}	0.634^{**}		0.404^{**}
			(0.273)	(0.286)		(0.174)
Control for						
Personnel Trait	N	Y	Y	Y	Y	Y

Family type	N	N	Y	Y	Y	Y
Parental Trait	N	N	Y	Y	Y	Y
State	N	N	N	Y	N	N
Observations	4,103	4,103	4,103	4,103	5,806	4,626
Log Likelihood	−703.842	−298.054	−272.275	−258.981	−413.264	−319.210
Akaike Inf. Crit.	1,413.684	624.109	624.550	697.962	892.528	704.420

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

Estimation are weighted by survey replication weights

N is unweighted sample size

Table 2: AME for Regression (3) using nonTopcode Sample

factor	AME	SE	z	p	lower	upper
Female	-0.01	0.01	-2.03	0.04**	-0.03	-0.00
Health insurance coverage	-0.05	0.02	-2.39	0.02**	-0.09	-0.01
Health insurance coverage(m)	-0.01	0.03	-0.28	0.78	-0.06	0.04
Health insurance coverage(f)	0.05	0.02	2.03	0.04**	0.00	0.10

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

Estimation are weighted by survey replication weights

I hereby confirm that the work presented has been performed and interpreted solely by myself except for where I explicitly identified the contrary. I assure that this work has not been presented in any other form for the fulfillment of any other degree or qualification. Ideas taken from other works in letter and in spirit are identified in every single case.

signature: