DO POOR FAMILIES DROP OUT OF SCHOOL BECAUSE OF LOW INCOME?

Junzhao Shi[1]
BSc Economics
3rd year
University College London

Explore Econ Undergraduate Research Conference
February 2020

[1] Special thanks to Professor Aureo de Paula, Dr Daniel Wilhelm, Mr Michele Giannola, Dr Gabriella Conti, Dr Ramin Nassehi, Dr Dunli Li, Mr Edwin Mu and Mr Gehui Jin for support and guidance.

Introduction:

Income inequality in the US have been rising for the last few decades, with the Gini coefficient increasing from 34.6 in 1979 to 41.5 in 2016 (World Bank estimates, 2019). There are a lot of discussions centred on how income inequality changes various aspects of the economy, such as consumption and expenditure behaviour and social mobility. Zooming onto the aspect of social mobility, education, being one of the gateways for kids from low-income families to climb up the social ladder, is typically allocated with high association with income inequality. In the case of rich families in the US, parents would send their kids to expensive private schools, where education resources are generally better than public schools, in terms of teacher-student ratio, study equipment and students' motivation for further studies (Figlio and Stone, 2000). Children from deprived backgrounds tend to have less opportunities to get into elite colleges, some of them are not even able to finish their secondary education. The factors that lead to students from low-income backgrounds dropping out of schools and colleges vary. Stinebrickner and Stinebrickner (2003) found that the primary reasons are beyond the direct costs of college but focused more on family environment. Different to colleges, secondary education generally do not incur as high cost as colleges. Foley, Gallipoli and Green (2014) found that cognitive ability and parental valuation of education plays a large role in high school drop-out decision. Contrary to many discussions, which placed emphasis on permanent family characteristics, this paper intends to analyse the effect of family income on the decision of school dropout, providing evidence on the contrary for public policies such as negative income tax scheme to support and motivate families and students to finish their education.

Previous Research:

Chevalier *et al.* (2013) used instrumental variables to account for endogeneity of paternal and maternal education and parental income. Akin to early research done by Harmon and Walker (1995), they first utilised the change in minimum school leaving age in the UK in the 1970s to attribute for the exogenous change in years of education. The second source of exogeneity is found similar to Angrist and Krueger (1991). Since the fact that children born in the months early in the school year enter school at a slightly older age than their peers, when they decide to dropout by the time they turn 16, they have less education than their younger counterparts, inducing the exogenous change in education. The third instrument is union status of parents, which was constructed by Shea (2000). Using the Labour Force Survey (LFS) and pooling households from 1993-2012, which corresponds to the families in which the parents are affected the school leaving age change, they found that paternal education affects daughters' further education decision the most and that maternal education plays a minor role. The effect of parental income even when instrumented by various instruments are shown to be not robust, leading to the conclusion that policies aiming to help children pursue further education by increasing family income may not be effective.

Some, on the other hand, have tried to use quasi-experimental approach. Akee *et al.* (2010) implemented a quasi-experimental approach using an obliged permanent transfer of profit from a newly opened casino on the Eastern Cherokee reservation in North Carolina to households with pre-existing American Indian status in 1997. By sampling both American Indian and non-Indian and showing that the two groups had similar income growth rate before 1997, they utilised difference-in-difference analysis to estimate the effect of exogenous change in income on educational attainment, high school completion and criminal records. The empirical results show that an additional $4,000 increase in income per year would allow kids on average to gain one more year of schooling. Nonetheless, it is subject to scrutiny that the results found are applicable to the whole American population, since only American Indians were induced with the exogenous increase in income. Factors such as heterogeneity in propensity spending on education among different ethnicities would create bias in the effect of income on education for the whole population.

It is also important to note some research relied on the change in public policies. Dahl & Lochner (2012) utilised the expansion of Earned Income Tax Credit (EITC) in the US in the late 1980s and 1990s as a source of exogenous variation in family income for low- and middle-income households. Using the varied amount of financial support families gained from the EITC as an instrumental variable for family income throughout the period, they evaluated the effect of income on children's educational achievement. It is found that a $1,000 increase in income raises test scores by 6 percent of a standard deviation in the short run. It should also be noted that the gains for low-income families are higher and more robust than the rest of the population. Løken (2010) applied a similar approach, using the family income during the Norwegian oil boom in the 1970s as a source of exogenous variation, instrumenting on family permanent income. As some regions were affected by the burgeoning oil industry much more than others, the control group is relied on observations from the neighbouring county during the same period. The IV result shows that there is no significant causal relationship between parental income and education, despite the change in income is regarded as permanent. However, it can be argued that families might move to the oil-boom affected region, seeking for higher income, causing selection on treatment.

Data and Empirical strategy:

The data that this paper relies on are from the Panel Study of Income Dynamics (PSID) in the United States, where the survey began in 1968, recorded a nationally representative sample of about 5,000 families residing in the United States. It covers the early temporary implementation of the Earned Income Tax Credit (EITC) when it was introduced in 1975. The survey data appear with one-year lag, meaning that the implementation in 1975 was recorded in the 1976 wave data. In this paper, I would keep it consistent with waves of survey, which means that the initial introduction was in year 1976 data. As the data are recorded on individual level, one would first identify and group individuals in the same family together to form family level data. I first define low-income households as those with income at the lowest quintile in 1975, for which the upper limit is $5,000 (Brookings Institution, 2019). These families are traced throughout the period from 1970 to 1978, where three periods of EITC implementations are recorded. Some families might have moved up the social ladder towards later period. However, this shouldn't be a worry, as the interest of this paper is to find out, with certain degree of social mobility, how income affects low-income households' educational decisions. As the EITC only allows families with dependent children, i.e. sons and daughters, step-children and further descendent, the sample is selected with families with pre-tax income less than $5,000 in 1975 and are eligible for the EITC. This gives a sample of 765 households.

The amount of EITC given by the government is strictly dependent on the amount of pre-tax earnings a family had in a given year. For families with earnings below $4,000, the EITC programme would sponsor 10% of the income, whereas if family earning before tax exceeds

$4,000, they would not be eligible for EITC. The simple relationship between the amount of EITC and pre-tax income is shown in figure 1.

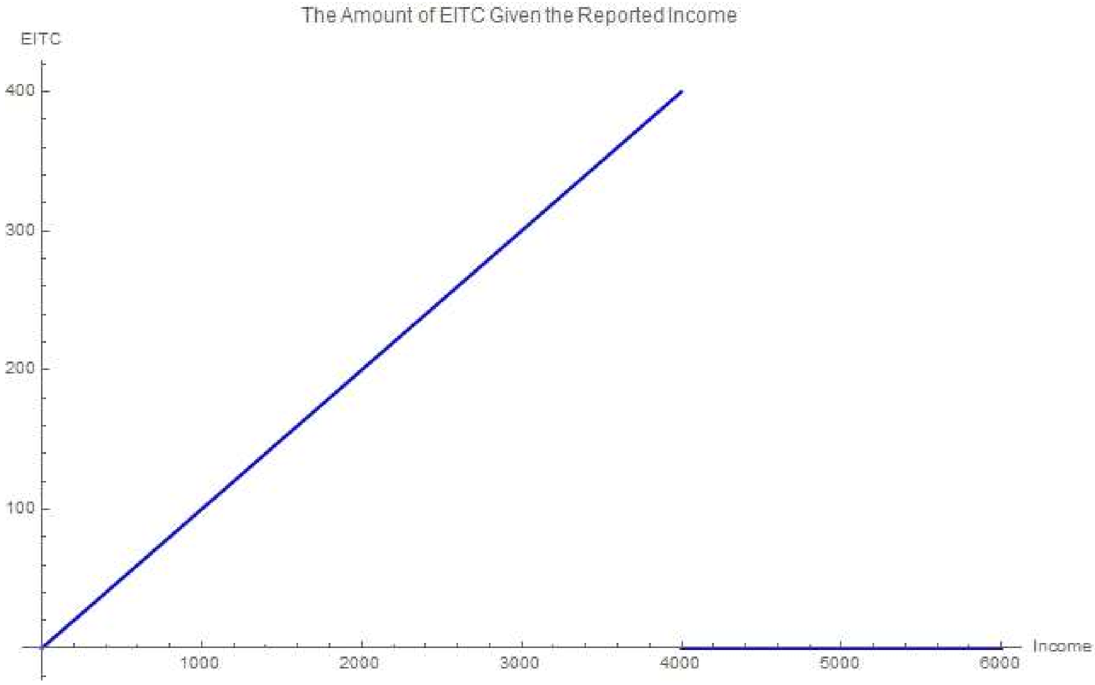


Figure 1.

Since the sample is focused on the bottom quintile of the US population, where most of the families' financial situation is hand-to-mouth. The model would then assume that a low-income family's school dropout decision is based on the current income, independent of previous income flow.

Preliminary Results:

The pooled OLS estimator is first presented in table 2 where the binary decision variable is regressed on family net income without controlling for other variables. This shows the negative statistically significant correlation between income and dropout decision. Notice that income is calculated in terms of $1,000 and that the interpretation for the coefficient would be an increase in $1,000 of income in 1975 is associated with 0.5% reduction in the probability of dropping out of school, which is economically insignificant. The second regression reports OLS estimates when controlling for other time-variant variables. One could observe that the coefficient on family net income becomes not only economically insignificant, but statistically insignificant. However, it is likely that the regression equation suffers from omitted variable bias, as family specific effects are shown in previous literature to have an important role in determining educational decisions. The third column reports the fixed effect model using first-differencing estimator. By accounting for time-invariant family fixed effect, the coefficient for family net income becomes more precisely estimated, yet the value only increased by 0.3%.

Table 2. OLS With and Without Time-Variant Control Variables and FD Results

School Dropout Decision	(1) OLS	(2) OLS Controlled for Time-Variant variables	(3) FD
Family Net income	-0.005***	-0.001	0.004**
	(0.0012)	(0.001)	(0.002)
Total Income Tax		-0.005	0.006*
		(0.004)	(0.004)
Yearly Property Tax		0.007**	-0.0009
		(0.003)	(0.0034)
Parental Age		0.003***	0.002*
		(0.0003)	(0.001)
Number of Kids		0.019***	0.021***
		(0.002)	(0.006)
Region		-0.012**	-0.004
		(0.005)	(0.013)
Constant		-0.030	N/A
		(0.016)	

Note: Robust standard errors are reported in the parentheses.

FEIV Results:

The fixed effect instrumental variable approach is reported in table 3, in which the first column shows that the estimation precision and the value for the effect of family net income on educational decision have both improved. Using the full dataset, one could observe that with $1,000 of increase in net income, the probability of school dropout decreases by 6.9%. However, the maximum amount of EITC given to households is $400. Thus, even with the maximum amount of EITC support, dropout probability would decrease by 3-4%. It is important to note that the null hypothesis for the coefficient for family net income would only be rejected at 10% of significance level, meaning the value is not very precisely estimated. Nevertheless, it is still interesting to see how different periods of implementation affects families' educational decision. In the third column, which reports the effect of income on educational decision right after the implementation of EITC, one could observe that the value is small relative to the analyses with later periods included and that the value is very imprecisely estimated. Towards later periods, as people start to expect EITC to be permanently implemented, their educational decision starts to be affected by net income. This finding is consistent with Blau (1999).

Table 3. FEIV Results with Time Heterogeneity

School Dropout Decision	(1) FEIV with Full Dataset	(2) FEIV with Dataset Until 1977	(3) FEIV with Dataset Until 1976
Family Net income	-0.069*	-0.072**	-0.038
	(0.040)	(0.037)	(0.025)
Total Income Tax	-0.011	-0.020	-0.016
	(0.038)	(0.036)	(0.019)
Yearly Property Tax	0.009	0.008	0.006
	(0.007)	(0.007)	(0.007)
Parental Age	0.0006	0.0006	0.001
	(0.0013)	(0.0013)	(0.001)
Number of Kids	0.016**	0.017**	0.018***
	(0.007)	(0.007)	(0.007)
Region	0.0005	0.0007	-0.001
	(0.0143)	(0.0144)	(0.014)

Note: Robust standard errors are reported in the parentheses.

Conclusion:

This paper evaluated the effect of income on educational decision of low-income households. There is moderate evidence suggests that an increase in income by $400 would lead to 3-4% of reduction in dropout possibility. However, the moderate result would not serve as a strong evidence for policy makers to use income stimulating methods to encourage school attendance. Additionally, one-period surge in income, such a lottery win or temporary tax refund, would have little effect on families' educational decision.

References:

Akee, R., Copeland, W., Keeler, G., Angold, A., Costello, E., (2010). Parents' Incomes and Children's Outcomes: a Quasi-Experiment Using Transfer Payments from Casino Profits. *American Economic Journal: Applied Economics* 2(1): pp. 86-115.

Angrist, J., Krueger, A., (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics* 106(4): pp. 979-1014.

Chevalier, A., Harmon, C., O'Sullivan, V., Walker, I., (2013). The Impact of Parental Income and Education on the Schooling of Their Children. *Journal of Labor Economics*: 2:8.

Dahl, G., Lochner, L., (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review* 102(5): pp. 1927-1956.

Figlio, D., Stone, J., (1999). Are Private Schools Really Better? *Research in Labor Economics* 18: pp. 115-140.

Foley, K., Gallipoli, G., Green, D., (2014). Ability, Parental Valuation of Education, and the High School Dropout Decision. *Journal of Human Resources* 49(4): pp. 906-944.

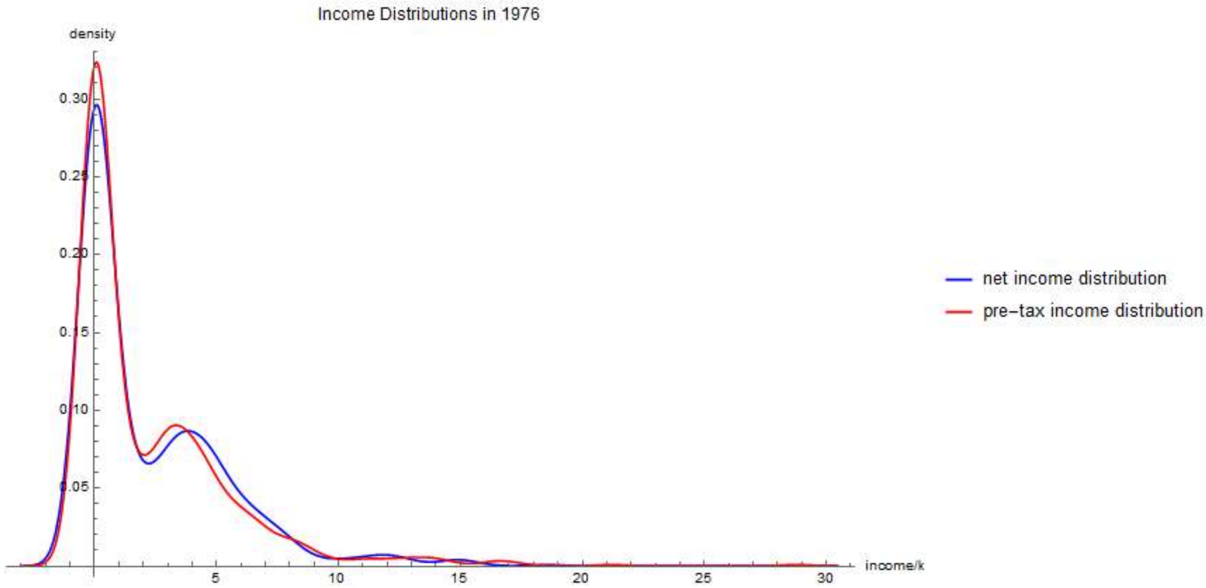
Harmon, C., Walker, I., (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *American Economic Review* 85(5): pp. 1278-1286.

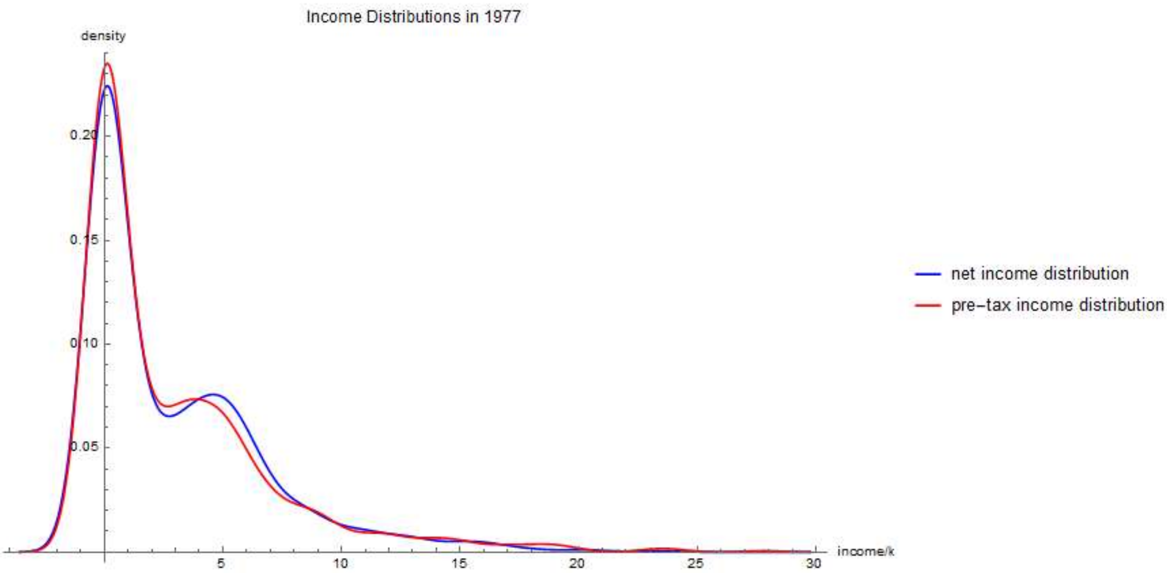
Løken, K., (2010). Family Income and Children's Education: Using the Norwegian Oil Boom as a Natural Experiment. *Labour Economics* 17(1): pp. 118-129.

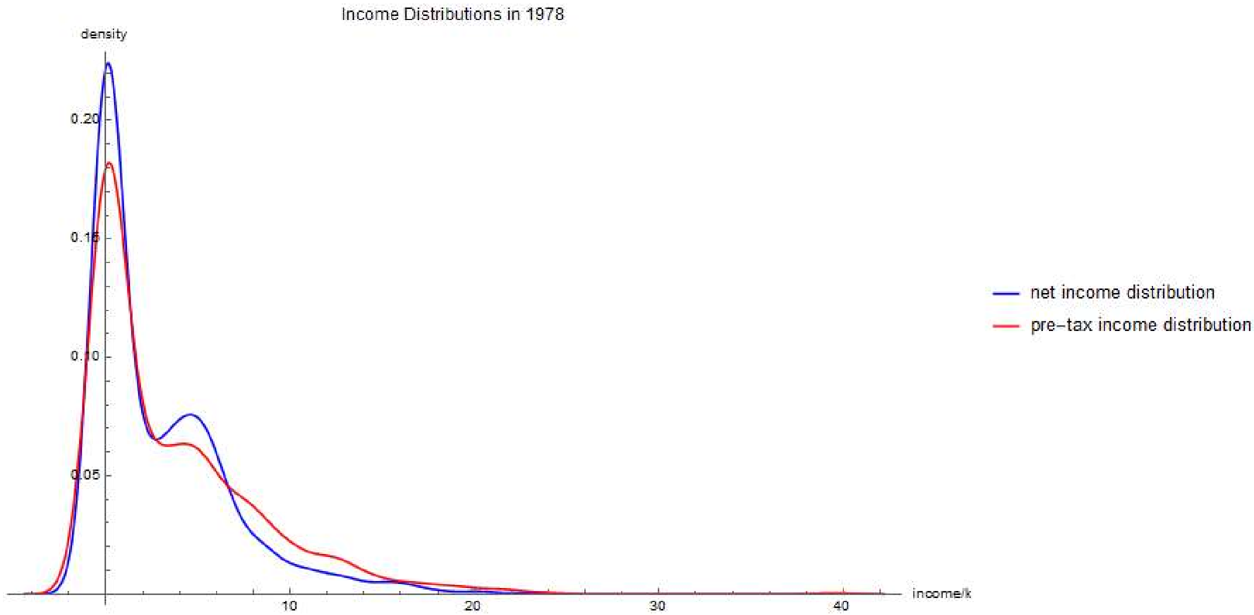
Shea, J., (2000). Does Parents' Money Matter? *Jounal of Public Economics* 77: pp. 155-184.

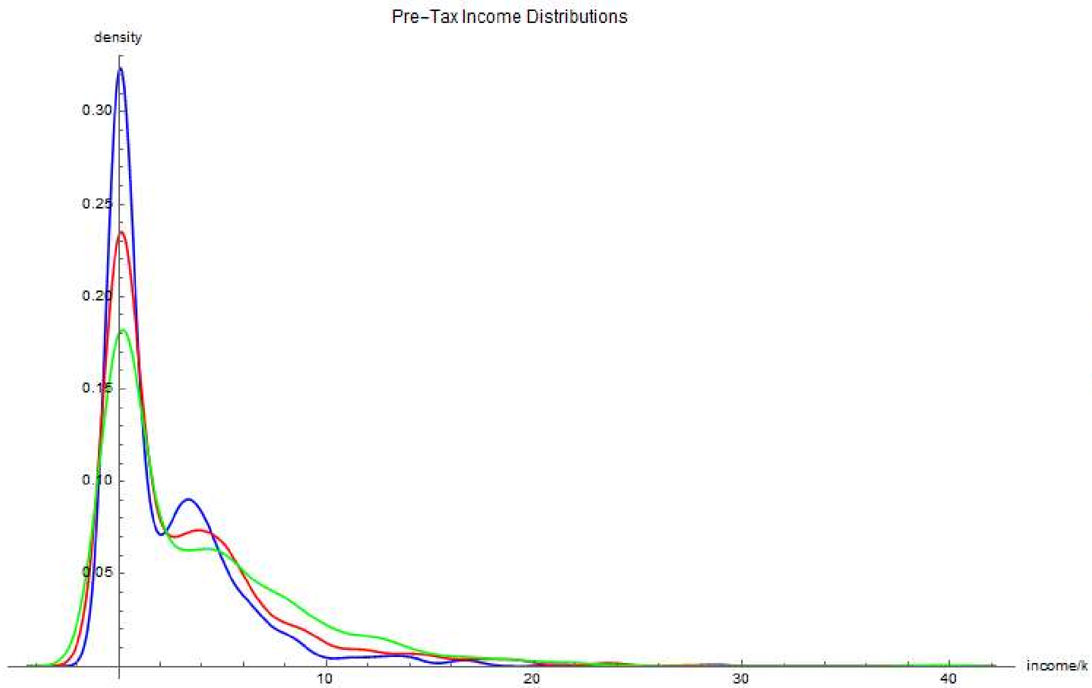
Stinebrickner, R., Stinebrickner, T., (2003). Understanding Educational Outcomes of Students from Low-Income Families – Evidence from a Liberal Arts College with a Full Tuition Subsidy Program. *Journal of Human Resources* 38(3): pp. 591-617.

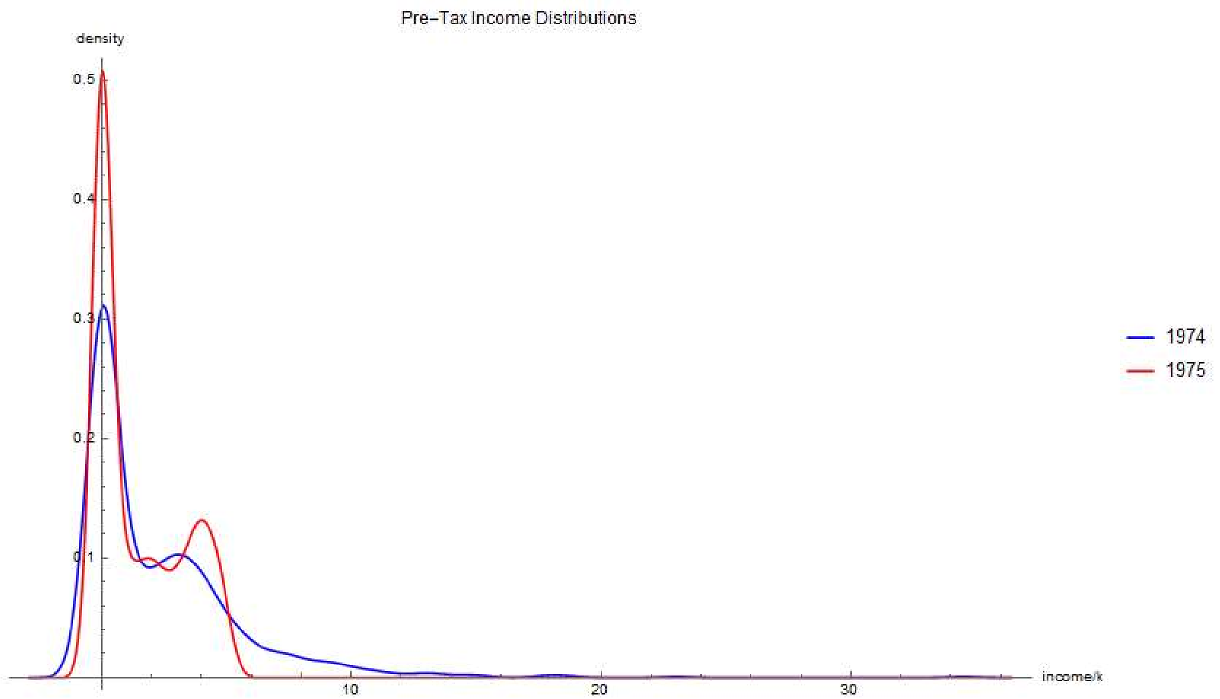
Appendix:

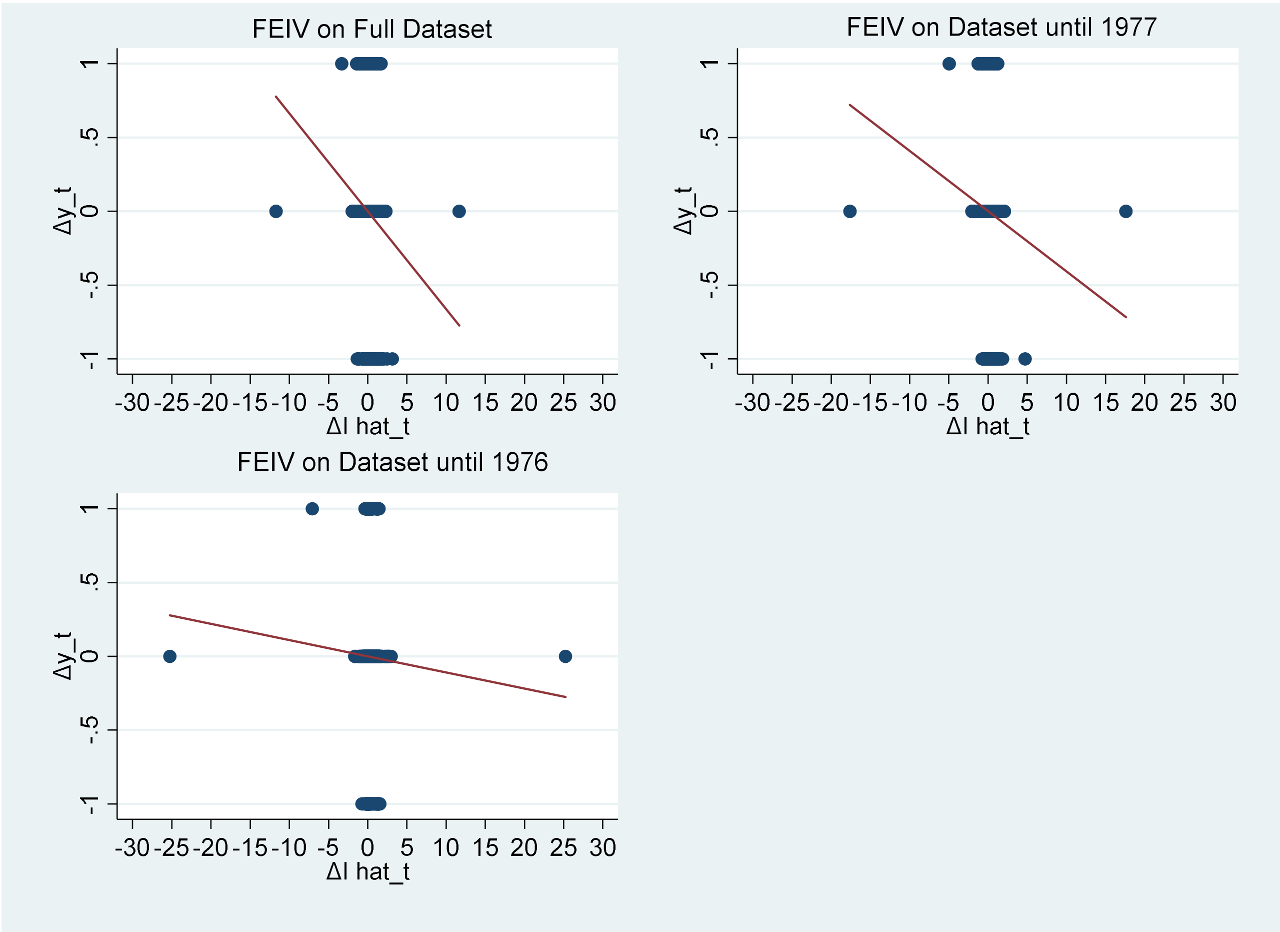

Income Distributions in 1976
density
0.30
0.25
0.20
0.15
0.10
0.05
5
10
15
20
25
30
income/k
net income distribution
pre-tax income distribution


Income Distributions in 1977
density
0.20
0.15
0.10
0.05
5
10
15
20
25
30
income/k
net income distribution
pre-tax income distribution


Income Distributions in 1978
density
0.20
0.15
0.10
0.05
10
20
30
40
income/k
net income distribution
pre−tax income distribution


Pre−Tax Income Distributions
density
0.30
0.25
0.20
0.15
0.10
0.05
10
20
30
40
income/k
1976
1977
1978


Pre−Tax Income Distributions
density
0.5
0.4
0.3
0.2
0.1
10
20
30
income/k
1974
1975


FEIV on Full Dataset
FEIV on Dataset until 1977
FEIV on Dataset until 1976
Δy_t
Δl hat_t
-1
-.5
0
.5
1
-30 -25 -20 -15 -10 -5 0 5 10 15 20 25 30

Descriptive Data for Family Net and Pre-tax Income

Year	*Mean/k*	*Median/k*	*Standard Deviation*
	Family Net income		
1970	2.64	1.673	0.11
1971	2.53	1.75	0.11
1972	2.37	0.95	0.12
1973	2.33	0.93	0.12
1974	2.10	1	0.13
1975	1.40	0.3	0.06
1976	2.24	0.55	0.11
1977	2.90	1.21	0.14
1978	3.57	2	0.16
	Family Pre-tax Income		
1970	2.78	1.69	0.12
1971	2.66	1.75	0.12
1972	2.49	0.96	0.13
1973	2.44	0.93	0.13
1974	2.26	1	0.12
1975	1.42	0.3	0.06
1976	2.23	0.5	0.12
1977	2.96	1.12	0.15
1978	3.67	1.9	0.17

Note: data from the Panel Study of Income Dynamics (PSID). Mean and median are reported in dollars.

Regression equation derivation:

A family's binary decision to let kids drop-out of school can be modelled as follows,

$$y_{it} = X_i'\alpha + W_{it}'\beta + I_{it}\gamma + u_i + \epsilon_{it},$$

where the subscript i denotes family identity and t denotes time. X_i is a vector of observed time-invariant factors, W_{it} is a vector consists of observed time-variant variables, I_{it} is family net income, u_i is unobserved family fixed effect and ϵ_{it} is the residual error term. To eliminate the family fixed effect, one could implement first-differencing method as follows,

$$\Delta y_{it} = \Delta W_{it}'\beta + \Delta I_{it}\gamma + \Delta\epsilon_{it}, where$$

$$\Delta y_{it} = y_{it} - y_{it-1} = (W_{it}' - W_{it-1}')\beta + (I_{it} - I_{it-1})\gamma + (\epsilon_{it} - \epsilon_{it-1})$$

One would then need to find an appropriate instrumental variable for ΔI_{it}. The family net income I_{it} can be elaborated in terms of the following equation,

$$I_{it} \equiv PI_{it} + a_{it} + \tau_{it}$$

Where PI_{it} is the pre-tax income of family i at time t, a_{it} is the amount of financial support that been given by the government for the family i under the EITC programme and τ_{it} is the total amount of income tax deducted from the pre-tax income. For simplicity, the model would treat income tax as exogenous. Pre-tax income, on the other hand, would be treated as endogenous in the model in the following specification,

$$PI_{it} = Z_i'\delta_1 + L_{it}'\delta_2 + \eta_{it}$$
$$Z_i \subset S, S = \{X_i, u_i\},$$
$$l_{it} \subset K, K = \{W_{it}\}$$

η_{it} is modelled as an unobserved random shock of pre-tax income with conditional expectation taken as zero,

$$E[\eta_{it}|Z_i, L_{it}] = 0, \forall t$$

Then family net income has the following relationship with the variables defined above,

$$I_{it} = Z_i'\pi_0 + L_{it}'\pi_1 + a_{it}\pi_2 + \tau_{it}\pi_3 + \phi_{it} \quad (*)$$
$$\phi_{it} = \eta_{it} + \omega_{it}$$

One could further assume that the random shock to pre-tax income η_{it} is conditionally independent to the residual term ω_{it}, i.e. $E[\eta_{it}\omega_{it}|Z_{it}, L_{it}] = 0, \forall t$.

However, since the way the model suggests that pre-tax income is related to both time-variant Z_i and invariant L_{it} variables, the amount of financial support from the EITC programme, which is dependent on pre-tax income, is dependent on Z_i and L_{it} as well, i.e.

$$a_{it} = f(PI_{it}) = f(Z_i, L_{it}, \eta_{it})$$

This means that $COV(a_{it}, \phi_{it}) \neq 0$. As the regressor a_{it} correlates with the residual, simply using $(*)$ to estimate the regression coefficients could be biased. Since the variable of interest is ΔI_{it}, it is possible to first-difference the first-stage regression, i.e.

$$\Delta I_{it} = \Delta L_{it}'\pi_1 + \Delta a_{it}\pi_2 + \Delta\tau_{it}\pi_3 + \Delta\eta_{\text{it}} + \Delta\omega_{\text{it}}$$

It is important to note that linear estimation, which takes the conditional expectation of ΔI_{it} can be written as follows,

$$E[\Delta I_{it}|Z_{it}, L_{it}, a_{it}, \tau_{it}] = E[\Delta L_{it}'\pi_2 + \Delta a_{it}\pi_3 + \Delta\tau_{it}\pi_4 + \eta_{\text{it}} - \eta_{\text{it}-1} + \Delta\omega_{\text{it}}|Z_{it}, L_{it}, a_{it}, \tau_{it}]$$
$$= \Delta L_{it}'\pi_2 + \Delta a_{it}\pi_3 + \Delta\tau_{it}\pi_4 + E[\eta_{\text{it}} - \eta_{\text{it}-1}|Z_{it}, L_{it}, a_{it}, \tau_{it}] + E[\Delta\omega_{\text{it}}|Z_{it}, L_{it}, a_{it}, \tau_{it}],$$

where $E[\eta_{\text{it}} - \eta_{\text{it}-1}|Z_{it}, L_{it}, a_{it}, \tau_{it}] = E[\eta_{\text{it}}|Z_{it}, L_{it}, a_{it}, \tau_{it}] - E[\eta_{\text{it}-1}|Z_{it}, L_{it}, a_{it}, \tau_{it}] = 0$ as assumed above.

Thus, using linear estimation with first-differencing model could prevent the random shock η_{it} to confound with estimators in the regression.

Thus, the empirical strategy of the fixed effect instrumental variable (FEIV) can be summarised as follows.

$$\Delta y_{it} = \Delta L_{it}'\rho_1 + \Delta I_{it}\rho_2 + \Delta\tau_{it}\rho_3 + \Delta\mu_{it}$$
$$\Delta I_{it} = \Delta L_{it}'\pi_1 + \Delta a_{it}\pi_2 + \Delta\tau_{it}\pi_3 + \Delta\phi_{\text{it}}$$