UNIVERSITY OF SOUTHAMPTON

DEPARTMENT OF ECONOMICS

SCHOOL OF ECONOMIC, SOCIAL & POLITICAL SCIENCES

DISSERTATION: RESEARCH TOPICS

Returns to Education

Full name of author:

(31946747)

Presented for B.Sc. (Social Sciences) in Economics

10 March 2022

or

28 April 2022

I declare that this dissertation is my own work, and that where material is obtained from published or unpublished work, this has been fully acknowledged in the references.

Signed: 31946747, 09/03/2023

Investigating Returns to Education using Parental Education as an Instrumental Variable through Two-Stage Least Squares (2SLS) Regression Technique

Contents

Word Count: 1974

Introduction

Understanding the education-earnings relationship is vital for both individual investments in schooling and national policy. Modelling returns to education promotes human capital growth according to Keynes's theory of economic growth [1] and Ricardo's theory of comparative advantage [2]. Furthermore, many empirical studies show that education indirectly reduces unemployment and increases occupational access for those who participate (Cohn & Addison, 1997; Psacharopoulos, 1985).

Returns to education are affected by a variety of characteristics, such as family background, ability, or institutional components, rather than being a fixed parameter in the population (Card, 1999). Using the NLSYM (Survey of Young Men) dataset for 1977, this literature builds on Jacob Mincer's (1974) human capital earnings function (hereafter, HCEF) by exploring the use of exogenous family background

[1] Keynes argues that investing in education enables the acquisition of new abilities and knowledge, leading to the creation of new goods, services, and technologies (Brown et al., 2010). Human capital is seen as a major driver of economic growth and development particularly in knowledge-based economies, where it opens up new opportunities for entrepreneurship and economic activity.

[2] Education enables the acquisition of skills that are of more valuable compared to others, suggesting nations *and* individuals should specialise in what they are relatively good at (Judy & D'amico, 1997).

covariates, introduced as instrumental variables (hereafter, IVs), to decompose the causal relationship between education and earnings. This approach aims to provide policymakers and educators with greater interpretability for returns to education and inform policy decisions on how best to allocate resources to promote economic growth and quality of life.

Through use of two-stage least squares (hereafter, 2SLS), the research question of this study is to test the effect of parental education as the IV on estimates of returns to education.

Findings show that parental education has a significant, positive relationship with their children's education outcomes, returns are 12.09%, or, relatively, 70% higher than that of the OLS methodology (7.59%); consistent with previous research by Shi (2016), who finds a higher estimate for the 2SLS estimator. The Wu-Hausman test confirms that the OLS estimator is endogenous and exhibits heteroskedasticity, while the 2SLS estimator is preferred, supporting the use of parental education as an adequate IV for future research in this area.

Economic Analysis

Model Specification

According to Mincer's HCEF, the log of individual earnings (y) can be broken into a linear education term and a quadratic experience term.

$$\ln(y) = \beta_0 + \beta_1 b + \beta_2 e + \beta_3 e^2 + \varepsilon \tag{1}$$

where β_0 is a the expected log earnings when education and experience are both zero (constant), β_1 is the "return to education" coefficient, β_2 is the linear effect of experience on earnings, β_3 is the quadratic effect of experience on earnings (used as the return to experience generally diminishes with age), and ε is the error term.

Willis (1986) asserts that if the number of years of completed schooling accurately reflects an individual's education, and if each additional year has a constant effect on labour market returns, then coefficient β_1 fully internalises the returns to education. The first assumption has face validity more in the United States than European countries with multiple education streams (Card, 1999). Accordingly, the NLSYM dataset provides data for United States men entering the workforce (monitoring 14-to-24-year-olds after a ten-year-period), and a diverse set of variables relative to other census results from the time (Perez & Hirschman, 2009).

We recreate a simple HCEF variant using ordinary least-squares (OLS), estimating the log-level weekly way in 1976 against education, experience, and experience squared, as shown in Figure 1. Adjusted R-squared shows that 19.5% of the variance in earnings is explained, with the coefficient of education indicating 9.3% return per additional year of education, while experience has marginally weaker effect, and this significance is supported with t-statistic values.

```
. reg lwage educ exper expersq

      Source |       SS           df       MS      Number of obs   =     3,010
-------------+----------------------------------   F(3, 3006)      =    243.99
       Model |  116.049688         3  38.6832293   Prob > F        =    0.0000
    Residual |  476.591923     3,006  .158546881   R-squared       =    0.1958
-------------+----------------------------------   Adj R-squared   =    0.1950
       Total |  592.641611     3,009  .196956335   Root MSE        =    .39818

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0931707   .0035802    26.02   0.000     .0861508    .1001906
       exper |   .0897828   .0070636    12.71   0.000     .0759328    .1036328
     expersq |  -.0024859   .0003377    -7.36   0.000    -.0031481   -.0018237
       _cons |   4.468541   .0686899    65.05   0.000     4.333857    4.603224
------------------------------------------------------------------------------
```

Figure 1

Card (1999) finds that simple variants of the HCEF can explain 20-35% of the observed earnings variation. They suggest that any significant improvements to the initial model will come from building more flexible interactions between education and experience. This can be done (1) by incorporating more exogenous variables to explain the causal relationship, and (2) by better interpreting the endogeneity in the existing variables.

Data Analysis

Appendix 1.1. shows the summary of a selection of statistics from the NLSYM dataset, which began in 1966 with 5525 men aged 14-24. In 1976, 71% of the original sample remained (falling to 61% by 1981), while representing a decade of change and over a quarter reporting 15+ years of education (Card, 1999).

Most notably, sampling bias exists in the data; the maximum years of education for all surveyed individuals is 18 years, leading to clustering at the maximum value (18) as some certificates take indeterminately longer, resulting in downward bias (Card, 1999). Moreover, Appendix 1.2. shows over-sampling for men in the South and the black population, relative to nationally representative samples (41% versus 32% for region, and 28% versus 10% for race). Stratified or random sampling techniques would ensure that the sample is representative of the population (Kish, 1965), and corresponding variables are used as controls for all models in this report.

OLS Regression

In Figure 2, OLS regression is implemented in three parts. Geographical components "Lived in Standard-Metropolitan-Statistical-Area" and "Lived in South" (both 1976) are highly statistically significant, while "4-year college proximity" and "single-mother" variables are insignificant. Conceivably, this may be due to endogeneity between variables. Parts (2 & 3) are denoted as:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 smsa + \beta_5 south + e \tag{2}$$

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 smsa + \beta_5 south + \beta_6 married + \beta_7 momdad14 + \beta_8 sinmom14 + e \tag{3}$$

We find from Eq. (3) (see Appendix 3.1.) that the full set of family background variables aren't jointly significant, though "married" has a strong relationship with earnings. Returns to education are 7.6% per year, consistent with Card (1993). Notably, with more variables, the unobserved error decreases with a greater standard error.

	(1) lwage	(2) lwage	(3) lwage
educ	0.0932***	0.0813***	0.0759***
	(0.00358)	(0.00351)	(0.00350)
exper	0.0898***	0.0837***	0.0730***
	(0.00706)	(0.00677)	(0.00674)
expersq	-0.00249***	-0.00220***	-0.00190***
	(0.000338)	(0.000324)	(0.000320)
nearc4		0.0187	0.0180
		(0.0162)	(0.0160)
smsa		0.145***	0.159***
		(0.0167)	(0.0164)
south		-0.172***	-0.166***
		(0.0148)	(0.0146)
married_			0.157***
			(0.0156)
momdad14			0.0618**
			(0.0223)
sinmom14			0.00224
			(0.0298)
Constant	4.469***	4.607***	4.570***
	(0.0687)	(0.0680)	(0.0684)
Observations	3010	3010	3003

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Figure 2

Instrumental Variable Analysis

The HCEF assumes that education is free and that students do not simultaneously gain experience (or earn) while enrolled (Heckman et al., 1998). This suggests that within the existing model, components of educational quality are not internalised. Decomposing this part of the model can increase estimates significantly (Chetty et al., 2014). To address endogeneity we use IVs to control for the correlation between the endogenous variables and the error term with 2SLS. During first-stage, the IVs predict the endogenous variables. During second-stage, the predicted values are used as regressors.

A robust earnings-education instrument necessitates identifying a source of exogenous variation in education choices, uncorrelated with the dependent variable and error, but correlated with the predictor. Family background is a promising candidate as parental education and incomes can give their children better access to informed advice and more expensive opportunities (e.g. tutoring, postgraduate), afterall higher school quality is associated with higher earnings (Card & Krueger, 1992).

Mazumder (1997) shows that parental education is highly correlated with schooling outcomes as families with high genetic "ability", such as IQ, may be more motivated to achieve labour market success. This induces positive correlation between parental education ("fatheduc", "motheduc") and unobserved determinants of education[3]. Card (1999) shows each additional year of schooling for either parent raises their child's completed education by 0.2 years, and also explains 30% of the observed variation in General Social Survey (GSS) data.

Figure 3 shows negative correlation between education and experience, corresponding to Mincer's "potential experience" equation, which mechanically relates the variables (visualised also in Figure 4). As mentioned previously, multicollinearity due to over-sampling ensures we control using race and south dummy variables.

```
. correlate lwage educ exper nearc4 fatheduc motheduc
(obs=2,220)
```

	lwage	educ	exper	nearc4	fatheduc	motheduc
lwage	1.0000					
educ	0.2763	1.0000				
exper	0.0723	-0.6239	1.0000			
nearc4	0.1323	0.1258	-0.0626	1.0000		
fatheduc	0.1887	0.4692	-0.3571	0.1367	1.0000	
motheduc	0.1987	0.4396	-0.3163	0.0773	0.6315	1.0000

Figure 3

Griliches (1979) qualifies parental education as a instrumental variable candidate, finding that parental education affects earnings primarily via their effect on the level of achieved schooling, suggesting exogeneity with wages. However, high correlation between maternal and paternal education may indicate multicollinearity, which may cause inflated standard errors when used as IVs. Therefore, it is important to assess the adequacy of these instruments with Adjusted R-squared. Alternative IVs are explored in other papers[4]. While Griliches (1977) and Becker (1964) suggest such biases could be smaller than other

[3] One approach would be to use twin studies to reduce "ability bias" and "institutional bias", which could upward-skew the OLS regressions if individuals with higher ability choose to acquire more schooling, thereby providing a more direct estimation of the returns to education (Bonjour, et al., 2003). However, twin studies may still suffer from some limitations, such as the standard error due to small sample sizes and the assumption that twins have similar educational experiences, which may not always be the case (Miller, et al., 1995).

[4] (see Card, 1999).

measurement biases (up to 10% of the variance), but targeting endogeneity is still important for studies which build upon and improve interpretability of the HCEF framework, as emphasised by Polachek (2007) and Card (1999).

A distributional analysis of family background is conducted in Figure 4, showing that modern 24-to-34-year-old males stay in education longer than their parents. Implicitly, this suggests that returns to education may be diminishing longitudinally, as the existing workforce in real terms earn the same as the previous generation, despite better education (Denison, 1962; Mankiw, 1992).

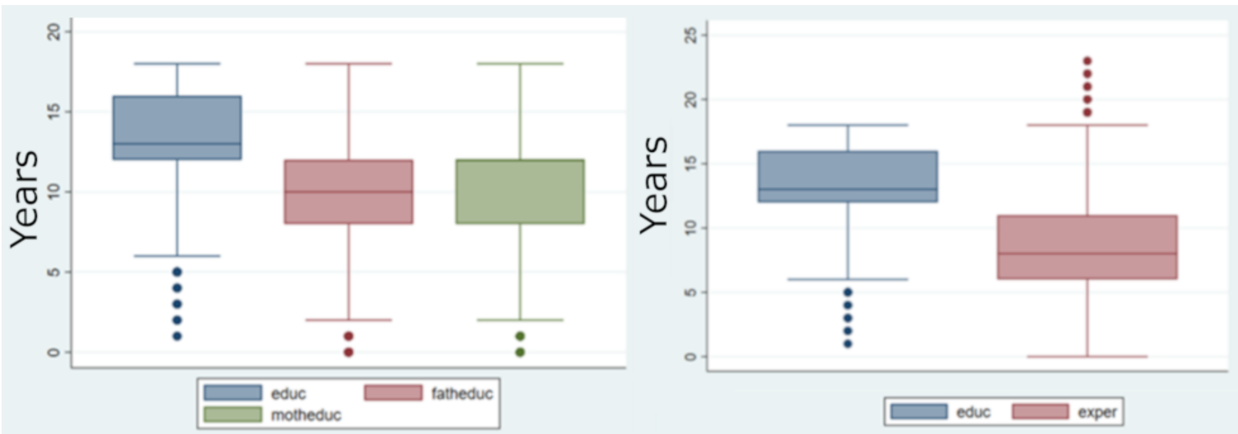

Figure 4

Results

First-stage

We can use the first stage to discern the relative endogeneity for variables in the model. For the variables selected, we run an OLS regression against education in the form Eq.(4) where the parental education variables are the instrumental variables (Z).

$$educ = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 smsa + \beta_4 south + \beta_5 married + \beta_6 momdad14 + \beta_7\, sinmom14 \; + \boldsymbol{\beta_8 motheduc} \; + \boldsymbol{\beta_9 fatheduc} + v \tag{4}$$

The results present parental education as a strong candidate for an IV, given strong t-statistics (above threshold of 2 for 5% significance level) and positive correlation with education. However, measurement error is a concern as “fatheduc” and “motheduc” are missing for 22.9% and 11.7% of the males surveyed, respectively, which reduces the usable data to only 73.1% (2,200 of 3,010 total 1976 observations) via sampling bias, without a corresponding variable to control with.

Consequently, bias may exist due to heteroskedasticity, which would reduce certain estimator robustness if present. In Figure 5 (right), the results of the Breusch-Pagan/Cook-Weisberg test for hetero-skedasticity in STATA indicate that the probability of observing a significant (larger than 61.94) test statistic under the null hypothesis is effectively zero, so we reject the null hypothesis. This may invalidate the statistical inference of the OLS model.

```
. eststo: reg educ fatheduc motheduc exper expersq smsa south married momdad14 sinmom14

      Source |       SS           df       MS      Number of obs   =     2,215
-------------+----------------------------------   F(9, 2205)      =    231.34
       Model |  7195.48995         9  799.498883   Prob > F        =    0.0000
    Residual |  7620.42247     2,205  3.45597391   R-squared       =    0.4857
-------------+----------------------------------   Adj R-squared   =    0.4836
       Total |  14815.9124     2,214  6.69192069   Root MSE        =     1.859

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    fatheduc |   .1184158   .0142815     8.29   0.000     .0904091    .1464224
    motheduc |   .1352176   .0168992     8.00   0.000     .1020776    .1683576
       exper |  -.3965165   .0387287   -10.24   0.000     -.472465   -.3205679
     expersq |   .0028976   .0019719     1.47   0.142    -.0009694    .0067646
        smsa |   .3427997   .0911184     3.76   0.000     .1641128    .5214866
       south |  -.1834107   .0844238    -2.17   0.030    -.3489691   -.0178523
    married_ |   .3077195   .0915368     3.36   0.001     .1282121    .4872269
    momdad14 |   .7847627   .1970211     3.98   0.000     .3983964    1.171129
    sinmom14 |   .9300499   .4914996     1.89   0.059    -.0338006      1.8939
       _cons |   12.89097   .3200471    40.28   0.000     12.26334     13.5186
------------------------------------------------------------------------------
(est3 stored)
```

```
. predict resid, residuals
(795 missing values generated)

. hettest resid

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: resid

H0: Constant variance

    chi2(1) =  61.94
Prob > chi2 = 0.0000
```

Figure 5 (left: firststage regression. right: heteroskedasticity test)

Parental education has a positive and significant impact on education, with each additional year of parental education resulting in 0.118-0.135 years of education. The adjusted R-squared value of 48% indicates that parental education explains a substantial portion of the variance in earnings, and when we remove mechanically related variables like experience and experience squared, they account for over half of the explanation (see Appendix 2).

Other research shows that using parental education as an IV in 2SLS regression can lead to stronger results, particularly for maternal education, as demonstrated by Shi (2016) using PSID data[5]. Shi finds that OLS estimates are downward-biased compared to IV estimates, implying parental education may be a strong instrumental variable[6].

Second-stage

Figure 6, column 3 is a second-stage regression against lwage using these two variables as IV, where the output of Eq.(2) is $\widehat{\boldsymbol{educ}}$:

$$\log(wage) = \beta_0 + \beta_1 \widehat{\boldsymbol{educ}} + \beta_2 exper + \beta_3 exper^2 + \beta_4 smsa + \beta_5 south + \beta_6 married + \beta_7 momdad14 + \beta_8 sinmom14 + v \tag{6}$$

[5] Additionally, proximity to a 4-year college has been found to be highly significant in predicting returns to education, as shown by Card (1993) using the NLSYM dataset.
[6] In their study, Shi (2016) finds the OLS estimate is downward-biased compared to the IV estimate.

	(1) lwage	(2) educ	(3) lwage
educ	0.0759***		0.121***
	(0.00350)		(0.0119)
exper	0.0730***	-0.397***	0.0977***
	(0.00674)	(0.0387)	(0.00964)
expersq	-0.00190***	0.00296	-0.00218***
	(0.000320)	(0.00197)	(0.000409)
nearc4	0.0180	0.192*	-0.00757
	(0.0160)	(0.0925)	(0.0194)
smsa	0.159***	0.281**	0.148***
	(0.0164)	(0.0958)	(0.0204)
south	-0.166***	-0.154	-0.132***
	(0.0146)	(0.0855)	(0.0182)
married_	0.157***	0.306***	0.140***
	(0.0156)	(0.0915)	(0.0193)
momdad14	0.0618**	0.779***	0.0214
	(0.0223)	(0.197)	(0.0420)
sinmom14	0.00224	0.883	0.0524
	(0.0298)	(0.492)	(0.102)
fatheduc		0.117***	
		(0.0143)	
motheduc		0.136***	
		(0.0169)	
Constant	4.570***	12.81***	3.837***
	(0.0684)	(0.322)	(0.197)
Observations	3003	2215	2215
Adjusted R-squared	0.290	0.484	0.229

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Figure 6

When parental education is taken as an exogenous determinant of schooling, the implied IV estimate for returns to education is 12.09% with a standard error of 0.0119, corresponding to a 68.7% increase over the corresponding OLS estimates. This indicates that endogeneity bias was present in the OLS regression, corresponding to 10% returns to education for 2SLS estimator by Shi (2016), against 5.87% for OLS (70.4% increase).

Comparing R-squared values for OLS (Appendix 3.1) with IV (Appendix 3.2), OLS explains 6% more of the relationship with a lower RMSE score, but also has 26% more datapoints. To determine whether the IV estimator is more efficient with smaller variance, we run the Wu-Hausman test (Figure 7), comparing the efficiency and consistency of the 2SLS and OLS estimators. We reject the null hypothesis (F-statistic=15.916, p-value=0.0001), indicating OLS is endogenous and inefficient, potentially due to exhibiting heteroskedasticity. Overall, parental education is an adequate IV and the 2SLS estimator is preferred.

```
. estat endog

  Tests of endogeneity
  H0: Variables are exogenous

  Durbin (score) chi2(1)          =  15.8811  (p = 0.0001)
  Wu-Hausman F(1,2204)            =  15.9163  (p = 0.0001)
```

Figure 7

Conclusion

In conclusion, evidence suggests that parental education has a significant, positive relationship with their children's educational outcomes, and translates into higher earnings through greater returns to education.

While the 2SLS estimator produces a higher estimate for returns to education than the OLS estimator, the latter explains slightly more of the relationship between education and earnings. The Wu-Hausman test confirms that the OLS estimator is endogenous and inefficient, and the Breusch-Pagan/Cook-Weisberg test suggests heteroskedasticity.

When using parental education “motheduc” and “fatheduc” as instruments for education, returns are 12.09%, or 70% higher than that of the OLS methodology (7.59%). Shi (2016) also found a higher estimate for the 2SLS estimator.

These findings support the use of parental education as an adequate instrumental variable for future research in this area. However, it is important to acknowledge missing data and measurement error, which may reduce statistical significance and increase endogeneity. Future research could address these gender and sampling limitations with GSS, PSID, and NSFH datasets, and further explore the implications of our findings for policy and practice in education and labor markets.

Appendix

(1.1.)

```
. sum lwage educ exper expersq black south motheduc fatheduc
```

Variable	Obs	Mean	Std. dev.	Min	Max
lwage	3,010	6.261832	.4437976	4.60517	7.784889
educ	3,010	13.26346	2.676913	1	18
exper	3,010	8.856146	4.141672	0	23
expersq	3,010	95.57907	84.61831	0	529
black	3,010	.2335548	.4231624	0	1
south	3,010	.4036545	.4907113	0	1
motheduc	2,657	10.34814	3.179671	0	18
fatheduc	2,320	10.00345	3.720737	0	18

(1.2.)

```
. sum south black
```

Variable	Obs	Mean	Std. dev.	Min	Max
south	3,010	.4036545	.4907113	0	1
black	3,010	.2335548	.4231624	0	1

```
. sum black if south == 0
```

Variable	Obs	Mean	Std. dev.	Min	Max
black	1,795	.1147632	.3188248	0	1

```
. sum black if south == 1
```

Variable	Obs	Mean	Std. dev.	Min	Max
black	1,215	.4090535	.4918616	0	1

(2.2.)

```
. reg educ smsa south married momdad14 sinmom14

      Source |       SS           df       MS      Number of obs   =     3,003
-------------+----------------------------------   F(5, 2997)      =     63.88
       Model |  2072.21172         5  414.442345   Prob > F        =    0.0000
    Residual |  19444.6681     2,997  6.48804407   R-squared       =    0.0963
-------------+----------------------------------   Adj R-squared   =    0.0948
       Total |  21516.8798     3,002  7.16751492   Root MSE        =    2.5472

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        smsa |   .9011322   .1049573     8.59   0.000     .6953365    1.106928
       south |   -.856986   .0968787    -8.85   0.000    -1.046942   -.6670305
    married_ |  -.3222018    .103589    -3.11   0.002    -.5253144   -.1190891
    momdad14 |   1.171563   .1502895     7.80   0.000     .8768823    1.466244
    sinmom14 |   .0215816   .2029382     0.11   0.915    -.3763308    .4194939
       _cons |   12.27305   .1814342    67.64   0.000      11.9173    12.62879
------------------------------------------------------------------------------
```

```
. reg educ motheduc fatheduc smsa south married momdad14 sinmom14

      Source |       SS           df       MS      Number of obs   =     2,215
-------------+----------------------------------   F(7, 2207)      =    113.20
       Model |  3914.18841         7  559.169773   Prob > F        =    0.0000
    Residual |   10901.724     2,207  4.93961215   R-squared       =    0.2642
-------------+----------------------------------   Adj R-squared   =    0.2619
       Total |  14815.9124     2,214  6.69192069   Root MSE        =    2.2225

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    motheduc |   .1978335   .0200507     9.87   0.000     .1585133    .2371537
    fatheduc |   .2095957   .0167028    12.55   0.000     .1768408    .2423505
        smsa |   .4572497   .1088287     4.20   0.000     .2438324     .670667
       south |  -.1057311   .1008662    -1.05   0.295    -.3035336    .0920715
    married_ |  -.2769031    .106508    -2.60   0.009    -.4857695   -.0680367
    momdad14 |   .5636411   .2352755     2.40   0.017     .1022566    1.025026
    sinmom14 |   .4134194   .5869528     0.70   0.481    -.7376182    1.564457
       _cons |   8.785869   .3111661    28.24   0.000      8.17566    9.396078
------------------------------------------------------------------------------
```

(3.1.)

```
. reg lwage educ exper expersq nearc4 smsa south married momdad14 sinmom14
```

Source	SS	df	MS
Model	172.705988	9	19.1895542
Residual	418.255129	2,993	.139744447
Total	590.961117	3,002	.196855802

Number of obs	=	3,003
F(9, 2993)	=	137.32
Prob > F	=	0.0000
R-squared	=	0.2922
Adj R-squared	=	0.2901
Root MSE	=	.37382

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0759241	.0035019	21.68	0.000	.0690578	.0827904
exper	.0729904	.0067411	10.83	0.000	.0597728	.086208
expersq	-.0018996	.0003197	-5.94	0.000	-.0025265	-.0012727
nearc4	.0179922	.0159575	1.13	0.260	-.0132966	.0492809
smsa	.1592968	.0164371	9.69	0.000	.1270676	.191526
south	-.165977	.0146059	-11.36	0.000	-.1946157	-.1373383
married_	.1570057	.015625	10.05	0.000	.1263689	.1876426
momdad14	.0618147	.0222881	2.77	0.006	.0181131	.1055162
sinmom14	.0022368	.0297921	0.08	0.940	-.0561782	.0606518
_cons	4.570028	.0683664	66.85	0.000	4.435978	4.704078

(3.2.)

```
. ivregress 2sls lwage (educ = motheduc fatheduc) exper expersq nearc4 smsa south married_ momdad14 sinmom14
```

Instrumental variables 2SLS regression

Number of obs	=	2,215
Wald chi2(9)	=	530.32
Prob > chi2	=	0.0000
R-squared	=	0.2322
Root MSE	=	.38496

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
educ	.1209355	.0118792	10.18	0.000	.0976526	.1442184
exper	.0977175	.009636	10.14	0.000	.0788313	.1166036
expersq	-.0021807	.0004092	-5.33	0.000	-.0029828	-.0013787
nearc4	-.0075708	.0193608	-0.39	0.696	-.0455172	.0303757
smsa	.1477614	.0203945	7.25	0.000	.1077889	.1877338
south	-.132451	.0181573	-7.29	0.000	-.1680386	-.0968635
married_	.1396598	.0193	7.24	0.000	.1018325	.177487
momdad14	.0213722	.0419528	0.51	0.610	-.0608538	.1035983
sinmom14	.0523662	.1022303	0.51	0.608	-.1480016	.252734
_cons	3.836652	.1966321	19.51	0.000	3.45126	4.222043

Instrumented: educ
Instruments: exper expersq nearc4 smsa south married_ momdad14 sinmom14
motheduc fatheduc

(4.0.)

```
ssc install estout

** Regression HCEF **

reg lwage educ exper expersq

** Correlation matrix **

correlate lwage educ exper nearc4 fatheduc motheduc

sum lwage educ exper expersq black south motheduc fatheduc

** Run regression experiments **

eststo clear

reg lwage educ exper expersq nearc4 black south
reg lwage educ exper expersq nearc4 black south motheduc
reg lwage educ exper expersq nearc4 black south motheduc fatheduc

** OLS **

eststo: reg lwage educ exper expersq
eststo: reg lwage educ exper expersq nearc4 smsa south
eststo: reg lwage educ exper expersq nearc4 smsa south married momdad14 sinmom14

esttab, label se

reg lwage educ exper expersq nearc4 smsa south married momdad14 sinmom14 motheduc fatheduc

** IV **

ivregress 2sls lwage (educ = motheduc fatheduc) exper expersq nearc4 smsa south married_ momdad14 sinmom14

estat endog
estat firststage

** Present Results **

eststo clear

eststo: regress lwage educ exper expersq nearc4 smsa south married_ momdad14 sinmom14
eststo: ivregress 2sls lwage (educ = motheduc fatheduc) exper expersq nearc4 smsa south married_ momdad14 sinmom14
eststo: regress educ exper expersq nearc4 smsa south married_ momdad14 sinmom14 fatheduc motheduc

esttab, label se ar2
* mtitle("OLS" "IV" "FS")
```

References

Becker, W. C. (1964). Consequences of different kinds of parental discipline. *Review of child development research*, *1*(4), 169-208.

Bonjour, D., Cherkas, L. F., Haskel, J. E., Hawkes, D. D., & Spector, T. D. (2003). Returns to education: Evidence from UK twins. American Economic Review, 93(5), 1799-1812.

Brown, P., Lauder, H., & Ashton, D. (2010). *The global auction: The broken promises of education, jobs, and incomes*. Oxford University Press.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Available at: https://www.nber.org/system/files/working_papers/w4483/w4483.pdf

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, *3*, 1801-1863.

Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of political Economy*, *100*(1), 1-40.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. American economic review, 104(9), 2633-2679.

Cohn, E., & Addison, J. T. (1998). The economic returns to lifelong learning in OECD countries. Education economics, 6(3), 253-307.

Denison, E. (1962). The Sources of Economic Growth in The United States and the Alternatives Before Us. By Edward F. Denison. New York: Committee for Economic Development, 1962. Pp. 308.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, 1-22.

Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. Journal of political Economy, 87(5, Part 2), S37-S64.

Heckman, J. J., Lochner, L. J., & Taber, C. R. (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. The Review of Economic Dynamics, 1(1), 1-58.

Judy, R. W., & D'amico, C. (1997). *Workforce 2020: Work and workers in the 21st century*. Hudson Institute, Herman Kahn Center, PO Box 26-919, Indianapolis, IN 46226; tele.
Kish, L. (1965). Survey sampling: John Willey and Sons. *Inc. NY*.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. The quarterly journal of economics, 107(2), 407-437.

Mazumder, B. (2005). Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. The Review of Economics and Statistics, 87(2), 235–255. http://www.jstor.org/stable/40042900

Miller, P., Mulvey, C., & Martin, N. (1995). What do twins studies reveal about the economic returns to education? A comparison of Australian and US findings. The American Economic Review, 85(3), 586-599.

Mincer, J. (1974). Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

Perez, A. D., & Hirschman, C. (2009). The Changing Racial and Ethnic Composition of the US Population: Emerging American Identities. Population and development review, 35(1), 1–51. https://doi.org/10.1111/j.1728-4457.2009.00260.x

Psacharopoulos, G. (1985) "Returns to Education: A Further International Update and Implications." Journal of Human Resources 20: 583-604.

Polachek, S. W. (2008). Earnings over the life cycle: The Mincer earnings function and its applications. *Foundations and Trends® in Microeconomics*, *4*(3), 165-272.

Shi, Z. (2016). Effects of parental education on Return to education.

Willis, Robert J. (1986), "Wage determinants: a survey and reinterpretation of human capital earnings functions", in: Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North Holland, Amsterdam)