Intergenerational Income Mobility in the UK

Shu Rou Ng
Financial Economics
School of Economics
University of Kent, July 2023

Abstract

This paper estimates the intergenerational income elasticity in the UK. Earlier UK studies used the OLS estimator and a rank-based estimator to estimate the intergenerational income elasticity, but their results were an underestimate because of lifecycle bias and attenuation bias. I use a longitudinal UK data and find that lifecycle bias is caused by scale mismeasurement by comparing OLS estimates and rank-based estimates. I find that taking the average of sons' earnings at different points as sons' lifetime earnings is important. Besides that, I also discover that sons' earnings at early ages suffer from lifecycle bias. Moreover, IGE decreases when controls for son's non-cognitive skills, cognitive skills, and qualifications are used. Lastly, I find that IV estimates are upward biased because parent's education and parent's employment as instruments for parental lifetime income might be positively correlated with sons' earnings.

Chapter 1: Introduction and Literature Review

1.1: Introduction

It is interesting to observe the extent to which income is passed down from generation to generation. Intergenerational income elasticity (IGE) is the measure of intergenerational income mobility. The higher the IGE, the greater the intergenerational income relationship across generations, and the lower the intergenerational mobility. A weak intergenerational income mobility suggests that a child income depends more on his or her parents' income. Blanden *et al.* (2004) discovered that the UK has a lower intergenerational income mobility compared to North American and Europe countries. In addition, intergenerational income mobility in UK is low compared to the Nordic countries (Björklund *et al.*, 2006). Previous paper found a 0.30 of IGE for men (Blanden *et al.*, 2004), while Gregg *et al*. (2017) found a higher intergenerational income elasticity, 0.43.

This research mainly focuses on discussing lifecycle bias and attenuation bias, which are the major intergenerational income elasticity estimation problems discovered by previous authors. The main objective of this research is to look at how the IGE changes at different periods across the lifecycle of an individual; and find the intergenerational income elasticity between sons and parents in the UK. Moreover, many previous UK studies used only the OLS approach and rank-based estimates to estimate the IGE, but this study provides new insight and tries to estimate the IGE by using the IV approach. It is of interest to know whether the methodology used in this research results in a higher or lower intergenerational income mobility in the UK compared to previous studies.

This study uses the OLS approach, rank-based estimation and IV approach. This research found that the UK's intergenerational income mobility is between 0.27 and 0.50. However, note that this research may not provide the true IGE because the results ignore self-employed individuals and people who are out of work. Moreover, this paper discovered that scale mismeasurement is likely the primary cause of the lifecycle bias. Furthermore, parent's education and parent's employment could be good proxies for parents' lifetime income.

This paper consists of chapter 1, which shows the introduction and literature review. Chapter 2 discusses data and related theories of intergenerational income elasticity. Next, chapter 3 explains the methodology used and talks about lifecycle bias and measurement error issues, followed by chapter 4 which presents the results using the OLS approach, rank-based estimation and IV approach. Finally, chapter 5 discusses limitations of this research and draws conclusion.

1.2: Literature Review

1.2.1: Lifecycle Bias and Attenuation Bias

As there are many research papers on intergenerational income mobility in the US, this literature review will focus more on USA's intergenerational income elasticity of father-son correlation. The literature review focuses on reviewing and evaluating the 'Intergenerational Income Mobility in the United States' (1992) by Solon Gary with respect to its methodology and main findings. This section mainly discusses attenuation bias, lifecycle bias and homogenous samples. Chapter 3 will further explain these biases in more details.

Earlier studies, for instance, Behrman and Taubman (1985) obtained an intergenerational correlation of 0.2 between parents' income and sons' earnings in the US. However, Solon (1992) concluded that the intergenerational income elasticity between sons and fathers in the United States was 0.4 or even higher. This indicates that earlier studies exaggerate the extent of intergenerational income mobility in the United States. Solon (1992) argued that earlier studies were heavily downward biased by measurement error and unrepresentative samples. The author used intergenerational data from the Panel Study of Income Dynamics (PSID).

Solon (1992) showed that attenuation bias in intergenerational income elasticity is caused by measurement error generated by transitory income fluctuations. Using a short-run proxy for long-run status will lead to measurement error because of transitory fluctuations in current income. Noise in current income, which is the variance of measurement error in income, produces an attenuation bias. Researchers should address this problem or else, IGE will be downward biased if attenuation bias exists. This is because measurement error causes the variance of current status to exceed the variance of permanent status. Due to transitory fluctuations and measurement error, using only a year of income contains a significant amount of noise (Mazumder, 2018).

The most common approach to solve this problem would be using a multi-year average of parental income (Hauser and Sewell, 1975). An alternative way to reduce or eliminate errors-in-variables attenuation bias is to use an instrument variables (IV) approach to predict parents' income based on parental characteristics (Solon, 1992 and Dearden *et al.*, 1997). According to Piraino (2007), "The idea is that the instruments will possibly suffer less from transitory variation than the single-year measures of income, thus representing a better proxy for long-run economic status". Both methods have been widely used by researchers.

Another substantive measurement issue that was not mentioned much by Solon (1992) but highlighted in many intergenerational mobility papers is lifecycle bias. As the correlation between current income and lifetime income varies across the life cycle, it indicates that producing consistent estimates of the IGE requires taking into account both parents and sons age. Haider and Solon (2006) stated that if earnings are measured too early in the life cycle, the gap between low and high lifetime earners will be underestimated compared to what it will be in midlife, and this will therefore understate the true IGE. For example, Reville (1995, cited in Haider and Solon, 2006) showed the estimated IGE was around 0.25 when the sons' earnings are observed in their twenties, but the estimates begin to approach 0.5 in their thirties. Haider and Solon (2006) suggested using an income variable around mid-life for both children's and parents' income to mitigate the lifecycle bias because it produced more precise estimates of changes over time. However, Nybom and Stuhler (2016) argued that "The validity of this approach rests on an assumption that does not hold in Swedish data."

The OLS results in Solon's article showed that the elasticity between sons' earnings and fathers' earnings of five-year averages was around 0.4 or higher, which indicated less mobility than earlier research. This seemed to be consistent with Zimmerman's (1992) study. On the other hand, Mazumder (2005) and Haider and Solon (2006) stated that even estimates based on an average fathers' income across five years has some attenuation bias. Mazumder (2005) reported that the estimated of IGE is biased down by approximately 30% based on father's five-year average income. Mazumder (2005) proposed using average income data of 20 to 25 years because lengthening the fathers' earnings window obtained a high reliability rate (almost no attenuation bias). For instance, Mazumder (2015) found that the estimated IGE in family income was greater than 0.6 in the United States by using up to 15-year averages of fathers' earnings. Similarly, Aakvik *et al.* (2012) found that compared to five-year averages, the estimates of IGE based on 15-year averages increased by around 20 to 30 percent, for both sons and daughters. Therefore, if researchers want to use multi-year average of parental income to mitigate the attenuation bias, it is suggested to use an average income data of 20 to 25 years instead of 5 years.

Solon (1992) has provided detailed and well analyzed insight on discussing whether father's years of education will be a valid instrument. Examples of instruments used in other literature include employment status, education and housing status (Blanden and Machin, 2008). Solon (1992) used fathers' years of schooling as a proxy for fathers' earnings, and found that the estimated IGE was 0.53. He emphasized that the IV estimates would be upward biased and may be arguably served as an upward inconsistent if father's education has a positive and independent effect on sons' economic status. Solon (1992) claimed that OLS and IV results produce the true value of intergenerational income elasticity.

Despite this, using father's education as an instrument for father's permanent income may result in upward inconsistency and produce upward bias for IV estimator. Mazumder (2005) argued that 0.53 in the article could potentially be a consistent estimate for the IGE and it could even be downward biased due to lifecycle effects. Nybom and Stuhler (2016) showed that the IV estimates in Sweden do not serve as an upper bound of the true IGE because it suffers from larger lifecycle effects than OLS estimates.

Grawe (2006) stated that fathers' earnings observed in their forties and sons in their late-twenties to mid-thirties were most precise. Haider and Solon (2006) suggested that sons' earnings were measured between the early thirties and mid-forties. Nonetheless, Rohenkohl (2019) has shown that income observed in the mid-thirties and forties would be a more accurate indicator of lifetime income.

Due to lifecycle bias, studies with young samples will have a lower IGE. In Solon's work, the sample mean for sons in 1984 was 29.6, while the sample mean for fathers in 1967 was 42. This indicated that Solon's work may contain lifecycle bias because he measured the sons' earnings at an earlier stage of the life cycle. Therefore, although an instrument variable being introduced in this article, the IV estimates were likely to be downward biased because of lifecycle bias. It can be concluded that a greater attenuation bias results from increasing transitory earnings variance over time.

1.2.2: Homogenous Sample

The second major concern of the author was unrepresentatively homogenous samples. Solon (1992) pointed out that homogenous samples would have lower variance in permanent income compared to population samples. In other words, the sample variance of income in effect understated the population variance. Some earlier intergenerational studies that relied on homogenous samples tend to underestimate the IGE. Examples of homogeneous samples could be samples collected from a specific city or region.

Even though homogeneous selection was on fathers, there would not be any problem in making a consistent estimation of the intergenerational income correlation if we could observe permanent income of the father. In contrast, Goldberger (1981) and Chung and Goldberger (1984) stated that if homogeneous selection was on sons, the estimate of IGE was inconsistent and likely to be downward biased. The homogenous fathers' sample will exacerbate the attenuation bias if we do not observe the fathers' permanent income (Solon, 1992).

Thus, researchers should consider avoiding samples that are homogenous when collecting data because it will lower the variation in permanent status and cause a greater downward bias. Therefore, homogenous samples aggravate the attenuation bias and lead to even greater downward bias.

Chapter 2: Data and Relevant Theories

2.1: Data and Sample restrictions

This research uses data from the British Cohort Study 1970 (BCS70), which follows the lives of 17,000 people born in the UK in 1970. As mentioned before, earlier studies in this field emphasized the necessity of observing parents and sons at the same stage of the lifecycle. Since all sons are observed in the same age group in BCS70, I therefore include parental age and age squared because parents are of different ages. Thus, this study uses the average of father age and mother age as parental age and restricts parental age between age 25 and age 60. Note that controlling the average age of father and mother will limit the sample to two-parent families.

Previous literatures mainly focused on intergenerational mobility between fathers and sons. Blanden *et al.* (2004) suggested that the changing influence of mothers' earnings partially explains the reduces the rise in IGE by a small amount. So, this research uses parental income to accommodate the importance of mothers' earnings. The reason why this paper only focuses on sons' earnings is because many daughters' earnings are missing compared to sons' earnings, and it is easier to make comparison with existing literatures. All earnings variables are in British pounds and deflated using the Consumer Price Index of every observed year (base year is 2015). They also measured in weakly gross pay in this study.

The dependent variable is sons' earnings, which are observed in five periods (age 26, 30, 34, 38 and 42). Sons' earnings data are collected from BCS70, where sons are asked to provide information on their gross pay (amount paid before deductions). Gregg *et al.* (2017) claimed that periods out of work must be accounted for when considering child's lifetime earnings because the IGE estimates tend to be greater when workless periods are included. Thus, if sons are out of work for over three periods, I remove the samples because their earnings do not represent accurate lifetime earnings. Since self-employment earnings data are all missing, samples in this study excludes self-employed sons, and only includes sons who are employed.

For parental income, it is measured from the combined gross income of parents when sons are aged 16, where BCS70 required parents to place their income within the appropriate band (there was eleven earnings bands). Then, this study takes the midpoint of each band as parental income. To reduce the effects of life-cycle bias that will affect the IGE estimates, this study restricts the sample to include only observations of parents' income when they are aged 30 to 60. Hence, it is crucial to note that after these sample restrictions, the sample sizes fall between 200 and 500 observations, which is a relatively small sample.

Both parent's employment and parent's education are observed when child aged 16. For parent's employment, I combine 'present employment situation of father' and 'present employment situation of mother' into a dummy variable.

Parent's employment	=0 if both parents 'Unemp and seeking work', 'Sick, will seek work', 'Looking after home', 'Permanently sick', 'Full-time student' and 'Retired'.
	=1 if one of them has 'Regularly employed', 'Casual/occasional week' and 'Other employment sit'.

=2 if both have ‘Regularly employed’, ‘Casual/occasional week’ and ‘Other employment sit’.

For parent’s education, parents are asked about ‘Holder of Degree, or Diploma, or Membership of Professional Institute (e g BSc, Bed, PhD, HND HNC, FAC FRICS, MIEE)’, ‘Other Qualification(s)’ and ‘No qualification(s)’. I first turn them into dummy variable:

(a) ‘Holder of Degree, or Diploma, or Membership of Professional Institute (e g BSc, Bed, PhD, HND HNC, FAC FRICS, MIEE)’	=1 if one of them has =2 if both have
(b) ‘Other Qualification(s)’	=1 if one of them has =2 if both have
(c) ‘No qualification(s)’	=1 if one of them has =2 if both have

Then, I combine three of them into one dummy variable, which is called parent’s education:

Parent’s education =0 if (c)==2

=0.5 if (c)==1

=1 if (a)==1 *or* (b)==1

=2 if (a)==2 *or* (b)==2 *or* (a)==1&(b)==1

2.2: Becker and Tomes model

The intergenerational income elasticity (IGE) is predicted to be positive, with an estimated value between zero and one. The intergenerational elasticity was explained by using Becker and Tomes (1986) model. This model shows that parents care about the welfare of their children, and they make investments. It predicts intergenerational coefficient of between zero and one. For example, assuming a parent must decide how much to invest in human capital of a child. A child will receive more as parents become richer. Becker and Tomes (1986) claimed that the more you invest in a person, the lower the marginal rates of return because investment costs eventually increase as his forgone earnings rise. Thus, it is arguably to say that what stops these coefficients and drives them down is because of decreasing returns to investment in the human capital of the child.

Following O'Neill *et al.* (2007) and Lillard and Reville (1999) which discussed Becker and Tomes (1986) model, where parents spend money on their own consumption and invest in their child's human capital. Equation (1) is a regression of a son's log earnings (y_t^*) on parental log earnings (y_{t-1}^*) and endowments (b_t). Equation (2) shows the correlation between a son's endowment (b_t) and parent's endowment (b_{t-1}). In the model, parents are altruistic, and they want to maximize utility or welfare of their children without reducing their own utility. If credit constraints do not exist, the human capital and earnings of children will only depend on the inheritability of endowments γ , and do not depend on parents' human capital. In other words, every child will receive the optimal level of human capital, which depends on b_t, so $\beta = 0$ in equation (1).

$$y_t^* = \beta\, y_{t-1}^* + \gamma\, b_t + e_t^y \tag{1}$$

$$b_t = \rho b_{t-1} + e_t^b \tag{2}$$

where e_t^y and e_t^b are IID random variables.

In contrast, poor parents fail to finance the optimal level of human capital if credit constraints exist. Then, the constraint weakens as income increases when b_t is constant, so only wealthy parents can invest more in their children and $\beta > 0$. Hence, the mobility of wealthy parents is greater because they can invest effectively in their children and better-endowed children would have higher expected earnings.

Chapter 3: Methodology

3.1: Ordinary Least Squares (OLS) Approach

Intergenerational income elasticity measures the relationship between the lifetime incomes of parents and sons across generations. All income variables are measured in log. Equation (3) shows the OLS regression of sons' lifetime earnings (y_i^{son*}) on parents' lifetime earnings ($y_i^{parent*}$), where β is the estimated intergenerational income elasticity (IGE).

$$y_i^{son*} = \alpha + \beta y_i^{parent*} + \varepsilon_i \tag{3}$$

where i denotes the family and ε_i is an error term which is assumed to be uncorrelated with parents' lifetime income $cov(y_i^{parent*}, \varepsilon_i) = 0$. Due to data limitations, it is a major challenge for researchers to estimate both parents' and sons' lifetime income.

Proxies for y_i^{son*} and $y_i^{parent*}$ are sons' current earnings and parents' current income. Equation (4) shows that sons' current earnings (y_{it}^{son}) proxy for sons' lifetime earnings in period t.

$$y_{it}^{son} = y_i^{son*} + u_{it} \tag{4}$$

where the error term, u_{it} represents the transitory fluctuation around lifetime income due to measurement error and short-term transitory variations. We assume that u_{it} and v_{it} are not correlated with each other and are also uncorrelated with y_{it}^{son} and y_{it}^{parent}.

Equation (5) shows the parents' current income proxy for parents' lifetime income. Its structure is the same as equation (4),

$$y_{it}^{parent} = y_i^{parent*} + v_{it} \tag{5}$$

where v_{it} is the error term.

3.1.1: Lifecycle Bias

Jenkins (1987) highlighted that earnings trajectories across the lifecycle have considerable heterogeneity, which differs by family background. Nybom and Stuhler (2016) also stated that heterogeneity in individuals' age-earnings profile is important. As discussed in section 1.2, true IGE estimates will be underestimated if earnings are observed too early in the lifecycle. The following part will follow what was done by Gregg *et al.* (2017) who discussed Solon's (1992) work. Even though lifecycle bias affects both parents and sons, the explanation below focuses only on sons for simplicity. λ_t in equation (6) is a measure of sons' earnings at a certain point in time, which differs from their lifetime earnings over the lifecycle.

$$y_{it}^{son} = \lambda_t y_i^{son*} + u_{it} \tag{6}$$

Assuming there is no error in parental income in equation (7),

$$y_{it}^{son} = \alpha + \beta y_i^{parent*} + e_{it} \tag{7}$$

So, note that $y_i^{parent*} = y_{it}^{parent}$ because no error in parental income. Then, equation (7) yields the probability limit of β:

$$plim\hat{\beta} = \frac{Cov(y_{it}^{son}, y_{it}^{parent})}{var(y_{it}^{parent})} = \frac{Cov(\lambda_t y_i^{son*} + u_{it}, y_{it}^{parent})}{var(y_{it}^{parent})}$$

$$= \frac{Cov(\lambda_t \beta y_i^{parent*} + u_{it}, y_{it}^{parent})}{var(y_{it}^{parent})}$$

$$= \frac{\lambda_t \beta \ var(y_{it}^{parent})}{var(y_{it}^{parent})} + \frac{var(u_{it}, y_{it}^{parent})}{var(y_{it}^{parent})}$$

$$= \lambda_t \ \beta + \frac{Corr(y_i^{parent*}, u_{it}) \ \sigma_{u_{it}}}{\sigma_{y^{parents*}}} \tag{8}$$

Böhlmark and Lindquist (2006) stated that λ_t in equation (8) refers to a population parameter that is correlated with the shape of age-earnings profiles, which changes across cohort, sex and nation. Allowing for heterogeneity in age-earnings profiles across the lifecycle generates lifecycle bias, for example, different types of workers have different shape of age-earnings profiles (Böhlmark and Lindquist, 2006). If $\lambda_t = 1$, then the measurement error is approximately classical and hence $\hat{\beta}$ will result in a consistent estimate of β if $Corr(y_i^{parent*}, u_{it}) = 0$. However, Nybom and Stuhler (2016) argued that $Corr(y_i^{parent*}, u_{it}) = 0$ from equation (8) usually does not hold because deviations from average income trajectories related with family and individual characteristics leads to inconsistent estimates of IGE when $\lambda_t = 1$.

Due to these issues, this study first focuses on estimating intergenerational income elasticity at different points over the lifecycle. Since lifecycle bias might arise from heterogeneity across lifecycle income profile and Böhlmark and Lindquist (2006) claimed that control for parental age will not remove lifecycle bias. Most estimates in this research still follows Haider and Solon (2006) and control for parental age and age squared.

3.1.2: Attenuation Bias (Measurement Error)

Another measurement issue is attenuation bias (measurement error), where only parental income variable will contain measurement error and includes transitory shocks at any point in time. It is crucial to note that measurement error in sons' earnings do not cause inconsistency in the model (Nybom and Sulther 2016; Haider and Solon, 2016). To show how parental income contain measurement error, the following section follows again the work of Gregg *et al.* (2017). In equation (9), assuming there is no measurement error in the sons' earnings variable,

$$y_i^{son*} = \alpha + \beta y_{it}^{parent} + e_{it} \tag{9}$$

where y_{it}^{parent} is parents' annual income.

Note that $y_{it}^{son} = y_i^{son*}$ because no error in sons' earnings variable. Then, the probability limit of $\hat{\beta}$ is:

$$plim\hat{\beta} = \frac{Cov\left(y_{it}^{parent}, y_i^{son*}\right)}{var\left(y_{it}^{parent}\right)} = \frac{Cov\left(y_{it}^{parent}, \alpha + \beta y_i^{parent*}\right)}{var\left(y_{it}^{parent}\right)}$$

$$= \frac{Cov\left(y_i^{parent*} + v_{it}, \alpha + \beta y_i^{parent*}\right)}{var\left(y_i^{parent*}\right) + var(v_{it})}$$

$$= \frac{\beta var\left(y_i^{parent*}\right)}{var\left(y_i^{parent*}\right) + var(v_{it})} = \beta \frac{\sigma^2_{y^{parent*}}}{\sigma^2_{y^{parent*}} + \sigma_v^2} \tag{10}$$

As shown in equation (10), the attenuation bias arises from the measurement error in parental income (σ_v^2) due to variance of measurement error in parental income. Shortly speaking, inaccurately measuring parents' permanent income causes attenuation bias. A constant attenuation bias implies that earnings persistence estimates will not change because of proportional growth in permanent and transitory earnings variance (Grawe, 2003). The attenuation bias depends on the variance of the measurement error in parental income, so the bias will increase when the variance of the measurement error in parental income increases. For instance, if parental income contains measurement error, noise will be generated on parental income (where the noise will not be correlated with children's income). Then, the coefficient as shown in equation (10) goes down because the denominator grows while the numerator remains the same. Thus, this suggests that noise in parents' current income cause parents' permanent income to be measured with error and produces attenuation bias that lower IGE.

The increase in transitory earnings variance leads to a greater attenuation bias and reduces persistence estimates. Grawe (2003) noted that transitory earnings variance is higher when both parents and sons are older in later periods. Hence, findings are inconsistent if attenuation bias arises because measurement error causes the estimated IGE to be different from the true IGE. The OLS estimates will be downward biased and served as a lower bound estimate of the actual IGE. Therefore, it is suggested to use multi-average income or IV approach to solve this measurement error problem.

3.2: Rank-based Estimation

Besides than OLS estimator, many previous studies (Rohenkohl, 2019; Gregg *et al.*, 2017) used the rank-based estimation, which is another approach to estimate intergenerational income elasticity. The rank-based estimation approach employs the regression of the rank of sons' earnings on the rank of parental income. It helps to remove the difference in variation between two measures. Section 4.2.3 will discuss more about rank-based estimates.

3.3: IV Approach (2SLS)

Researchers can use multi-average parental income or the IV approach to reduce attenuation bias as described in section 2. Due to data limitations, this research is only able to use IV approach. An appropriate instrumental variable should be correlated with the endogenous variables (relevance condition), but uncorrelated with the error term (exogeneity condition). In other words, the exogeneity condition states that an instrument can only affect dependent variable through endogenous variables. The IV estimates are not consistent if either one of the conditions fails to hold. Blanden *et al.* (2013) showed that social class is not an appropriate instrument for permanent income.

Estimate equation (11) to find the IV estimates, where y_i^{son*} is the multi-year average of sons' earnings. y_{it}^{parent} is the endogenous variable that shall be instrumented by parental characteristics because unable to observe parents' permanent income. The IV estimates in this research use parent's education and parent's employment as proxies for parents' lifetime income. Parents age and age squared are exogenous variables.

$$y_i^{son*} = \alpha + \beta y_{it}^{parent} + Parents\ age + \ Parents\ age^2 + \ \varepsilon_i \tag{11}$$

$$\widehat{y_{it}^{parent}} = \pi_0 + \pi_1\ parents'\ education + \ \pi_2\ parents'\ employement + u \tag{12}$$

where π shows the relationship between parental characteristics and parental income. In the first stage, we use equation (11) to run an OLS regression for y_{it}^{parent} on all instruments, all exogenous variables, and an intercept to get the fitted values. Then, the second stage is to regress y_{it}^{parent} on the predicted values of all endogenous variables, exogenous variables and an intercept using OLS. The IV estimation in this research is conducted using Stata (version 17).

To explain the consistency of IV estimates, this part will follow Solon's (1992) work. For simplicity, I add only parents' education (E_i) into equation (3) (it is the same for parents' employment):

$$y_i^{son*} = \beta_1 y_i^{parent*} + \beta_2 E_i + \varepsilon_i \tag{13}$$

Estimate equation (3) results in omitted variable bias if equation (13) shows the true relationship between sons' earnings and parental income, because equation (3) excludes parent's education. The relationship between β_1 and β_2 is:

$$\rho = \beta_1 + \frac{\beta_2\ Cov(E_i, y_i^{parent*})}{\sigma_y^2} = \frac{\beta_1 + \beta_2 \lambda \sigma_E}{\sigma_y} \tag{14}$$

where λ shows the correlation between E_i and $y_i^{parent*}$, and σ_E^2 represents the variance of E_i.

As using only one year of parental income causes attenuation bias. To solve the attenuation bias, we use parents' education as a proxy for parents' current income. Assuming that E_i is uncorrelated with v_{it} and u_{it} results in:

$$plim\ \hat{\rho}_{IV} = \frac{Cov(E_i, y_{it}^{son})}{Cov\left(E_i, y_{it}^{parent}\right)}$$

$$= \frac{Cov\left(E_i, \beta_1 y_{it}^{parent} + \ \beta_2 E_i + \ \varepsilon_i + u_{it} - \ \beta_1 v_{it}\right)}{Cov(E_i, y_{it}^{parent})}$$

$$= \frac{\beta_1 + \beta_2\sigma_E^2}{\lambda\sigma_E\sigma_y}$$

$$= \frac{\beta_1 + \beta_2\sigma_E}{\lambda\sigma_y}$$

$$= \frac{\beta_1 + \beta_2\lambda\sigma_E}{\sigma_y} + \beta_2(\frac{\sigma_E}{\lambda\sigma_y} - \frac{\lambda\sigma_E}{\sigma_y})$$

$$= \rho + \beta_2\sigma_E\left(\frac{1 - \lambda^2}{\lambda\sigma_y}\right) \tag{15}$$

Note that error in E_i does not affect the consistency of the estimates. Equation (15) proves that IV estimates are consistent only if the instrument variable, which is parent's education in this case is not correlated with sons' earnings because instruments cannot affect dependent variable directly ($\beta_2 = 0$) or parents' education and parents' permanent income have perfectly positive correlation ($\lambda = 1$). The problem that this research is facing is $\beta_2 > 0$, where parent's education and employment might be positively correlated with sons' earnings. If $\beta_2 > 0$, IV estimates will be upward biased, because it is reasonable to assume that sons who have high-educated parents are expecting to make more money than those who have low-educated parents. Similarly, for parent's employment, it is considerable to say that sons whose parents are upper-class workers earn more than those who are lower-class workers.

Hence, the expected result in this research is the IV estimates would provide as an upper bound of actual IGE because the parental characteristics used to predict parents' permanent income have a positive effect on sons' earnings in this research. Note that IV estimates will not provide as an upper bound of the actual IGE if it suffers from larger lifecycle effects than OLS estimates.

Chapter 4: Findings

4.1: Descriptive Statistics

Figure 1

	Mean	Standard deviation	Min	Max	N
Parental income (age 16)	554.764	357.234	52.995	1622.281	713
Sons’ earnings (age 26)	339.586	136.037	14.535	1453.488	927
Sons’ earnings (age 30)	896.715	4184.347	1.058	115342.8	1005
Sons’ earnings (age 34)	927.834	1105.111	50.277	18275.52	988
Sons’ earnings (age 38)	1026.765	1053.726	59.032	15643.45	835
Sons’ earnings (age 42)	1066.251	3786.491	1.161	117065.6	1029

Notes: Observations exclude full-time and part time self-employed sons and sons who are out of work for over 3 periods.

Figure 1 shows descriptive statistics for the main variables used in this research. All income variables are transformed from annual gross earnings to weekly gross earnings. To show the comparison between sons’ earnings and parental income, income variables are adjusted for CPI. The mean shows that sons’ earnings at age 26 have lower earnings compared to parental income observed when their sons are aged 16. This is because sons’ earnings are observed at an earlier stage of the lifecycle. This suggests that sons at a young age earned less than their parents. On average, sons at age 30, 34, 38 and 42 have higher earnings than their parents, which means that individuals’ living conditions have improved across generations in absolute terms. The mean shows that sons’ earnings increase considerably by 557.129 from age 26 to 30 and starts to rise steadily at age 30 to 42.

4.2.1: OLS Estimates

Figure 2

	(1) OLS	(2) OLS	(3) Rank-based	(4) OLS (5-95% winsorization)	N	(5) OLS	(6) Rank-based	N
Age 26	0.200***	0.201***	0.344***	0.180***	388	0.241***	0.319***	356
	(0.037)	(0.038)	(0.049)	(0.026)		(0.048)	(0.051)	
Age 30	0.278***	0.277***	0.349***	0.250***	417	0.344***	0.330***	385
	(0.042)	(0.043)	(0.046)	(0.035)		(0.052)	(0.048)	
Age 34	0.266***	0.264***	0.356***	0.267***	413	0.303***	0.312***	378
	(0.043)	(0.044)	(0.046)	(0.036)		(0.059)	(0.049)	
Age 38	0.326***	0.312***	0.402***	0.344***	356	0.382***	0.371***	323
	(0.046)	(0.048)	(0.048)	(0.042)		(0.055)	(0.051)	
Age 42	0.358***	0.362***	0.386***	0.334***	427	0.458***	0.359***	390
	(0.047)	(0.049)	(0.044)	(0.040)		(0.070)	(0.046)	
Average	0.272***	0.273***	0.345***	0.263***	271	0.321***	0.300***	249
	(0.043)	(0.044)	(0.047)	(0.038)		(0.057)	(0.049)	

Notes: Column (1) presents the OLS estimates with no controls, whereas column (2) controls for the age and age^2 of parents at the time that income is observed. Column (3) to (6) all control for the age and age^2 of parents. Column (3) presents the rank-based estimates. Column (4) shows OLS estimates after 5-95% winsorization. Column (5) and (6) presents OLS estimates and rank-based estimates of sons' earnings regress on two-period averages of parental income (age 10 and 16). Robust standard errors are shown in parenthesis.

***= significant at 1% where P-value < 0.01

**= significant at 5% where P-value < 0.05

*= significant at 10% where P-value < 0.10

First, we use the OLS approach to estimate intergenerational income elasticity at different points across son's lifecycle, where the independent variable is parental income at age 16. Note that the estimates from column (1) to (5) suffer from attenuation bias, because only using a single year of parental income as the dependent variable. The point of doing this is to focus on looking at how estimates change over time and the difference between OLS estimation and rank-based estimation.

The equation for point estimate is $y_{it}^{son} = \alpha + \beta y_{it}^{parent} + e_{it}$. Column (1) shows OLS estimates with no controls. Column (2) presents OLS estimates controlling for parental age and age squared. IGE estimates are slightly lower in column (2) compared to column (1), except for estimates at age 26 and 42. However, there is only a difference of 0.001 at age 26 and 0.004 at age 42, which implies only a little change in intergenerational income mobility when include parents age and age squared.

Blanden *et al*. (2004) stated that there was a decline in mobility due to an increase in income inequality in the UK since the late 1970s. In column (2), IGE starts at a very low estimate of 0.201 at age 26 and increases significantly to 0.277 at age 30, indicating a dramatic decrease in the UK intergenerational income mobility. Gregg *et al.* (2017) found that IGE increased slightly starting from age 30, suggesting that income mobility declined overtime, or that the persistence of intergenerational inequality increased overtime. However, column (2) shows that IGE decreases from age 30 to age 34, but there is not much difference. The IGE then increases to 0.312 at age 38 and peaks at 0.362 at age 42. The IGE at age 42 is very close to Belfield *et al.* (2017) results, who used the same data and procedure as this research, but controls for a quadratic of fathers' age found the IGE was 0.358 at age 42.

The accuracy of average annual income as a representation of lifetime income relies on the ages at which annual income is measured. To reduce the transitory fluctuations of earnings and the impact of very low-earnings periods, it is suggested to use a multi-year average of sons' current income to reduce lifecycle bias (Nybom and Stuhler, 2016). Hence, to find the UK's intergenerational income elasticity, this research will use equation (9) where it regresses the average of sons' earnings across five periods on parental income at age 16. So, as shown in columns (1) and (2), OLS estimate with no additional controls is 0.272, while OLS estimate increases to 0.273 when controls for parents' age and age squared. The difference between them is extremely small, which is consistent with what Rohenkohl (2019) has discovered.

4.2.2: Winsorization

Few literatures pointed out that IGE estimates can be sensitive to the treatment of outliers and missing values (Dahl and DeLeire, 2008; Nybom and Suthler, 2016). Winsorization limits extreme values in the data to diminish the impact of spurious outliers. Rohenkohl (2019) found that winsorizing has a little impact on increasing the IGE estimates, indicating that part of the persistence in IGE estimates is due to extreme values of the distribution, while the rank-based estimates remain relatively stable to winsorization. The results for winsorization are also shown in Figure 2.

Column (4) shows the OLS estimates after 5-95% winsorization. With winsorization, IGE rises sharply from 0.180 at age 26 to 0.250 at age 30, and it continues to increase gradually. The sharp increases from age 26 to 30 might imply that there is lifecycle bias at sons' early ages. Comparing age 30 to 34 in column (4) and column (2), IGE decreases without winsorized data, while IGE increases with winsorized data. This implies that since parental income is the same over the lifecycle, it is arguably to say that the increases of IGE from 0.250 to 0.267 in column (4) is due to winsorization that removes the left tail of sons' earnings. In other words, some lowest sons' earnings are removed with winsorizing, so the IGE rises to 0.267. Even though IGE decreases from 0.344 at age 38 to 0.334 at age 42, this will not affect the intergenerational income mobility too much because there is no significant change between them.

Therefore, the OLS estimates of five-period averages of sons' earnings on parents' annual income after winsorization is 0.263, which shows only a 0.10 drop compared to OLS estimates without winsorizing. Overall, IGE estimates move closer once data are winsorized because winsorization removes some outliers.

4.2.3: Rank-based Estimates

Gregg *et al*. (2017) argued that the inaccuracy in the distance between sons' earnings (scale mis-measurement issues) drives the lifecycle bias, instead of the inaccuracy of rank ordering (positional accuracy issues). Positional inaccuracy refers to incorrectly ordering the individuals. Taking lifecycle bias as an example, if a survey is conducted before an individual realized the complete returns to their education, sons' earnings will be placed in a lower distribution than in later periods when sons' earnings have matured (Gregg *et al*., 2017). Conversely, scale mismeasurement refers to incorrectly measuring the difference between individuals' earnings. For instance, the scale of earnings gaps between low-educated individuals and high-educated individuals will be understated.

In this paper, earnings variables are used in levels (not log) for rank estimates. To look at the difference between OLS estimates and rank-based estimates, this part mainly focuses on discussing column (2) and (3). Since rank-based estimates remove the scale mis-measurement issues, it makes sense that rank-based estimates in column (3) are much higher than OLS estimates in column (2). At early ages, there is a large difference between the OLS estimates and rank-based estimates. Then, starting at age 34, the difference slowly narrows as sons are getting older. These reflect the argument of Gregg *et al.* (2017) who stated that OLS estimates are subject to more lifecycle bias than rank-based estimates, especially at a young age.

For point estimates, OLS estimate rises sharply at age 26, while rank-based estimate does not. Since rank-based estimation removes scale mismeasurement, it is reasonable to conclude that scale mismeasurement causes the majority of lifecycle bias in OLS estimates, instead of positional inaccuracy. Even though the robust standard errors in the OLS estimates in column (2) are slightly smaller than rank-based estimates in column (3), but OLS estimates suggest being downward biased due to life-cycle effects (larger lifecycle bias compared to rank-based estimates) because sons' earnings are observed at young age. Therefore, it would seem that rank-based estimates are not subject to much lifecycle bias as compared to OLS estimates.

As shown in column (3), the rank-based estimate increases relatively stable from age 30 to 38. There is a slight decrease from 0.402 at age 38 to 0.386 at age 42, which does not cause any significant difference at all. The rank-based estimates of five-period averages of sons' earnings on parental income at age 16 is 0.345, which indicates that an increase of 10 percentile point in the rank of parental income would result in a 3.45 percentile point increase in the rank of sons' earnings. Overall, rank-based estimates are more stable across the lifecycle compared to OLS estimates.

Next, Gregg *et al.* (2017) who used two different UK datasets (NCDS and BCS70), found that the age-earnings trajectories for both datasets are identical across cohorts in the UK. However, Jenkins (2009) found large differences in individual age-earnings trajectories from average trajectories of groups defined by education, gender, and cohort in UK data. Therefore, if Jenkins

(2009) was correct, it can be argued that not only heterogeneity in age-earnings profiles causes lifecycle bias in estimating UK's IGE, but also scale mismeasurement.

4.2.4: Averaging parental income for two periods

In figure 2, column (5) and (6) show a regression of sons' earnings at different points across the lifecycle on the average of parental income across two periods (age 10 and 16). The way this research measures parental income at age 10 is the same as how parental income are measured at age 16, where midpoint of each band is taken. The sample size slightly reduced because not all parents provided their income in both periods.

The pattern across lifecycle for OLS estimates in column (5) is similar to column (2). Again, this suggests that lifecycle bias in OLS estimates at early age are subject to more lifecycle bias, because the coefficient at age 26 is relatively small compared to rank-based estimates. Comparing the OLS estimates in column (5) and column (2), averaging parental income increases IGE estimate from 0.273 to 0.321. This could suggest that single-year parental income suffers more transitory income variance (larger attenuation bias) than two-year averages of parental income, so OLS estimates in column (2) are lower because it is underestimated. However, as mentioned in the literature review, using average of parental income across two periods will not entirely mitigate attenuation bias.

Therefore, the IGE in OLS estimates suffer from larger attenuation bias when using only a single year of parental income. This result is consistent with Gregg *et al.* (2017), where it showed OLS estimates with average parental income is higher than using only a single year of parental income.

4.3: Average sons' earnings at different periods

Figure 3

	(1)	(2)	(3)
OLS	0.239***	0.283***	0.323***
	(0.039)	(0.045)	(0.047)
Rank-based	0.343***	0.355***	0.380***
	(0.045)	(0.046)	(0.047)
N	339	312	322

Fig 3 Notes: All columns control for parents age and age^2. Dependent variable in column (1) is the average of sons' earnings at age 26, 30 and 34, while column (2) takes the average sons' earnings at age 30, 34 and 38. Column (3) uses the average of sons' earnings at age 34, 38 and 42. Robust standard errors are shown in parenthesis.
***= significant at 1% where P-value < 0.01
**= significant at 5% where P-value < 0.05
*= significant at 10% where P-value < 0.10

Figure 3 is estimating $y_i^{son*} = \alpha + \beta y_{it}^{parent} + e_{it}$, where the independent variable is parental income at age 16 for all columns. Note that, attenuation bias still exists because of single-year parental income. OLS estimate is 0.239 in column (1) and 0.323 in column (3). Comparing column (1) and column (3), column (1) shows that taking the average of sons' earnings at early ages results in a much lower IGE. These results reflect the fact that lifecycle bias would attenuate the IGE. Both OLS estimates and rank-based estimates increase with age, and rank-based estimates are relatively stable. Therefore, these estimates show that taking different points of sons' earnings as sons' lifetime earnings will change both OLS and rank-based estimates.

4.4:

Add Additional Controls

Figure 4

	(1) OLS estimates	(2) OLS estimate
Parental income	0.241***	0.209***
	(0.048)	(0.045)
Parents age	-0.066	-0.061
	(0.059)	(0.059)
Parents age squared	0.001	0.001
	(0.001)	(0.001)
Application	-0.002	0.001
	(0.005)	(0.005)
Clumsiness	-0.013**	-0.014***
	(0.006)	(0.005)
Extroversion	-0.006	-0.008
	(0.007)	(0.006)
Hyperactivity	0.005	0.008**
	(0.004)	(0.004)
Anxious (age 10)	0.002	-0.003
	(0.006)	(0.006)
Reading	0.002	-0.002
	(0.004)	(0.003)
Maths	0.007**	0.005
	(0.003)	(0.003)
Anxious (age 16)	0.086	0.193
	(0.121)	(0.131)
Bad GCSEs		0.220**
		(0.090)
CSE 2-5, other Scottish school qualification		0.077
		(0.100)
O levels, Good GCSEs		0.141*
		(0.076)
1 A-level or more than 1 AS-level		0.564***
		(0.125)
2 or more A-levels		0.079
		(0.128)
Diploma of HE		0.216*
		(0.115)
Degree, other degree level		0.448***
		(0.088)
Higher degree		0.410***
		(0.116)
N	224	224

Fig 4 Notes: The dependent variable in this figure is the average of sons' earnings across five periods (age 26, 30, 34, 38, 42) and independent variable is parental income at age 16. Both columns control for parents age and age squared. Column (1) presents the OLS estimates, controls for son's non-cognitive skills at age 10 and 16 and son's cognitive skills at age 10. Column (2) adds control of son's highest academic level at age 30. Robust standard errors are shown in parenthesis.

***= significant at 1% where P-value < 0.01

**= significant at 5% where P-value < 0.05

*= significant at 10% where P-value < 0.10

$$y_i^{son*} = \beta_0 + \beta_1 y_{it}^{parent} + \beta_2 non - cognitive\ skills + \beta_3 cognitive\ skills + e_{it} \quad (16)$$

$$y_i^{son*} = \beta_0 + \beta_1 y_{it}^{parent} + \beta_2 non - cognitive\ skills + \beta_3 cognitive\ skills + \beta_4 highest\ academic\ level + e_{it} \quad (17)$$

It is interesting to observe how IGE estimates will change when adding controls of son's non-cognitive skills, cognitive skills, and highest academic level (these variables are in linear form).[1] Variables for son's non-cognitive skills, including *application, clumsiness, extroversion* and *hyperactivity* at age 10; *anxious* at age 10 and 16, while cognitive skills include *maths* and *reading* at age 10. Son's highest academic level contains nine dummy variables.

At here, the independent variable is the five-period averages of sons' earnings. Column (1) estimates equation (16), while column (2) estimates equation (17). The coefficient interpretation of log-linear regression is $100 * \left[exp(\widehat{\beta_1}) - 1\right]$. Column (1) shows the coefficient of *clumsiness* and *maths* are statistically significant at 5% significance level. It indicates that one unit increase in *clumsiness* would decrease sons' earnings by 1.305%, while one unit increase in *maths* score would increase sons' earnings by only 0.724%. Both variables have very small impact on sons' earnings and the IGE drops from 0.273 (in figure 2, column 2) to 0.241.

[1] See Appendix A for more details about son's non-cognitive, cognitive skills and highest academic level.

After adding son's highest academic level as shown in column (2), the coefficient of *clumsiness* is statistically significant at the 1% significance level, interpreting that one unit increase in *clumsiness* would decrease sons' earnings by 1.42%. Then, the coefficient of *hyperactivity* is now statistically significant at 5% significance level, which indicates that one unit increase in *hyperactivity* would raise sons' earnings by 0.82%. For son's highest academic level, the coefficient of sons with *1 A-level or more than 1 AS-level*, *Degree* and *Higher Degree* are statistically significant at 5% significance level. It indicates that sons with *1 A-level or more than 1 AS-level* earn more than a son without it by 75.7%. Moreover, sons with *Degree* earn more than a son without *Degree* by 56.5%, while sons with a *Higher Degree* earns 50.74% more than those without a *Higher Degree*. Interestingly, the coefficient of *Bad GCSEs*, 0.220 is significant at 5% significance level and has a small positive effect on sons' earnings. However, this coefficient might not be accurate because there are only three individuals in the sample who have *Bad GCSEs*. Other than that, the coefficient of *O levels* and *Diploma of HE* is statistically significant at 10% significance level. Obviously, the IGE slightly decreases from 0.241 to 0.209 after adding son's qualifications.

Therefore, the IGE decreases from 0.273 (in figure 2, column 2) to 0.241 due to the significance of son's *clumsiness* and *maths and* diminishes further to 0.209 as the significance of son's *clumsiness*, *hyperactivity* and *highest academic levels.* Note that both 0.241 and 0.209 are statistically significant at 1% significance level. The coefficient of 0.209 implies that one unit increase in parental income would increase 2.09% in sons' earnings as an adult on average. Hence, intergenerational income mobility in the UK increases with the introduce of son's non-cognitive skills, cognitive skills and qualifications, indicating that sons' earnings depend less on their parental income compared to IGE with only parental age as controls.

4.5: IV Estimates

Figure 5

	(1)	N
Average	0.506***	137
	(0.105)	

Notes: Robust standard errors are shown in parenthesis.
***= significant at 1% where P-value < 0.01
**= significant at 5% where P-value < 0.05
*= significant at 10% where P-value < 0.10

Due to data limitations, we are unable to observe parents' lifetime income, and use only one year of parental income as a proxy for lifetime income, which causes attenuation bias. So, I propose an IV approach to estimate IGE. To make it easier to make comparison between OLS results and IV results, IV estimates will control for parents age and age squared only. For IV estimates, the independent variable is the average of sons' earnings across five periods and the dependent variable, parent's permanent income is instrumented by parent's employment and parent's education[2]. The IV estimate is 0.506, suggesting that a 10% increase in parental income, increase 5.06% in sons' earnings as an adult. However, as mentioned at section 3.3, parents' education and employment are very likely to be positively correlated with sons' earnings, so the IV estimate is upward bias.

Chapter 5: Conclusion

The findings in this research are subject to three limitations. Firstly, the income range for '£500 and over' is not clear enough to calculate the exact parental income at age 16. It would be better to have a range of '£500 and £1000' or '£500 and £1500', instead of using '£500 and over'. As

[2] See Appendix B for details about dummies.

this research uses the midpoint estimate of ranges, setting the highest point for the income range is crucial because it could affect the midpoint estimate, which will directly affect the IGE results. This is because the higher the mid-point for parental income, the lower the IGE estimates.

Secondly, this study restricts samples to two-parent families only, meaning that the influence of one-parent families has not been considered. Not only that, this research only estimates the intergenerational income elasticity between parents and sons, and not take into account the intergenerational income elasticity between parents and daughters.

Lastly, is the problem of workless individuals and self-employed individuals. This research ignores individuals who being unemployed for out of 3 periods and self-employed individuals. Gregg *et al.* (2017) found that IGE was understated by around 0.05 due to the exclusion of unemployed individuals. Hence, the measurement issues include (1) how to treat periods out of work, and (2) how to deal with zero earnings (self-employed individuals) when considering lifetime earnings. Clearly, there could be other factors such as luck and a child's personality that would affect the intergenerational income elasticity indirectly.

Rohenkohl (2019) used the OLS estimator and found that the IGE in the UK was between 0.25 and 0.27. Blanden *et al.* (2004) used BCS70 cohort and found that IGE was around 0.30 for men. Gregg *et al*. (2017) found an even higher estimate of 0.43 for the BCS70. The likely reason why the upper bound (0.50) in my research is higher than Gregg *et al*.'s (2017) findings is because I control parental age to reduce lifecycle bias, and I use IV estimates to reduce attenuation bias, whereas Gregg *et al*. (2017) only used average parental income for only two periods (suffer from attenuation bias).

To summarize, there are many factors to consider when measuring a country's intergenerational income mobility. Sample selection, lifecycle bias and attenuation bias are the main problems in estimating intergenerational income elasticity. Section 4.2.2 concludes that IGE estimates are closer after winsorizing because winsorization reduces the magnitude of some outliers. By comparing OLS estimates and rank-based estimates, the results in section 4.2.3 suggest that scale mismeasurement causes the lifecycle bias rather than positional inaccuracy. However, heterogeneity in income profiles may affect OLS estimates. Hence, further research should find out if there is heterogeneity in income profiles in the UK data. Moreover, section 4.2.4 proves that using single-year parental income has a lower IGE compared to two-year parental income but note that using two-year parental income does not mean attenuation bias is fully mitigated. Section 4.3 shows that taking the average of sons' earnings at different periods matters when estimating IGE across generations. Then, results from section 4.4 imply that there is high intergenerational income mobility when controls for son's non-cognitive skills, cognitive skills, and qualifications compared to results that without these controls.

In conclusion, the findings suggest that lifecycle bias still exist in point-in-time estimates of intergenerational income elasticity. The IV estimate of 0.506 could serve as an upper bound of the true IGE because instruments that used in this research are very likely to be positively correlated with parental income. Meanwhile, the OLS estimate of 0.273 provides a lower bound of the true IGE because the OLS estimates do not take into account the measurement error in parental current income. To sum up, the findings reveal that the intergenerational income mobility in the UK falls between 0.27 and 0.50. This result indicates that for every additional 10% of parental income advantage, only between 2.7% and 5.0% will be passed on to the next generation in the UK. Nonetheless, both OLS and IV estimates in this research still suffer from workless spells bias and using multi-average of sons' earnings will not totally mitigate lifecycle bias. Hence, it is concluded that the intergenerational income elasticity of 0.27-0.50 still understates the true IGE, meaning that sons' earnings will be less dependent on parental income.

6: Bibliography and Appendices

6.1: Bibliography

Aakvik, A., Jacobsen, K, Nilsen, Ø. A. and Vaage, K., (2012). 'Intergenerational Earnings Mobility Revisited: Estimates Based on Lifetime Earnings', *The Scandinavian Journal of Economics*, 114(1), pp. 1–23. Available at: http://www.jstor.org/stable/41407774 (Accessed: 1 December 2022)

Becker, G. S. and Tomes, N. (1986). 'Human Capital and the Rise and Fall of Families', *Journal of Labor Economics,* 4(3), pp. S1-S39. Available at: https://www.jstor.org/stable/2534952 (Accessed: 12 April 2023)

Behrman, J. and Taubman, P. (1985). 'Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model', *The*

Review of Economics and Statistics, 67(1), pp. 144–151. Available at: https://doi.org/10.2307/1928446 (Accessed: 24 October 2022)

Belfield, C., Crawford, C., Greaves, E., Gregg, P. and Macmillan, L. (2017). *Intergenerational income persistence within families*, London: The IFS. Available at: https://ifs.org.uk/publications/intergenerational-income-persistence-within-families (Accessed: 1 October 2022)

Björklund, A., Bratsberg, B., Eriksson, T., Jäntti, M., Naylor, R., Österbacka, E., Raaum, O. and Røed, K. (2006). *American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States.* IZA DP No. 1938. Available at: https://www.iza.org/publications/dp/1938/american-exceptionalism-in-a-new-light-a-comparison-of-intergenerational-earnings-mobility-in-the-nordic-countries-the-united-kingdom-and-the-united-states (Accessed: 7 November 2022)

Blanden, J. and Machin, S. (2008). 'Up and Down the Generational Income Ladder in Britain: Past Changes and Future Prospects', *National Institute Economic Review*, 205, pp. 101–116. Available at: https://doi.org/10.1177/0027950108096594 (Accessed: 12 April 2023)

Blanden, J., Goodman, A., Gregg, P. and Machin, S. (2004). 'Changes in intergenerational mobility in Britain', *Generational Income Mobility in North America and Europe.* pp. 122-146. Available at: https://doi.org/10.1017/CBO9780511492549.007 (Accessed: 8 February 2023)

Blanden, J., Gregg, P. and Macmillan, L. (2013). 'Intergenerational persistence in income and social class: the effect of within-group inequality', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 176(2), pp. 541-563. Available at: https://www.jstor.org/stable/23355204 (Accessed:13 November 2022)

Böhlmark, A. and Lindquist, M. J. (2006). 'Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden', *Journal of Labor Economics*, 24(4), pp. 879-896. Available at: https://doi.org/10.1086/506489 (Accessed: 2 January 2023)

Butler, N., Bynner, J.M., University of London. Institute of Education. Centre for Longitudinal Studies. (2016). *1970 British Cohort Study: Ten-Year Follow-Up, 1980*. [data collection]. *6th Release.* UK Data Service. SN: 3723, DOI: http://doi.org/10.5255/UKDA-SN-3723-7

Butler, N., Dowling, S., Osborn, A., University of London. Institute of Education. Centre for Longitudinal Studies. (2016). *1970 British Cohort Study: Five-Year Follow-Up, 1975*. [data collection]. *5th Release.* UK Data Service. SN: 2699, DOI: http://doi.org/10.5255/UKDA-SN-2699-4

Bynner, J., Butler, N., University of London, Institute of Education, Centre for Longitudinal Studies. (2022). *1970 British Cohort Study: Age 16, Sweep 4, 1986*. [data collection]. *9th Edition.* UK Data Service. SN: 3535, DOI: http://doi.org/10.5255/UKDA-SN-3535-6

Bynner, J.M., University of London. Institute of Education. Centre for Longitudinal Studies. (2016). *1970 British Cohort Study: Twenty-Six-Year Follow-Up, 1996*. [data collection]. *5th Release.* UK Data Service. SN: 3833, DOI: http://doi.org/10.5255/UKDA-SN-3833-3

Chung, C. F. and Goldberger, A. S. (1984). 'Proportional Projections in Limited Dependent Variable Models', *Econometrica*, 52(2), pp. 531–534. Available at: https://doi.org/10.2307/1911502 (Accessed: 24 October 2022)

Dahl, M. and DeLeire, T. (2008). *The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data*, University of Wisconsin–Madison, Institute for Research on Poverty, Discussion Paper No. 1342-08. Available at: https://www.semanticscholar.org/paper/THE-ASSOCIATION-BETWEEN-CHILDREN-%27S-EARNINGS-AND-%27-Dahl-DeLeire/8d4f43b8707fb46b7b27af287ee2c9d84ab5b4fd (Accessed: 3 November 2022)

Dearden, L., Machin, S. and Reed, H. (1997). 'Intergenerational Mobility in Britain', *The Economic Journal*, 107(440), pp. 47–66. Available at: https://www.jstor.org/stable/2235270 (Accessed: 14 November 2022)

Goldberger, A. S. (1981). 'Linear regression after selection', *Journal of Econometrics*, 15(3), pp. 357-366. Available at: https://doi.org/10.1016/0304-4076(81)90100-7 (Accessed: 7 November 2022)

Grawe, N, D. (2003). *Life cycle bias in the estimation of intergenerational earnings persistence*. Available at: https://www.semanticscholar.org/paper/Life-Cycle-Bias-in-the-Estimation-of-Earnings-Grawe/b2546db995a2fd1be60415730517fa7b5aa64b7b (Accessed: 4 November 2022)

Grawe, N. D. (2006). 'Lifecycle bias in estimates of intergenerational earnings persistence', *Labour Economics*, 13(5), pp. 551-570. Available at: https://doi.org/10.1016/j.labeco.2005.04.002 (Accessed: 2 November 2022)

Gregg, P., Macmillan, L. and Vittori, C. (2017). 'Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK', *Oxford Bulletin of Economics and Statistics*, 79(1), pp. 79-100. Available at: https://doi.org/10.1111/obes.12146 (Accessed: 17 November 2022)

Haider, S. and Solon, G. (2006). 'Life-Cycle Variation in the Association between Current and Lifetime Earnings', *American Economic Review,* 96(4), pp. 1308-1320. Available at: https://www.aeaweb.org/articles?id=10.1257/aer.96.4.1308 (Accessed: 28 November 2022)

Hauser, R. M. and Sewell, W. H. (1975). E*ducation, Occupation, and Earnings: Achievement in the Early Career*. New York: Academic Press.

Jenkins, S. (1987). 'Snapshots versus movies: 'Lifecycle biases' and the estimation of intergenerational earnings inheritance', *European Economic Review*, 31(5), pp. 1149-1158. Available at: https://doi.org/10.1016/S0014-2921(87)80010-7 (Accessed: 5 February 2023)

Jenkins, S. P. (2009). 'Spaghetti Unravelled: A Model-Based Description of Differences in Income-Age Trajectories', *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain*, pp. 170-202. Available at: https://doi.org/10.1093/acprof:oso/9780199226436.003.0007 (Accessed: 12 February 2023)

Lillard, L. A. and Reville, R. T. (1999). *Intergenerational Mobility in Earnings and Occupational Status*. Available at: http://www-personal.umich.edu/~charlieb/Selected-Papers/Life_Cycle/Lillard-Reville-Mobility.pdf (Accessed: 17 November 2022)

Mazumder, B. (2015). *Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data,* FRB of Chicago Working Paper No. 2015-04. Available at: http://dx.doi.org/10.2139/ssrn.2620727 (Accessed: 22 October 2022)

Mazumder, B. (2018). 'Intergenerational Mobility in the US: What We Have Learned from the PSID', *The ANNALS of the American Academy of Political and Social Science*, 680(1), pp. 213-234. Available at: https://doi.org/10.1177/0002716218794129 (Accessed: 21 December 2022)

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), pp. 235-255. Available at: https://doi.org/10.1162/0034653053970249 (Accessed: 11 November 2022)

Nybom, M. and Stuhler, J. (2016). 'Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation', *The Journal of Human Resources*, 51(1), pp. 239-268. Available at: https://www.jstor.org/stable/24736005 (Accessed: 12 April 2023)

O'Neill, D., Sweetman, O. and Van de gaer, D. (2007). 'The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility', *The Journal of Economic Inequality 5*, pp. 159–178. Available at: https://doi.org/10.1007/s10888-006-9035-7 (Accessed: 5 February 2023)

Office for National Statistics. (2022). 'Consumer price inflation, historical estimates and recent trends, UK: 1950 to 2022'. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/consumerpriceinflationhistoricalestimatesandrecenttrendsuk/1950to2022 (Accessed: 20 December 2022)

Office for National Statistics. (2022). 'CPI INDEX 00: ALL ITEMS 2015=100'. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7bt/mm23 (Accessed: (Accessed: 20 December 2022)

Piraino, P. (2007). 'Comparable Estimates of Intergenerational Income Mobility in Italy', *The B.E. Journal of Economic Analysis & Policy*, 7(2). Available at: https://doi.org/10.2202/1935-1682.1711 (Accessed: 20 October 2022)

Rohenkohl, B. (2019). *Intergenerational Income Mobility in the UK: New evidence using the BHPS and Understanding Society.* Working Papers 2019017, The University of Sheffield, Department of Economics. Available at: https://www.sheffield.ac.uk/media/3812/download (Accessed: 10 November 2022)

Solon, G. (1992). 'Intergenerational Income Mobility in the United States', *The American Economic Review,* 82(3), pp. 393-408. Available at: https://www.jstor.org/stable/2117312 (Accessed: 10 November 2022)

University of London, Institute of Education, Centre for Longitudinal Studies. (2019). *1970 British Cohort Study: Thirty-Eight-Year Follow-Up, 2008-2009*. [data collection]. *5th Edition.* UK Data Service. SN: 6557, DOI: http://doi.org/10.5255/UKDA-SN-6557-4

University of London, Institute of Education, Centre for Longitudinal Studies. (2020). *1970 British Cohort Study: Thirty-Four-Year Follow-Up, 2004-2005*. [data collection]. *5th Edition.* UK Data Service. SN: 5585, DOI: http://doi.org/10.5255/UKDA-SN-5585-4

University of London, Institute of Education, Centre for Longitudinal Studies. (2022). *1970 British Cohort Study: Age 42, Sweep 9, 2012*. [data collection]. *3rd Edition.* UK Data Service. SN: 7473, DOI: http://doi.org/10.5255/UKDA-SN-7473-3

University of London. Institute of Education. Centre for Longitudinal Studies. (2016). *1970 British Cohort Study: Twenty-Nine-Year Follow-Up, 1999-2000*. [data collection]. *4th Release.* UK Data Service. SN: 5558, DOI: http://doi.org/10.5255/UKDA-SN-5558-3

Zimmerman, D. J. (1992). 'Regression Toward Mediocrity in Economic Stature', *The American Economic Review,* 82(3), pp. 409-429. Available at: https://www.jstor.org/stable/2117313 (Accessed: 23 November 2022)

Appendices

Appendix A

Variables for son's non-cognitive skills, including *application, clumsiness, extroversion* and *hyperactivity* at age 10; *anxious* at age 10 and 16; cognitive skills such as *maths* and *reading* at age 10.

Non-cognitive skills:

<u>Teacher reported at age 10</u>

Application: 10 items

- Scale: what extent does the child accept the goals of the school curriculum.
- Is given to daydreaming.
- Cannot concentrate on any particular task, even though the child may return to it frequently.
- Becomes bored during class.
- Becomes confused or hesitant when given a complex task.
- Is easily distracted.
- Pays attention to what is being explained in class.
- Is forgetful when given a complex task.
- Shows lethargic and listless behaviour.
- Fails to finish things he starts.

Clumsiness: 3 items

- Trips or falls easily or bumps into objects to other children.
- Is noticeably clumsy in formal or informal games.
- Manipulates small objects easily with his/her hands.

Extroversion: 5 items

- When something important has happened, does the child endeavour to tell his or her friends about it?
- When something important has happened, does the child endeavour to tell his or her teacher about it?
- When talking to friends, is the child (compared to the rest of the class) talkative?
- When talking to teacher, is the child normally (compared to the rest of the class) talkative?
- Scale: Extrovert-Introvert.

Hyperactivity: 6 items.

- Is excitable, impulsive.
- Shows restless or over-active behaviour.
- Squirmy and fidgety.
- Hums or makes other odd vocal noises at inappropriate times.
- Given to rhythmic tapping or rhythmic kicking during class.
- Has twitches, mannerisms or tics of the face or body.

Anxious: 9 items

- Is fearful or afraid of new things or situations.
- Cries for little cause.
- Behaves ‘nervously’.
- Is fussy or over-particular.
- In relations with others appears to be miserable; unhappy tearful or distressed.
- Becomes obsessional about unimportant tasks.
- Is sullen or sulky.
- Truants from school.
- Fearful in movements, requires much encouragement to move faster.

<u>Mother reported at age 16</u>

Anxious: 8 items

- Often worried, worries about many things.
- Tends to do things on own rather solitary.
- Often appears miserable, unhappy, tearful or distressed.
- Tends to be fearful or afraid of new things or new situations.
- Is fussy or overparticular.
- Becomes obsessional about unimportant things.
- Is sullen or sulky.
- Cries for little cause.

Cognitive skills:

<u>Child reported at age 10</u>

Maths: Friendly Maths Test score

(d) ***Friendly Maths Test (FMT):*** The lack of a fully acceptable mathematics test appropriate for ten year olds led to the development of a special test for the *BCS70 Ten-year Follow-up*. This was done in

collaboration with Colin Appleton and John Kerley, specialists in primary mathematics. It was piloted in two halves in Bristol schools each on 400 children. It consisted of a total of 72 multiple choice questions and covered in essence the rules of arithmetic, number skills, fractions, measures in a variety of forms, algebra, geometry and statistics.

Reading: Edinburgh Reading Test score

(c) ***Shortened Edinburgh Reading Test (ERT) (Self-completion):*** The *Edinburgh Reading Test* is a test of word recognition and the shortened version used in this study was made up of items extracted from the full Edinburgh Reading Test after consultation with its authors (Godfrey Thomson Unit, 1978). Items were carefully selected to cover a wide age range of ability from seven to thirteen years in a form suitable to straddle the ten-year cohort. Particular attention was paid to the lower limit to allow a score to be allocated for very poor readers. The shortened test contained 67 items which examined vocabulary, syntax, sequencing, comprehension and retention.

Appendix B

Highest academic levels:

<u>Sons reported at age 30</u>

There are nine dummy variables where each interprets different academic level.

Dummy 0	Label = None
Dummy 1	Label = Bad GCSEs
Dummy 2	Label = CSE 2-5, other Scottish school qualification
Dummy 3	Label = O levels, Good GCSEs
Dummy 4	Label = 1 A level or more than 1 AS level at gr
Dummy 5	Label = 2 or more A-levels
Dummy 6	Label = Diploma of HE
Dummy 7	Label = Degree, other degree level
Dummy 8	Label = Higher degree