Cross-national comparisons of intergenerational mobility: are the earnings measures used robust?

John Jerrim [1]

Álvaro Choi [2]

Rosa Simancas Rodríguez[3]

1 Institute of Education, University of London

2 Institut d'Economia de Barcelona, University of Barcelona

3 University of Extremadura

October 2013

Abstract

Academics and policymakers have shown great interest in cross-national comparisons of intergenerational earnings mobility. However, producing reliable estimates of earnings mobility is not a trivial task. In most countries researchers are unable to observe earnings information for two generations. They are thus forced to rely upon imputed data instead. In this paper we consider the robustness of the 'two-sample two-stage least squares' (TSTSLS) methodology that is frequently applied within the earnings mobility literature. Our results suggest that the TSTSLS imputation procedure typically produces poor approximations to long-run earnings, leading to large biases in estimates of intergenerational associations. We hence conclude that TSTSLS estimates should not be used in cross-national comparisons of intergenerational earnings mobility. When we exclude such studies from international comparisons, key findings from this literature no longer hold.

Key Words: Social mobility, cross-national comparison, two sample two stage least squares, permanent earnings

JEL codes: I20, I21, I28.

Contact Details: John Jerrim (J.Jerrim@ioe.ac.uk) Department of Quantitative Social Science, Institute of Education, University of London, 20 Bedford Way London, WC1H 0AL

Acknowledgements: We would like to thank John Micklewright for helpful comments on an initial draft.

1. Introduction

The transfer of social status across generations is an issue of great social and political concern. Policymakers have shown particular interest in cross-national comparisons of intergenerational *earnings* mobility - the link between the 'permanent' (long-run) earnings of fathers and the 'permanent' (long-run) earnings of their sons. For instance the 'Great Gatsby Curve', a simple scatterplot showing a strong cross-national association between income inequality and intergenerational earnings mobility, has received a great deal of attention in the United States (e.g. Economic Report of the President 2012; Center for American Progress 2012; Krueger 2012; Corak 2012; The Economist 2012; The White House 2013). The same is true in the United Kingdom, where government officials and the media frequently discuss how Britain has extremely low levels of social (earnings) mobility by international standards.

However, producing reliable estimates of intergenerational earnings mobility, which can be legitimately compared across countries, is not a trivial task (Solon 1992; Blanden 2013). Ideally, long-run earnings information is needed in each country for two generations (e.g. fathers and sons). Yet in many countries earnings data is only available for a single generation (e.g. for sons only). This is a major problem, as the key explanatory variable (father's earnings) is not observed at all. A number of recent papers have attempted to overcome this problem by imputing father's earnings using a 'two-sample two-stage least squares' (TSTSLS) approach. A full list of papers is provided in Appendix A[1]. Figure 1 illustrates that for 21 countries included in a recent review of earnings mobility by Corak (2012), around half (11) have applied the TSTSLS methodology (those with white bars). This is a striking result; it highlights just how important TSTSLS is to the intergenerational earnings mobility literature, particularly when it comes to cross-national comparisons.

Figure 1

In this paper we analyse four high quality datasets from two countries, the United Kingdom and the United States, to mimic how the TSTSLS approach is applied within the intergenerational earnings mobility literature. We then compare TSTSLS imputed earnings to

[1] Some of the papers cited state that they have used two-sample instrumental variables (TSIV). TSIV and TSTSLS are numerically distinct estimators, though asymptotically equivalent, with the latter being computationally easier and more efficient - see Nicoletti and Ermisch 2008 and Inoue and Solon 2010. Moreover, as Inoue and Solon (2010) note 'of the many empirical researchers who have since used a two-sample approach nearly all have used the two-sample two-stage least squares' estimator.

actual observed measures of long-run earnings to investigate the robustness of this approach. Specifically, this paper aims to:

(i) Investigate the quality of TSTSLS imputations of long-run earnings

(ii) Assess the reliability of intergenerational associations based upon this methodology

(iii) Establish, by implication, the credibility of international comparisons of intergenerational earnings mobility.

Our results suggest that:

- The correlation between TSTSLS predictions ($\hat{X}_{TSTSLS}$) and actual observed long-run earnings (X_{AVG}) is rather weak (typically r < 0.5).
- The difference between imputed and observed long-run earnings is not simply a matter of 'random noise'
- Intergenerational associations based upon this methodology are likely to be overestimated (although this cannot be automatically assumed).
- Academics and policymakers should therefore exercise a great deal of caution when interpreting cross-national comparisons of intergenerational earnings mobility, where TSTSLS imputations of father's earnings have been widely applied.

The paper now proceeds as follows. Section 2 describes the earnings mobility estimation problem and provides an overview of the TSTSLS imputation approach. Section 3 provides an overview of the British Household Panel Survey (BHPS) and the Labour Force Survey (LFS) datasets. It also describes our empirical methodology. In section 4 we compare TSTSLS imputations of men's earnings to actual observed measures using the BHPS dataset. We also investigate the robustness of intergenerational associations, focusing on the link between father's earnings and their children's educational plans. Conclusions and implications for future research follow in section 5.

2. The estimation problem

When investigating intergenerational earnings mobility, economists would ideally like to estimate the following Ordinary Least Squares (OLS) regression model:

$$Y_{True} = \alpha + \beta . X_{True} + \varepsilon \qquad (1)$$

Where:

X_{True} = (Log) permanent earnings of parent (e.g. father)

Y_{True} = (Log) permanent earnings of offspring (e.g. sons)

The parameter estimate of interest from (1) is $\hat{\beta}$. This is the estimated 'intergenerational earnings elasticity' - the most frequently used measure of intergenerational earnings mobility used in the cross-national comparative literature. (The intergenerational correlation, 'r', is an alternative measure, which re-scales $\hat{\beta}$ to take into account differences in income inequality across the two generations. Although Björklund and Jantti (2009) note that this measure has significant advantages, it is less frequently reported than the income elasticity). However direct estimation of (1) is not usually possible. This is because of the very demanding data requirements; information is needed on entire career earnings of parents and their offspring (e.g. for fathers and their sons). Thus X_{True} and Y_{True} are unobserved, with proxy measures used in their place. This can lead to bias in $\hat{\beta}$ if the proxy's miss-measure the constructs of interest. Although this is potentially true for both X_{True} and Y_{True}, the former has received by far the most attention in the existing literature[2]. It is also the focus of this paper.

Our survey of the literature suggests that four proxies for X_{True} are frequently used:

(a) X_{Single} = Father's earnings observed within a single – year ('current earnings')
(b) X_{IV} = Current father's earnings used in conjunction with an instrumental variable
(c) X_{AVG}= Father's earnings averaged over a number of years
(d) X_{TSTSLS} = Imputed father's earnings based upon other observable characteristics

Within the earnings mobility literature, option (a) is considered unsatisfactory. This is because earnings observed for an individual within any given year are likely to be subject to 'transitory' fluctuations (i.e. X_{Single} is a 'noisy' measure of X_{True}). Consequently, the intergenerational elasticity (β) is likely to be *underestimated*. The second option (X_{IV}) can potentially overcome this problem, though a credible instrument often cannot be found.

---

[2] If measurement error and transitory fluctuations in the dependent variable (Y) are random, OLS estimation of (1) continues to produce unbiased estimates of the intergenerational income elasticity (although less efficiently than using perfectly measured data). However, the same does not hold true for the explanatory variable (X), where such 'classical' measurement error leads to attenuation bias. This is a key reason why measurement error in X has been the focus of the income mobility literature. Although Haider and Solon (2006) suggest that non-random measurement error in Y can also lead to biased estimates, they indicate that this is likely to be small if sons' earnings are measured at approximately age 40.

Despite obvious problems, parental education and occupation are often the IV's chosen, with β *overestimated* as a result (Dearden et al 1997). The third approach (X_{AVG}) is hence typically preferred. Estimates of β will continue to be downwardly biased, but there will be convergence towards the true population parameter as the number of years averaged over increases. Although five consecutive years of parental earnings data is often used (Solon 1992; Vogel 2008; Björklund and Chadwick 2003; Hussein et al 2008; Corak and Heisz 1999), more than ten may be needed to sufficiently reduce this bias if there is substantial auto-correlation in the transitory component of earnings over time (Björklund and Jantti 2009; Mazumder 2005).

However it is the fourth and final proxy (X_{TSTSLS}) that is the focus of this paper. As noted in the introduction, this has been used to create intergenerational earnings mobility estimates for a number of countries (e.g. Australia, France, Italy, Spain, Japan, United Kingdom, Switzerland, China, Chile, Brazil) where researchers face an even more serious problem; within the dataset under investigation, no information is available on parental earnings at all. Thus simply replacing X_{True} in (1) with X_{Single}, X_{IV} or the preferred X_{AVG} is not possible, meaning academics have to turn to X_{TSTSLS} instead.

Within the earnings mobility literature, this TSTSLS approach is often described within an instrumental variable framework (e.g. Lefranc and Trannoy 2005; Nuñez and Miranda 2011). However, we believe it is more appropriate to consider the method applied as a cold-deck imputation (Nicoletti and Ermisch 2008) or 'generated regressor' (Murphy and Topel 1985; Wooldridge 2002:115) procedure. It can be summarised as follows. A researcher has access to two datasets: (i) the 'main' sample and (ii) the 'auxiliary' sample. The researcher wishes to estimate equation (1) above using the main sample. Unfortunately, (1) cannot be estimated directly as X_{True} is unobserved, and there is no readily available proxy (X_{Single} or X_{AVG}) to use in its place. However, the main dataset does contain a series of additional characteristics (Z) which one would expect to be associated with X_{True} (e.g. parental education, parental occupation). These Z characteristics are often called the 'instrumental variables' in the earnings mobility literature, though we believe 'imputer variables' is a more appropriate term.

Now say that the second 'auxiliary' sample (i) contains a measure of *current* earnings[3] (ii) is drawn from the same population as the main sample and (iii) contains the same 'imputer' Z variables as the main sample. The following OLS regression model can then be estimated using this auxiliary sample:

$$X = \alpha + \gamma . Z + u \qquad (2)$$

Where:

X = Current earnings in the auxiliary dataset

Z = The imputation variables (e.g. parental education, parental occupation)

Generating the following prediction equation:

$$\hat{X}_{TSTSLS} = \alpha + \hat{\gamma} . Z \qquad (3)$$

As the Z characteristics are also observed for individuals in the main sample, (3) can be used to replace unobserved permanent earnings (X_{True}) with the linear prediction ($\hat{X}_{TSTSLS}$). This means (4) can be estimated using the main sample instead of (1):

$$Y = \alpha + \beta . \hat{X}_{TSTSLS} + \varepsilon \qquad (4)$$

However, (4) will only produce reliable estimates of β if the imputed proxy ($\hat{X}_{TSTSLS}$) is closely related to 'true' permanent earnings (X_{True}). This will depend upon: (i) whether the main and auxiliary samples are drawn from the same population; (ii) the ability of Z (imputer variables) to predict earnings; (iii) whether the Z characteristics are measured in the same way in the two datasets; (iv) the auxiliary dataset sample size. The aim of this paper is to empirically investigate whether these assumptions hold, with a focus on points (ii) and (iv) above[4].

3. Data

[3] It is not clear why applications of the TSTSLS methodology use current earnings in this 'first – stage' regression rather than a measure of long-run earnings. Our presumption is that, although the latter would be preferable, it is rarely available, and so current earnings are used in their place.

[4] In most applications, researchers rely upon children's reports of their parents socio-economic characteristics in the main dataset. The difficulty with relying upon such reports has been widely discussed in the sociological literature (Looker 1989; Jerrim and Micklewright 2012). This issue is not investigated in detail here, where parental reports of their own characteristics are used within both the main and the auxiliary dataset. We are therefore likely to underestimate the potential difficulties with implementing the TSTSLS imputation procedure.

Our analysis draws upon two large, high quality British datasets: The British Household Panel Survey (BHPS) and the Labour Force Survey (LFS). The former acts as our 'main' sample and the latter the 'auxiliary' sample. We have chosen to focus upon the BHPS due to its large sample size, the detailed information available on respondents' earnings over a number of years, its widespread use, public accessibility, and the availability of youth supplement data to allow estimation of certain intergenerational associations. Appendix B describes two US datasets (the National Longitudinal Survey of Youth 1979, NLSY79, and the Current Population Survey, CPS) which we use to supplement our analysis.

BHPS data

The BHPS is a nationally representative longitudinal sample of British households. Data were initially collected in 1991 (wave 1) via a stratified clustered sample design. Annual face-to-face interviews have been conducted with all household members over the age of 16 up to 2008 (wave 18). The original sample size was 5,050 households, containing information on 9,092 individuals (a response rate of 74 percent). Sample members have been followed as they move address. New people joined the BHPS cohort when they started sharing the same household as a permanent sample member. Throughout our analysis we focus upon male respondents who have labour market earnings recorded in at least five BHPS survey waves. This leaves a total of 3,080 observations. We apply the 2008 longitudinal enumerated weight to adjust for non-random non-response.

Table 1 illustrates the number of labour market earnings observations available for these 3,080 individuals. Three-quarters have data available from eight or more years, with more than half having data from ten years or more. To create a long-run ('permanent') earnings measure we first of all inflate data to 2010 prices. Next, we divide respondents reported annual labour market earnings by the number of hours they work in a typical week. This gross hourly pay variable is then averaged for each respondent across all available survey waves. We call this derived variable X_{BHPS}. Blanden, Gregg and Macmillian (2013) have created a comparable 'parental income' measure for the BHPS using a similar approach.

Table 1

It is important to note that the variable we have derived actually refers to long-run average *earnings* (labour market income only). This is different to long-run *income*, which also includes interest, dividends and social security payments (amongst other things). We have intentionally chosen to focus on earnings as much of the existing intergenerational mobility literature actually focuses on this concept rather than income (e.g. Solon 1992; Hussain, Munk and Bonke 2008). This is particularly true of studies where the TSTSLS approach has been applied; the 'first-stage' prediction equation has almost always been specified with earnings from work as the dependant variable (therefore imputing father's earnings into the main dataset). Hence we believe that TSTSLS estimates actually capture intergenerational *earnings* mobility (rather than income mobility) and the approach taken in this paper is consistent with this view.

A second issue is that X_{BHPS} is still not an exact measure of respondents' permanent career earnings. This is because we only have access to between 5 and 18 years of data for each individual (see Table 1) rather than their entire 40 - 50 year career. Hence it may be more appropriate to consider X_{BHPS} as akin to the preferred time-average proxy (X_{AVG}) used in the most robust studies of intergenerational earnings mobility. Consequently we are not able to investigate measurement error in the TSTSLS imputations per se. Rather we consider how the TSTSLS imputations compare to the best long-run earnings measures currently used in the intergenerational mobility literature. To check the robustness of our results, we repeat our analyses using US data, where it is possible to average earnings data over an even greater number of years. Selected results from this supplementary analysis shall be presented where appropriate (full details can be found in Appendix B).

As part of the BHPS respondents have also been asked detailed questions about their current occupation and educational attainment. We use the one digit version of the SOC 2000 codes provided by BHPS, which places sample members into one of the following nine groups: (i) Managers / senior officials ; (ii) Professionals; (iii) Associate professionals (iv) Administration; (v) Skilled trade; (vi) Services; (vii) Sales; (viii) Plant and machine operative; (ix) Elementary occupations. With regards to education, respondents were asked about the qualifications that they hold. Using the information provided, the survey organisers have derived a 'highest academic qualification' variable. We combine the top two categories (higher degree and first degree) to maximise comparability with the LFS (see sub-section below) leaving the following groups: (i) Bachelor degree and higher; (ii) Other higher education; (iii) A-Levels; (iv) O-Levels; (v) CSE; (vi) None. These are the key imputation variables that will be used in our application of the TSTSLS technique.

From wave 4, the BHPS has also collected information from 11 – 15 year old children within respondent households. This included questions on whether the child expects to continue in education beyond age 16 (the minimum school leaving age in the UK). This data is used to test the robustness of intergenerational associations. Information is drawn from the final BHPS wave (2008) or the most recent available. We are able to link a total of 917 youths to fathers who have at least five labour market earnings observations available and who took part in the final BHPS wave. The BHPS youth weight is applied during this part of the analysis.

Labour Force Survey (auxiliary dataset)

We use numerous rounds of the Labour Force Survey (LFS) as our auxiliary dataset. This is cross-sectional data, collected by the UK Office for National Statistics, and has been designed to provide a nationally representative snapshot of the UK labour force. We pool information across all LFS waves between 2006 and 2008 to ensure a large sample size (we discuss the importance of the auxiliary dataset sample size in more detail below). The sample is then restricted to male respondents between the ages of 18 and 65. This leaves a total of 76,291 observations. As part of the LFS, respondents were asked a series of questions about their earnings and hours of work. The survey organisers have used this information to derive a gross hourly pay variable, which we adjust into real 2010 prices. We will use this information as the dependent variable in our 'first-stage' prediction equation (we have tested the robustness of our findings to using annual earnings instead, with little substantive change to our results. See also estimates using US data in Appendix B). The person weight, which helps to compensate for non-response and grosses the sample up to population estimates, is applied throughout.

The LFS also contains detailed information on respondents' current occupation and qualifications. The former is recorded as four digit SOC 2000 codes, the same schema as used in the BHPS. We also create a one digit (nine groups) version of this schema, as described for the BHPS in the previous sub-section. With regards to education, we convert the 30 categories provided into the same six schema used for the BHPS (we have experimented with alternative mappings and have found our substantive conclusions to be largely unchanged). A 'highest academic qualification' variable is then derived.

Methodology

The LFS is used to impute long-run earnings into the BHPS following the TSTSLS approach. The twist, of course, is that the BHPS also contains an actual measure of respondents' long-run earnings ($X_{AVG} = X_{BHPS}$) as described above. This means that we can compare $\hat{X}_{TSTSLS}$ to X_{AVG} to assess the 'quality' of the imputed earnings variable.

We begin by estimating a simple log-earnings regression model using the LFS (auxiliary) sample:

$$X_{LFS} = \alpha + \gamma . Z_{LFS} + u$$

Where 'X_{LFS}' refers to the natural logarithm of LFS hourly earnings. From this we generate the following prediction equation:

$$\hat{X}_{TSTSLS} = \hat{\alpha} + \hat{\gamma} . Z$$

There are several candidates to include as the Z (imputer) variables. Appendix A provides details on those typically used in the literature. There are three common choices: (i) broad education level only (ii) broad occupation only (iii) both broad education and broad occupation. We produce estimates using (i), (ii) and (iii) to investigate how this influences our results. A fourth model including broad education and very detailed occupational information (four digit SOC 2000 codes) will also be estimated. Although such a detailed 'first – stage' regression has only occasionally been used in the literature (e.g. Leigh 2007), we want to know whether this leads to a substantial improvement in the prediction of long-run earnings. Basic demographic characteristics are also included in the prediction models, such as ethnicity, age and age^2 (it is standard procedure to include individuals of all ages in the first-stage regression, with age and age^2 covariates capturing the non-linear relationship between age and earnings. Age is then usually set to 40 when generating predictions). Estimates from these models can be found in Appendix C. We choose a relatively simple specification of the 'first stage' prediction model as this is consistent with existing practice within the income mobility literature (see Björklund and Jantti 1997, Piraino 2007 and Cervini – Pla 2011 and Appendix A for examples). We use these estimates to impute long-run earnings into the BHPS dataset following the TSTSLS approach:

$$\hat{X}_{TSTSLS} = \hat{\alpha} + \hat{\gamma} . Z_{BHPS}$$

Our first task is to then compare the TSTSLS predictions ($\hat{X}_{TSTSLS}$) with actual (time average) long-run earnings (X_{AVG}) for BHPS sample members. We do this in a number of ways. To begin, we compare simple descriptive statistics of the predicted ($\hat{X}_{TSTSLS}$) and observed (X_{AVG}) long-run earnings distributions. Second, we consider the correlation between observed long-run earnings and the various TSTSLS imputations. Third, we divide the two measures into quartiles and present cross-tabulations and categorical measures of association (e.g. Cohen's Kappa and the percentage correct). Fourth, we investigate whether there are systematic differences between $\hat{X}_{TSTSLS}$ and X_{AVG} in terms of observable characteristics. To check the robustness of our results, we replicate all the above using two US datasets (NLSY and CPS) with further details available in Appendix B. Additional robustness tests are presented in Appendix D.

Our second task will be to investigate the robustness of estimated intergenerational associations. Unfortunately the BHPS does not contain enough information on offspring's earnings to test the robustness of earnings mobility estimates. However, information regarding the educational intentions of the cohort members' children is included within the BHPS youth questionnaire. We therefore investigate the link between children's educational plans and father's earnings using the following simple linear probability model (substantive conclusions hold if a logistic regression model is estimated instead):

$$S = \alpha + \delta . X + \varepsilon$$

Where:

S = 1 if the child expects to undertake post-16 education (0 otherwise)

X = Father's log hourly earnings

This model is estimated seven times, using the following different approximations for father's long-run earnings (X):

- (i) Current earnings in 2008 (X_{Single})
- (ii) Time – averaged earnings ($X_{AVG} = X_{BHPS}$)
- (iii) Current earnings in conjunction with an instrumental variable (X_{IV})
- (iv) TSTSLS imputation model 1 (race, age, education)
- (v) TSTSLS imputation model 2 (race, age, broad occupation)
- (vi) TSTSLS imputation model 3 (race, age, education, broad occupation)

(vii) TSTSLS imputation model 4 (race, age, education, detailed occupation)

In this part of the analysis we restrict the sample to the 917 observations with the relevant data available. Our primary interest shall be the extent to which (i) and (iii) – (vii) under or overestimate the association between fathers' earnings and children's schooling intentions relative to (ii). In Appendix B we perform a similar analysis using the two US datasets – but focusing on the relationship between mothers' earnings and children's scores on a standardised achievement test.

Finally, we will investigate how the quality of the TSTSLS predictions changes as the auxiliary sample size decreases. There is significant heterogeneity in the auxiliary sample size used within the existing literature. For instance, Ferreira and Veloso (2006) have access to 253,798 observations, compared to 1,033 in Nicoletti and Ermisch (2008), 540 for Sweden in Björklund and Jantti (1997) and as few as 166 for Peru in Grawe (2004). We hypothesise that as the auxiliary sample size diminishes so too will the quality of the TSTSLS imputations of long-run earnings. This is because the TSTSLS predictions will have a degree of uncertainty (i.e. they will have associated standard errors) which is inversely related to the sample size. Hence when the number of observations in the auxiliary dataset declines so will the precision of the long-run earnings imputations. This will then lead to attenuation in the relationship between $\hat{X}_{TSTSLS}$ and X_{AVG}. To our knowledge, this point has not been recognised in the existing literature.

To empirically investigate this issue, we follow the seven steps outlined below:

Step 1 → 5,000 observations will be randomly selected from the LFS dataset

Step 2 → The TSTSLS prediction model will be re-estimated using these 5,000 observations

Step 3 → The TSTSLS imputations of father's earnings shall be updated (based upon the new prediction model estimated in step 2).

Step 4 → The correlation between imputed and observed long run earnings in the BHPS will be re-estimated

Step 5 →The association between fathers' imputed long-run earnings and their offspring's school intentions will be re-estimated.

Step 6 → A further 50 observations will be randomly dropped from the LFS dataset (leaving 4,950)

Step 7 → Step 2 to step 6 shall be repeated

The above process is continued until there are no observations left in the auxiliary (LFS) dataset. We then plot the relationship between the estimates produced in steps 4 and 5 against the auxiliary dataset sample size. This will demonstrate whether reducing the sample size does indeed lead to attenuated estimates.

4. Results

The quality of the TSTSLS earnings imputations

In Table 2 we present our comparison of imputed and observed long-run earnings. Panel (a) refers to our main analysis using UK data, while panel (b) presents supplementary results for the US (see Appendix B). As Nicoletti and Ermisch (2008) note, the quality of TSTSLS long-run earnings imputations are likely to improve as the R^2 of the 'first stage' equation increases (ceteris paribus). We therefore present the R^2 values from our first-stage prediction equations in the top row of Table 2. These typically fall between 0.30 and 0.40. In Appendix A we review all the studies that have applied the TSTSLS earnings imputation methodology and find that R^2 values of this magnitude are consistent with those in the literature (where this information is reported). Nevertheless, this level of statistical 'fit' is not particularly strong; less than half the variation in log earnings has been explained in the first stage equation. This provides the first indication that the quality of the TSTSLS earnings imputations may be quite low.

Table 2

The second row of Table 2 provides information on the variance of imputed and observed long-run earnings. Regardless of the first-stage imputation model used, the variance of long-run earnings is significantly underestimated. For instance, the variance of observed (time – average) long-run earnings is 0.22 log-points. This compares to just 0.09 log points using TSTSLS imputation model 3, and a value as low as 0.04 when using model 1. Underestimation of the long-run earnings variance is hence in the region of 50 to 80 percent.

Next, we turn to the strength of the association between imputed and observed measures of long-run earnings. Estimated correlation coefficients can be found in the third row of Table 2. Scatterplots are also presented in Figure 2, with observed (time – average) values on the x-axis and TSTSLS imputations on the y-axis. If the TSTSLS approach produced exact replicas of observed long – run earnings, all data points would sit on the 45 degree line, and the estimated correlation coefficient would equal one. Correlation coefficients less than one and scatter around the 45 degree line hence illustrate the extent to which X_{AVG} and $\hat{X}_{TSTSLS}$ disagree.

Figure 2

The correlation between observed and predicted long-run earnings is modest (at best). Depending upon the 'first-stage' prediction model used, the estimated correlation falls somewhere between 0.3 and 0.5. Focusing on TSTSLS imputation model 3, the most common specification used in the existing literature, the estimated correlation coefficient is just 0.5. This implies that the TSTSLS imputed proxy captures just 25 percent of the variation in observed (time average) long-run earnings; three-quarters is *not* accounted for. This point is further emphasised in Figure 2 – there is substantial scatter of points around the 45 degree line, with only weak evidence of any positive association. Although this holds true in both panel a (model 1) and panel b (model 4), there is some improvement when the more detailed first – stage regression specification has been used. Nevertheless, one may view these modest correlations as rather disappointing; they suggest that the TSTSLS imputations only produce a rather weak proxy for men's long-run (permanent) earnings. Results for the US firmly support this view (see Appendix B).

Many studies of intergenerational earnings mobility also present transition matrices; fathers' and sons' earnings are divided into four equal groups ('quartiles') which are then cross-tabulated. Bauer (2006), Piraino (2007) and Leigh (2007) are examples having imputed fathers' earnings using TSTSLS. But how often are fathers assigned to the 'right' earnings quartile? The answer to this question can be found in Table 3, where we cross-tabulate TSTSLS imputed income quartile (using imputation model 3) against the time-average income quartile. Panel (a) illustrates how this cross-tabulation would look in the case of perfect agreement while panel (b) demonstrates the pattern under random assignment. Results for the UK and US can be found in panels (c) and (d). The agreement between the two measures is clearly rather low. The main diagonal in Table 3 panel (c) contains values of approximately 50 or below, with the lowest values coming outside of the tails of the distribution (i.e. outside the top and bottom earnings quartile). For instance, of those BHPS sample members in the 3rd 'time average' earnings quartile (shaded light grey), there is a 22 percent chance of them being assigned to the bottom TSTSLS earnings quartile, 23 percent in the second quartile, 33 percent in the third quartile and 22 percent in the top. This is little different to the situation under random assignment shown in panel (b).

In Table 2 we summarise the extent of agreement between time-average (observed) and TSTSLS (imputed) income quartile by presenting Cohen's Kappa (fourth row) and the percentage agreement (fifth row). The former is a measure of 'inter-rater' reliability that adjusts for chance agreement, and is frequently used in the psychometric literature. To aid interpretation, we follow the rules of thumb in Landis and Koch (1977), who suggested that Kappa statistics between 0.01 to 0.20 indicates 'slight' agreement, 0.21 to 0.40 'fair', 0.41 to 0.60 'moderate', 0.61 to 0.80 'substantial' and 0.81 to 0.99 'almost perfect' agreement. The Kappa statistics presented in Table 2 are in the range 0.13 to 0.23 – suggesting that there is evidence of only 'slight' to 'fair' agreement between observed and imputed earnings quartiles. This is well below the 0.40 that many believe to be the minimum acceptable value (e.g. Fleiss 1981). One can also see that only 35 to 40 percent of BHPS sample members are placed in the same earnings quartiles using the two techniques. This once again illustrates their lack of comparability, and that the TSTSLS imputation procedure generates weak measures of long-run earnings.

Table 3

The 'error' in the TSTSLS earnings imputations is now considered in more detail using the UK data. Specifically, we attempt to establish whether the discrepancy between observed and imputed long-run earnings is associated with a set of observables characteristics. This will help to reveal whether the scatter about the 45 degree line in Figure 2 follows a particular pattern or is simply random 'noise'. In essence, we are exploring whether this discrepancy has properties similar to 'classical' measurement error. We create a new variable (D) which captures the difference between observed and imputed long-run earnings:

$$D = X_{AVG} - \hat{X}_{TSTSLS}$$

For ease of interpretation, D has been standardised to have a mean of 0 and a standard deviation of 1. A series of bivariate OLS regression models are estimated:

$$D = \alpha + \gamma . E + \varepsilon$$

each using one of the following explanatory (E) variables:

(i) Occupation
(ii) Education
(iii) Industry
(iv) Whether the respondent has a child who plans to stay in school beyond age 16

Results can be found in Table 4.

Table 4

There are a number of statistically significant parameter estimates in each panel. Moreover, these are often large in absolute value. Focusing upon model 3, one can see that the discrepancy between observed and imputed long-run earnings is 0.42 of a standard deviation bigger for men working in elementary occupations than for those who are senior officials. Similarly, there is a difference of around 0.42 of a standard deviation between men with no education compared to men with a bachelor's degree or higher. Table 4 also indicates that the prediction 'error' (D) is often associated with children's educational plans. This ranges from 0.13 of a standard deviation in model 3 to 0.26 in model 1. Together, Table 4 clearly illustrates that there are a number of observable factors associated with the prediction error. Therefore the difference between TSTSLS imputed earnings ($\hat{X}_{TSTSLS}$) and observed time-averaged earnings (X_{AVG}) cannot simply be thought of as random 'noise'.

The impact upon intergenerational associations

We have thus far established that TSTSLS imputations:

(i) significantly underestimate the variance of actual (observed) long-run earnings
(ii) are only modestly associated with observed long-run earnings
(iii) are *not* simply 'noisy' approximations of long-run earnings

Next we consider the influence this has upon estimated intergenerational associations using the UK data, focusing upon the relationship between father's earnings and their children's educational plans. Estimates from the simple linear probability model described in section 3 can be found in Figure 3.

Figure 3

When using earnings data from a single year (X_{Single}), the parameter estimate of interest equals 0.112 (left-hand most bar). This suggests that a one log-unit increase in father's hourly earnings leads to an 11 percentage point increase in their offspring expecting to continue their education beyond age 16. As discussed in section 2, one would expect this estimate to be downwardly biased. Our results confirm this – the second bar from the left of Figure 3 is when long-run (time-averaged) earnings are used instead ($X_{AVG} = X_{BHPS}$). The estimated coefficient is now 0.132 – an increase of approximately 17 percent. In contrast, the instrumental variable (X_{IV}) estimate is around 0.22 - roughly 75 percent higher than when the time-average approach is used. As Dearden et al (1997) note, X_{Single} and X_{IV} can therefore be used to bound X_{True}. However, this is likely to be of limited use in cross-national research, as the range of possible values is usually very wide.

Note that in all four estimates using the TSTSLS imputations are above those when the long-run time average method is used. Overestimation is particularly large when only race, age and education are included in the 'first-stage' prediction equation (model 1). The parameter estimate stands at 0.29 – more than double the 0.13 found when the time average approach has been used. When a measure of occupation is included in the first stage regression, the estimated intergenerational association falls to 0.21. However, although the upward bias has been reduced, it is still more than 50 percent above the preferred (time averaged) estimate. Indeed, even when very detailed occupational information is included in the prediction model, intergenerational associations are still overestimated by approximately one third. To test the robustness of these results, Appendix B presents analogous results using the two US datasets and a different dependent variable (children's maths test scores). We continue to find substantial overestimation of intergenerational associations, though with a notable decline in the upward bias as additional detail is added to the first stage prediction model.

Of course, it is important to remember that, due to data limitations, we have only been able to investigate the link between parental income and young people's educational expectations and cognitive test scores (see Appendix B). Would a similar upward bias emerge if offspring income were used as the dependent variable instead? Björklund and Jantti (1997:Table 2) provides some insight into this issue. Using a small US dataset (n ≈ 300), they find evidence that IV and TSTSLS estimates of the intergenerational income elasticity are upwardly inconsistent by approximately 30 percent (when using education and occupation in the first stage prediction equation – as per our 'model 3'). However, the confidence interval around this estimate is very wide, with the upper bound stretching above 50 percent. Nevertheless such upward inconsistency of estimates is clearly in-line with the evidence we have presented in Figure 3. Together this suggests that intergenerational associations are likely to be significantly overestimated in current applications of TSTSLS relative to time-averaging. At this point, it is worth recalling Figure 1. Notice that countries with stronger intergenerational associations are the ones where the TSTSLS approach has been applied, rather than the time-averaging approach (i.e. the white bars all tend to be towards the bottom half of Figure 1). This could be the genuine ranking of countries, or it could be due to the inconsistency in the TSTSLS estimates described above. Although we recognise that some cross-national comparisons have adjusted TSTSLS estimates downwards in an attempt to take the upward inconsistency into account (e.g. Blanden 2013), the reality is that such adjustments are difficult to make as the size of the bias is actually unknown (and will depend, amongst other things, on the predictive ability of the imputer variables used). We shall discuss one of the implications of this finding in the conclusion to the paper.

Before doing so, it is worth considering whether intergenerational associations based upon TSTSLS imputations are always an overestimate of the time average approach (meaning one can treat them as an upper bound). We argue that although overestimation is likely, this does not necessarily have to hold. Indeed, when the sample size in the auxiliary dataset is very small, intergenerational associations may be *underestimated* (particularly when detailed information is included in the first stage prediction equation).

This point is illustrated in Figures 4 and 5, where we plot the relationship between the auxiliary dataset sample size and:

- The correlation between imputed and observed long-run earnings (left-hand panel)
- The estimated association between imputed earnings and children's educational plans (right hand panel).

Figure 4 refers to when TSTSLS imputation model 3 has been used while Figure 5 refers to imputation model 4. Starting with the left-hand panel of Figure 4, the correlation between observed and imputed long-run earnings is stable at around 0.48 when there are approximately 1,000 observations or more in the auxiliary dataset. However, there is some evidence of attenuation when the sample size starts to fall below this level (the sharp drop towards the left hand side of the graph). The left-hand panel of Figure 5 is consistent with this result. In particular, note how the correlation between observed and imputed long-run earnings falls from approximately 0.5 when there are around 5,000 auxiliary dataset observations to below 0.40 when there are 500 or fewer. This also highlights how a detailed imputation model combined with a small auxiliary dataset can be particularly problematic.

The right hand panel of Figure 4 and Figure 5 illustrates the impact that this has upon estimated intergenerational associations. Two horizontal lines have been superimposed on these graphs. The uppermost line illustrates the estimated intergenerational association using all 76,291 LFS observations to generate the TSTSLS earnings imputations. The lower line refers to the estimated intergenerational association using observed time-average earnings (X_{AVG}). When there are more than approximately 1,000 observations in the auxiliary dataset, estimates of the link between fathers' earnings and children's educational plans seem to be quite stable. However, when the sample size starts to drop below this level, estimates begin to fall. Indeed, in both Figure 4 and Figure 5 there are points below the lower superimposed line. This illustrates how TSTSLS imputations of father's earnings can lead to underestimation of intergenerational associations relative to the 'time-average' approach (while noting that, due to the issues discussed earlier in this section, overestimation is more likely). More generally, our experimentations with different imputation models and different random number seeds suggest that estimates based upon the TSTSLS approach become quite erratic as the auxiliary dataset becomes small. This is not only due to the problem of attenuation described above, but also because sampling variation within the first-stage prediction equation becomes quite large. We thus advise readers to be particularly cautious when interpreting TSTSLS intergenerational estimates where a relatively small auxiliary dataset has been used (e.g. Grawe 2004; Piraino 2007; Andrews and Leigh 2009; Bidisha 2013).

5. Conclusions

Intergenerational earnings mobility is a topic of great academic and political concern. However, producing reliable estimates of earnings mobility is not a simple task. In many countries earnings data cannot be linked across two generations. Consequently, several studies have had to impute information on fathers' earnings using the two-sample two-stage least squares (TSTSLS) approach. This paper has considered the quality of the TSTSLS imputed earnings data and the robustness of intergenerational associations based upon this methodology. Using four rich datasets from two developed countries, we have shown that

TSTSLS imputed earnings (X_{TSTSLS}) are only moderately associated with actual observed measures of long-run earnings (X_{AVG}), and that the former are not simply a 'noisy' approximation of the latter. Moreover, simple regression models using TSTSLS imputed earnings data can result in severely biased parameter estimates. We consequently conclude that academics and policymakers should exercise a great deal of caution when interpreting intergenerational associations where the TSTSLS imputation procedure has been applied.

These findings have important implications for international comparisons of intergenerational earnings mobility, where estimates are frequently treated as cross-national comparable, even when different empirical methodologies have been applied. For instance, Piraino (2007) estimates earnings mobility in Italy using TSTSLS, but then compares results to Sweden and the United States where time-average earnings have been used. Despite clear differences in methodology, it is claimed that '*new internationally comparable estimates of the degree of intergenerational mobility in Italy*' are produced - with the paper entitled '*Comparable Estimates of Intergenerational Income Mobility in Italy*' [emphasis our own]. We have shown that such strong statements are difficult to justify, as any variation found across countries could simply be due to differences in methodological approach.

More generally, when TSTSLS estimates are excluded from international comparisons, key findings from the earnings mobility literature no longer hold. One important example is the supposedly strong cross-national relationship between income inequality and intergenerational earnings mobility (often illustrated by the 'Great Gatsby Curve' – see http://www.whitehouse.gov/blog/2013/06/11/what-great-gatsby-curve). This graph has been widely discussed by leading policymakers (e.g. The White House 2013, The Sutton Trust 2013) and international media (e.g. The Economist 2012; The New York Times 2012), with the OECD (2011) summing up the conventional wisdom that '*intergenerational earnings mobility is low in countries with high inequality*.' We reproduce the 'Great Gatsby Curve' in the left hand panel of Figure 6 using all 'preferred' income inequality and earnings mobility estimates from three recent international comparative studies (Björklund and Jantti 2009, Corak 2012 and Blanden 2013). This includes a total of 45 earnings mobility estimates covering 22 developed and developing nations (note some have multiple estimates available). The much cited association between income inequality (x-axis) and earnings mobility (y-axis) is demonstrated by the steep median regression line (with the correlation coefficient standing at 0.75).

Figure 6

But how robust is this relationship? Of the 45 estimates plotted in Figure 6, 20 involve the problematic TSTSLS approach. Moreover there are severe methodological problems, in terms of cross-national comparability, in another three (this includes New Zealand where the sample is not nationally representative, Singapore where children's reports of parental income are used, and the UK where earnings measured at a single point in time is used in conjunction with an instrumental variable). The right-hand panel of Figure 6 illustrates what happens to the 'Great Gatsby' relationship once such studies have been removed (leaving just 22 observations for 8 countries where the 'time-average' method has been applied)[5]. Evidence of a general relationship between income inequality and earnings mobility is now very weak. There is almost no gradient to the fitted regression line, with the correlation fluctuating between 0.11 and 0.48 (depending on whether the US is treated as an outlier). There are of course many possible explanations for this apparent lack of relationship, including attenuation bias, confounding from omitted variables or there possibly being a non-linear association. Nevertheless, it is clear that empirical evidence in support of strong statements like '*countries with more intergenerational mobility also tend to have lower point-in-time income inequality*' (Economic Report of the President 2012, p176) is actually rather limited.

Future work within the earnings mobility literature should take the issues raised in this paper into account. The same methodology *must* be applied to all data for each country under consideration. This is the only way that reliable and robust cross-country comparisons of earnings mobility will be produced. This is unfortunately not the case in most existing studies, meaning one is unable to distinguish genuine cross-country variation from statistical noise. We are consequently forced to conclude that, despite the large volume of papers discussing this topic over the last decade, relatively little is currently known about how intergenerational earnings mobility really compares across countries (while noting that there is reasonably strong evidence to suggest that the US is 'exceptional' – see Jantti et al 2006).

[5] Each study included meets three very basic criteria (i) there must be at least one earnings observation in the sons generation; (ii) there must be at least three earnings observations in the father's generation (to allow for time-averaging) (iii) the study must be nationally representative. We believe this represents a minimal standard for cross-national comparability of income mobility estimates, and note that there remain a number of other methodological and data issues (e.g. missing data, life-cycle bias) that could still lead to spurious differences being observed. Likewise we have not tackled serious issues surrounding the comparability of the income inequality measures plotted along the x-axis.

Our key recommendation is that future work should focus on producing more robust and reliable estimates of earnings mobility that can be legitimately compared across countries. Such endeavour would be much more valuable than researchers trying to explain why there are "differences" between certain countries, when these "differences" may not really exist.

References

Andrews, D. and Leigh, A. 2009. 'More inequality, less social mobility.' *Applied Economics Letters* 16: 1489 - 1492.

Aaronson, D. and Mazumder, D. 2008. 'Intergenerational Economic Mobility in the United States, 1940 to 2000.' *Journal of Human Resources* 43(1): 139 – 172.

Bauer, P. 2006. 'The intergenerational transmission of income in Switzerland: A comparison between natives and immigrants.' WWZ Discussion Paper 06/ 01. Accessed 30/04/13 from http://ideas.repec.org/p/bsl/wpaper/2006-01.html

Bidisha, S.; Das, A. and McFarlane, A. 2013. 'Microeconometric analysis of earnings mobility of immigrants and ethnic minorities: evidence from the UK.' *Journal for Labour Market Research* 46(1): 35-42.

Björklund, A. and Chadwick, L. 2003. 'Intergenerational income mobility in permanent and separated families.' *Economics Letters* 80: 239–246.

Björklund, A. and Jantti, M. 1997. 'Intergenerational income mobility in Sweden compared to the United States.' *The American Economic Review* 87(5): 1009 –1018.

Björklund, A. and Jantti, M. 2009. 'Intergenerational economic inequality.' Pp. 491 - 521 in The Oxford Handbook of Economic Inequality, edited by Wiemer Salverda, Brian Nolan and Timothy Smeeding. Oxford: Oxford University Press.

Blanden, J. 2013. 'Cross-national rankings of intergenerational mobility: A comparison of approaches from economics and sociology.' *Journal of Economic Surveys* 27(1):38–73.

Blanden, J.; Gregg, P. and Macmillian, L. 2013. ‘Intergenerational persistence in income and social class: the effect of within-group inequality.’ *Journal of the Royal Statistical Society (Series A)* 176(2): 541-563.

Center for American Progress. 2012. ‘How to slide down the ‘Great Gatsby Curve’: inequality, life chances, and public policy in the United States.’ Accessed 15/04/13 from http://www.americanprogress.org/wp-content/uploads/2012/12/CorakMiddleClass.pdf

Cervini – Pla, M. 2011. ‘Intergenerational earnings and income mobility in Spain.’ MPRA working paper 34942. Accessed 13/12/12 from http://mpra.ub.uni-muenchen.de/34942/

Corak, M. and Heisz, A. 1999. ‘The intergenerational income mobility of Canadian men: evidence from longitudinal tax data.’ *Journal of Human Resources* 34(3): 504–533.

Corak, M. 2012. ‘Inequality from generation to generation: the United States in comparison’, University of Ottawa working paper. Accessed 13/12/12 from http://milescorak.files.wordpress.com/2012/01/inequality-from-generation-to-generation-the-united-states-in-comparison-v3.pdf

Dearden, L.; Machin, S. and Reed, H. 1997. ‘Intergenerational mobility in Britain.’ *Economic Journal* 107(440):47-66.

Dunn, C. 2007. ‘The intergenerational transmission of lifetime earnings: evidence from Brazil.’ *B.E. Journal of Economic Analysis and Policy* 7(2): article 2.

Economic Report of the President. 2012. ‘Economic report of the President.’ Accessed 16/04/13 from http://www.nber.org/erp/ERP_2012_Complete.pdf

Ermisch, J. Jantti, M. and Smeeding, T. 2012. *Inequality from Childhood to Adulthood: A Cross-National Perspective on the Transmission of Advantage*. New York: Russell Sage Foundation.

Ferreira, S. and Veloso, F. 2006. ‘Intergenerational mobility of wages in Brazil.’ *Brazilian Review of Econometrics* 26(2): 181 - 211.

Fleiss, J. 1981. *Statistical methods for rates and proportions*. New York: John Wiley.

Fortin, N. and Lefebvre, S. 1998. ;Intergenerational income mobility in Canada.' Pp 51 -63 in *Labour Markets, Social Institutions, and the Future of Canada's Children,* edited by Miles Corak. Ottawa: Statistics Canada.

Gong, H.; Leigh, A. and Meng, X. 2012. 'Intergenerational income mobility in urban China.' *The Review of Income and Wealth* 58(3): 481 – 503.

Grawe, N. 2004. 'Intergenerational mobility for whom? The experience of high and low earning sons in international perspective.' Pp. 269 - 310 in *Generational Income Mobility in North America and Europe*, edited by Miles Corak. Cambridge: Cambridge University Press.

Haider, S. and Solon, G. 2006. 'Life-cycle variation in the association between current and lifetime earnings.' *American Economic Review* 96(4): 1308-1320.

Hussain, M; Munk, M. and Bonke, J. 2009. 'Intergenerational earnings mobilities: how sensitive are they to income measures?' *Journal of Income Distribution* 18(3):79-92.

Inoue, A., and Solon, G. 2010. 'Two-sample instrumental variables estimators.' *The Review of Economics and Statistics* 92(3): 557 - 561.

Jantti, M.; Røed, K.; Naylor, R.; Björklund, A.; Bratsberg, B.; Raaum, O.; Österbacka, E. and Eriksson, T. 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 discussion paper 1938. Accessed 30/04/13 from http://ftp.iza.org/dp1938.pdf

Jerrim, J. and Micklewright, J. 2012. 'Socioeconomic gradients in children's cognitive Skills: are cross-country comparisons robust to who reports family background?' Institute of Education, Department of Quantitative Social Science working paper 12/06. Accessed 01/05/13 from http://ideas.repec.org/p/qss/dqsswp/1206.html

Krueger, A. 2012. 'The rise and consequences of inequality.' Presentation made to the Center for American Progress, January 12th 2012 Washington. Accessed 13/12/12 from http://www.whitehouse.gov/sites/default/files/krueger_cap_speech_final_remarks.pdf

Landis, J. and Koch, G. 1977. 'The measurement of observer agreement for categorical data.' *Biometrics* 33(1): 159-174.

Lefranc, A. and Trannoy, A. 2005. ‘Intergenerational earnings mobility in France: is France more mobile than the US?’ *Annales d’Economie et Statistique* 78(2): 57 – 77.

Lefranc, A.; Ojima, F. and Yoshida, T. 2010. ‘The intergenerational transmission of income and education: A comparison of Japan and France.’ Pp. 229 - 253 in Quality and Inequality in Education: Cross-national Perspectives, edited by Jaap Dronkers. Springer.

Lefranc, A. 2011. ‘Educational expansion, earnings compression and changes in intergenerational economic mobility: evidence from French cohorts, 1931-1976.’ Paper presented to the ESPE 2010 conference. Accessed 13/12/12 from http://espe.conference-services.net/resources/321/2017/pdf/ESPE2010_0583_paper.pdf

Lefranc, A.; Ojima, F. and Yoshida, T. 2013. ‘Intergenerational earnings mobility in Japan among sons and daughters: levels and trends.’ *Journal of Population Economics* DOI: 10.1007/s00148-012-0464-2

Leigh, A. 2007. ‘Intergenerational economic mobility in Australia.’ *The B.E. Journal of Economic Analysis & Policy*: 7(2): article 6.

Looker, D. 1989. ‘Accuracy of proxy reports of parental status characteristics.’ *Sociology of Education* 62(4): 257-276.

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.

Mocetti, S. 2007. ‘Intergenerational earnings mobility in Italy.’ *The B.E. Journal of Economic Analysis & Policy* 7(2): article 5.

Murphy, K. and Topel, R. 1985. ‘Estimation and inference in two-step econometric models.’ *Journal of Business and Economic Statistics* (3)4: 370 – 379.

Ng, I. 2007. ‘Intergenerational Income Mobility in Singapore.’ *The B.E. Journal of Economic Analysis & Policy* 7(2): Article 3.

Ng, Irene, Xiaoyi Shen and Kong Weng Ho. 2009. 'Intergenerational earnings mobility in Singapore and the United States. *Journal of Asian Economics* 20(2): 110-19.

Nicoletti, C. and Ermisch, J. 2008. 'Intergenerational earnings mobility: changes across cohorts in Britain.' *B.E. Journal of Economic Analysis and Policy* 7(2): article 9.

Nuñez, J. and Miranda, L. 2011. 'Intergenerational income and educational mobility in urban Chile.' *Estudios de Economia* 38(1): 195 – 221.

OECD. 2011. 'Divided we stand: why inequality keeps rising' Accessed 29/04/13 from http://www.oecd.org/els/soc/49170768.pdf

Piraino, P. 2007. 'Comparable estimates of intergenerational income mobility in Italy.' *The B.E. Journal of Economic Analysis & Policy* 7(2): article 1.

Seng, Y.C. 2012. 'Intergenerational Income Mobility in Singapore.' Singapore Ministry of Finance Research Report. Accessed 30/04/13 from http://app.mof.gov.sg/data/cmsresource/Feature%20Articles/Intergenerational%20Income%20Mobility%20In%20Singapore_13%20Jan%202012.pdf

Solon, G. 1992. 'Intergenerational income mobility in the United States.' *The American Economic Review* 83(3): 393 – 408.

The Economist. 2012. 'Having your cake: less inequality does not need to mean less efficiency.' Accessed 30/04/13 from http://www.economist.com/node/21564421

The New York Times. 2012. 'The Great Gatsby Curve.' Accessed 30/04/2013 from http://krugman.blogs.nytimes.com/2012/01/15/the-great-gatsby-curve/

The Sutton Trust. 2013. 'Moving up the Great Gatsby Curve.' Accessed 30/04/13 from http://www.suttontrust.com/news/news/moving-up-the-great-gatsby-curve/

The White House. 2013. 'What is the Great Gatsby Curve?' Accessed 01/07/2013 from http://www.whitehouse.gov/blog/2013/06/11/what-great-gatsby-curve

Ueda, A. 2009. 'Intergenerational mobility of earnings and income in Japan.' *The B.E. Journal of Economic Analysis & Policy* 9(1): Article 54.

Ueda, A. and Sun, F. 2013. 'Intergenerational economic mobility in Taiwan.' Institute of Research in Contemporary Political and Economic Affairs working paper 1306. Accessed 01/07/2013 from http://www.waseda-pse.jp/file/genseiken/WP1306%EF%BC%88%E4%B8%8A%E7%94%B0%EF%BC%89%E4%BF%AE%E6%AD%A3%E5%BE%8C.pdf

Vogel, T. 2008. Reassessing intergenerational mobility in Germany and the United States: the impact of differences in lifecycle earnings patterns. SFB Discussion Paper 2006-055, Humbolt University, Berlin. Accessed 30/04/2013 from http://www.wiwi.hu-berlin.de/professuren-en/vwl/wtm2/Employees/vogel/vogelmobility

Wilkinson, R. and Pickett, K. 2010. *The spirit level: why equality is better for everyone*. Penguin Books.

Wooldridge, J. 2002. *Econometric analysis of cross-section and panel data*. Massachusetts: MIT Press.

Table 1. Number of earnings observations for the BHPS cohort members

Number of earnings observations	%
5	8
6	8
7	8
8	13
9	9
10	13
11	4
12	4
13	4
14	4
15	4
16	5
17	5
18	13
n	**3,080**

Notes:

Source: Author calculations using the BHPS dataset

Table 2. Comparison of observed and imputed long-run earnings

(a) United Kingdom

	Observed	**Model 1**	**Model 2**	**Model 3**	**Model 4**
R-Squared	-	0.30	0.39	0.42	0.49
Variance	0.22	0.05	0.08	0.09	0.12
Correlation between imputed and observed long-run earnings	-	0.35	0.46	0.48	0.53
Kappa statistic	-	0.13	0.14	0.20	0.23
Percentage correct	-	35	36	40	42
Sample size (BHPS)	2,506	2,489	2,479	2,462	2,467
Sample size (LFS)	-	69,548	69,548	69,548	69,548

(b) United States

	Observed	**Model 1**	**Model 2**	**Model 3**	**Model 4**
R-Squared	-	0.32	0.33	0.37	0.43
Variance	0.62	0.12	0.11	0.15	0.23
Correlation between imputed and observed long-run earnings	-	0.48	0.41	0.51	0.54
Kappa statistic	-	0.15	0.12	0.23	0.28
Percentage correct	-	38	35	43	47
Sample size (NLSY79)	3,624	3,624	3,624	3,624	3,624
Sample size (CPS)	-	529,414	529,414	529,414	529,414

Notes:

i. Source: Authors’ calculations using BHPS, LFS, NLSY79 and CPS datasets

ii. R-squared is in reference to the first-stage prediction equation

iii. Model 1 – 4 indicates which TSTSLS imputation specification has been used. See section 3 for further details.

iv. Results presented refer to men only. For females in the US, see Appendix B

Table 3. Cross-tabulation of observed and predicted earnings quartile

(a) Perfect agreement

		Predicted quartile			
		Bottom	2nd	3rd	Top
Observed quartile	Bottom Quartile	100	0	0	0
	2nd Quartile	0	100	0	0
	3rd Quartile	0	0	100	0
	Top Quartile	0	0	0	100

(b) Random assignment

		Predicted quartile			
		Bottom	2nd	3rd	Top
Observed quartile	Bottom Quartile	25	25	25	25
	2nd Quartile	25	25	25	25
	3rd Quartile	25	25	25	25
	Top Quartile	25	25	25	25

(c) UK

		Predicted quartile				
		Bottom	2nd	3rd	Top	n
Observed quartile	Bottom Quartile	44	29	17	10	617
	2nd Quartile	34	31	23	12	630
	3rd Quartile	22	23	33	22	614
	Top Quartile	5	13	30	52	601

(a) US (males)

		Predicted quartile				
		Bottom	2nd	3rd	Top	n
Observed quartile	Bottom Quartile	53	24	17	6	1,115
	2nd Quartile	30	31	27	12	872
	3rd Quartile	23	22	33	22	745
	Top Quartile	7	13	29	51	668

Notes:

i. Figures refer to row percentages.
ii. The final column (n) refers to unweighted sample sizes
iii. Associated kappa statistics are 0.20 (England) and 0.23 (US)
iv. Source: Authors' calculations using TSTSLS prediction model 3 (see section 3 for further details).

Table 4. Relationship between prediction error and selected characteristics

Panel A. Social class

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Social class (Ref: Senior officials)								
Professional occupations	-0.236*	0.096	-0.018	0.099	-0.012	0.099	-0.06	0.097
Associate professionals	-0.447*	0.089	-0.067	0.091	-0.072	0.091	-0.115	0.093
Administrative occupations	-0.641*	0.118	0.296*	0.121	0.269*	0.121	0.248*	0.126
Skilled trade occupations	-0.521*	0.075	0.201*	0.082	0.198*	0.08	0.164*	0.077
Personal service occupations	-1.008*	0.163	0.297*	0.15	0.263	0.16	0.262	0.158
Sales and customer service	-1.203*	0.15	-0.087	0.139	-0.105	0.147	-0.164	0.15
Process, plant and machine operatives	-0.557*	0.083	0.417*	0.087	0.382*	0.086	0.380*	0.085
Elementary occupations	-0.756*	0.084	0.386*	0.092	0.415*	0.09	0.361*	0.088

Panel B. Education

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Education (Ref: Degree or higher)								
Other higher education	0.499*	0.103	0.159	0.101	0.481*	0.101	0.433*	0.103
A-Level	0.414*	0.081	-0.039	0.08	0.358*	0.08	0.293*	0.081
O-Level	0.285*	0.074	-0.125	0.075	0.299*	0.075	0.266*	0.077
CSE	0.343*	0.104	-0.387*	0.118	0.236*	0.115	0.234*	0.109
None	0.320*	0.084	-0.073	0.08	0.424*	0.081	0.417*	0.083

Panel C. Whether the youth expects to stay in school

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Youth stay in school (Ref: No)								
Yes	0.262*	0.102	0.239*	0.094	0.133	0.093	0.138	0.089

Panel D. Industry

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Industry (Ref: Wholesale and retail)								
Agriculture and Fishing	-0.244	0.225	0.376	0.206	0.219	0.23	0.496*	0.216
Mining	0.825*	0.178	1.127*	0.186	1.087*	0.196	0.655*	0.178
Manufacturing	0.386*	0.082	0.524*	0.083	0.473*	0.083	0.217*	0.086
Utilities	0.146	0.324	0.32	0.355	0.272	0.353	-0.143	0.367
Construction	0.278*	0.1	0.524*	0.106	0.469*	0.105	0.17	0.105
Hotels and restaurants	-0.523*	0.182	-0.547*	0.191	-0.526*	0.189	-0.267	0.161
Transport / communications	0.382*	0.1	0.751*	0.103	0.711*	0.103	0.504*	0.104
Finance	0.823*	0.16	0.653*	0.15	0.634*	0.152	0.049	0.172
Real Estate / business	0.341*	0.103	0.392*	0.102	0.302*	0.103	0.035	0.108
Public administration and defence	0.444*	0.111	0.511*	0.1	0.457*	0.103	0.259*	0.106
Education	0.340*	0.116	0.489*	0.11	0.314*	0.116	0.218	0.122
Health and social work	0.053	0.136	0.327*	0.135	0.215	0.14	0.098	0.141
Other personal service	-0.098	0.132	0.035	0.155	-0.05	0.15	-0.053	0.154

Notes:

i. Results from a series of bivariate regressions.

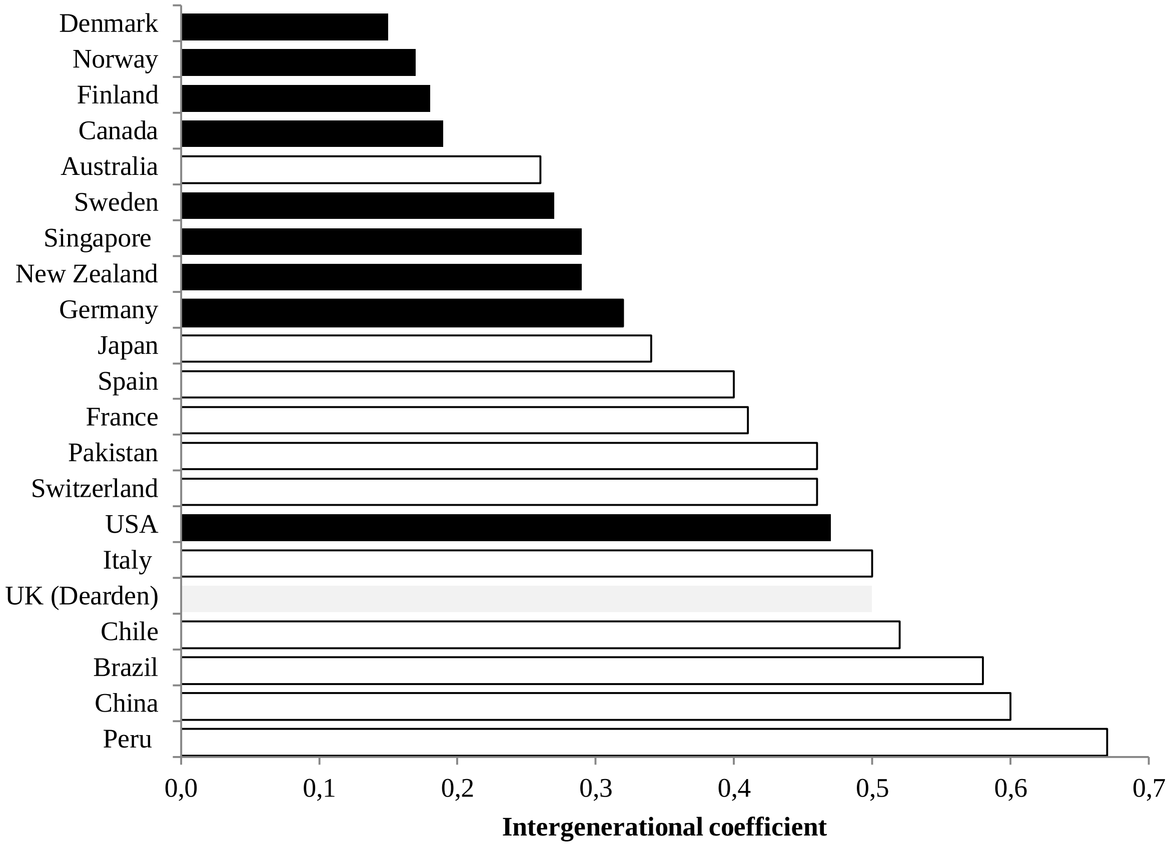
ii. * indicates statistical significance at the five percent level.

iii. All figures refer to standard deviation differences in relation to the reference group.

iv. Model 1 – model 4 refer to the different TSTSLS imputation model used.

v. Source: Authors’ calculations using the BHPS dataset.

Figure 1. International comparison of intergenerational earnings mobility

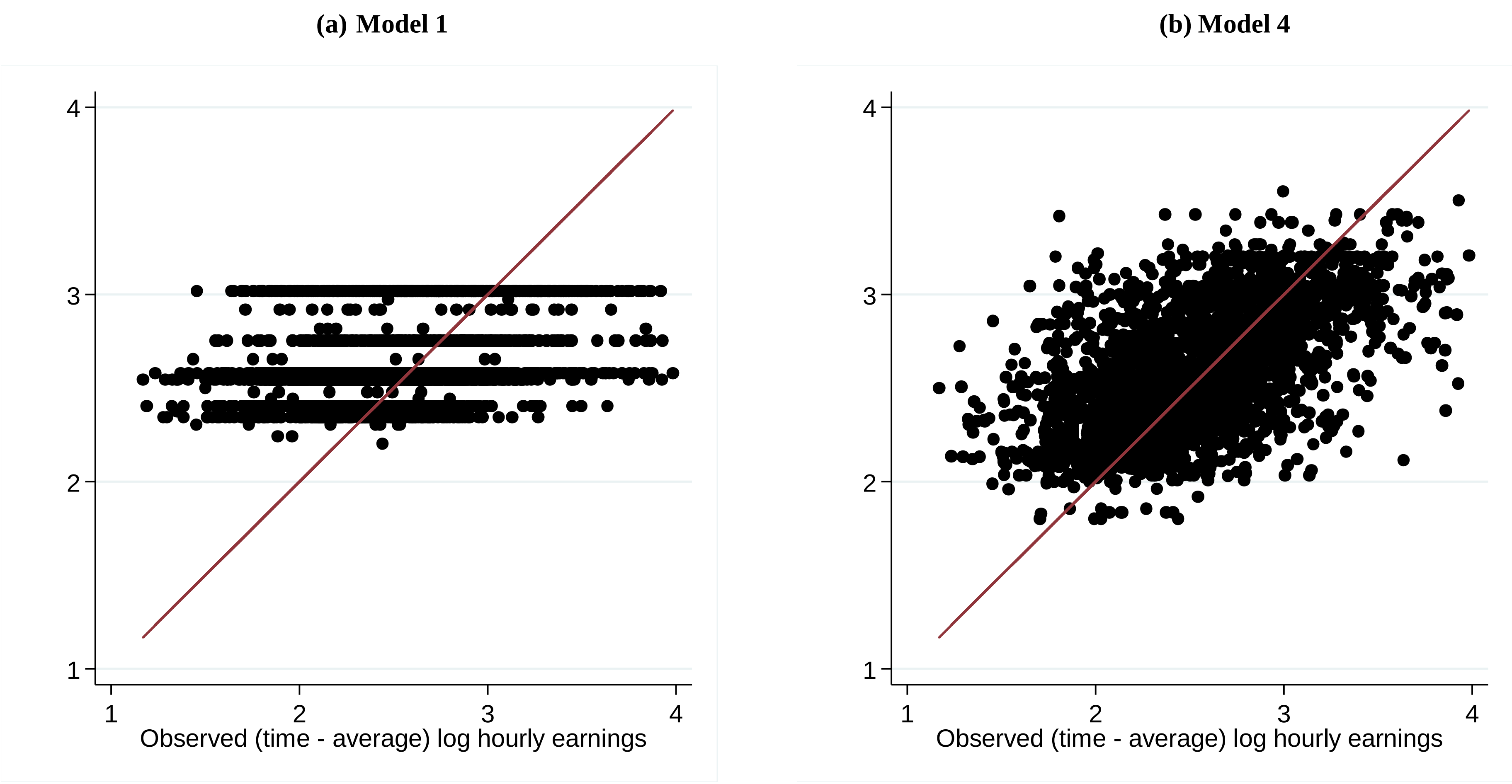


Notes:

i. Estimates drawn from Corak (2012). Argentina has been excluded as the source could not be found. The estimate for Singapore found in Corak (2012) is based upon Ng (2009). However the Ng (2009) study relied upon children's reports of parental income and ad-hoc adjustments to the estimated income elasticity. We have chosen to replace this with a more recent study by Seng (2012) which we believe to be more methodologically robust.

ii. The colour of the bar indicates the estimation strategy used. Black bars indicate where OLS regression with time-average parental earnings has been used. White bars are where the TSTSLS approach has been applied. Estimates for UK based upon a (single sample) instrumental variable approach and so shaded in light grey.

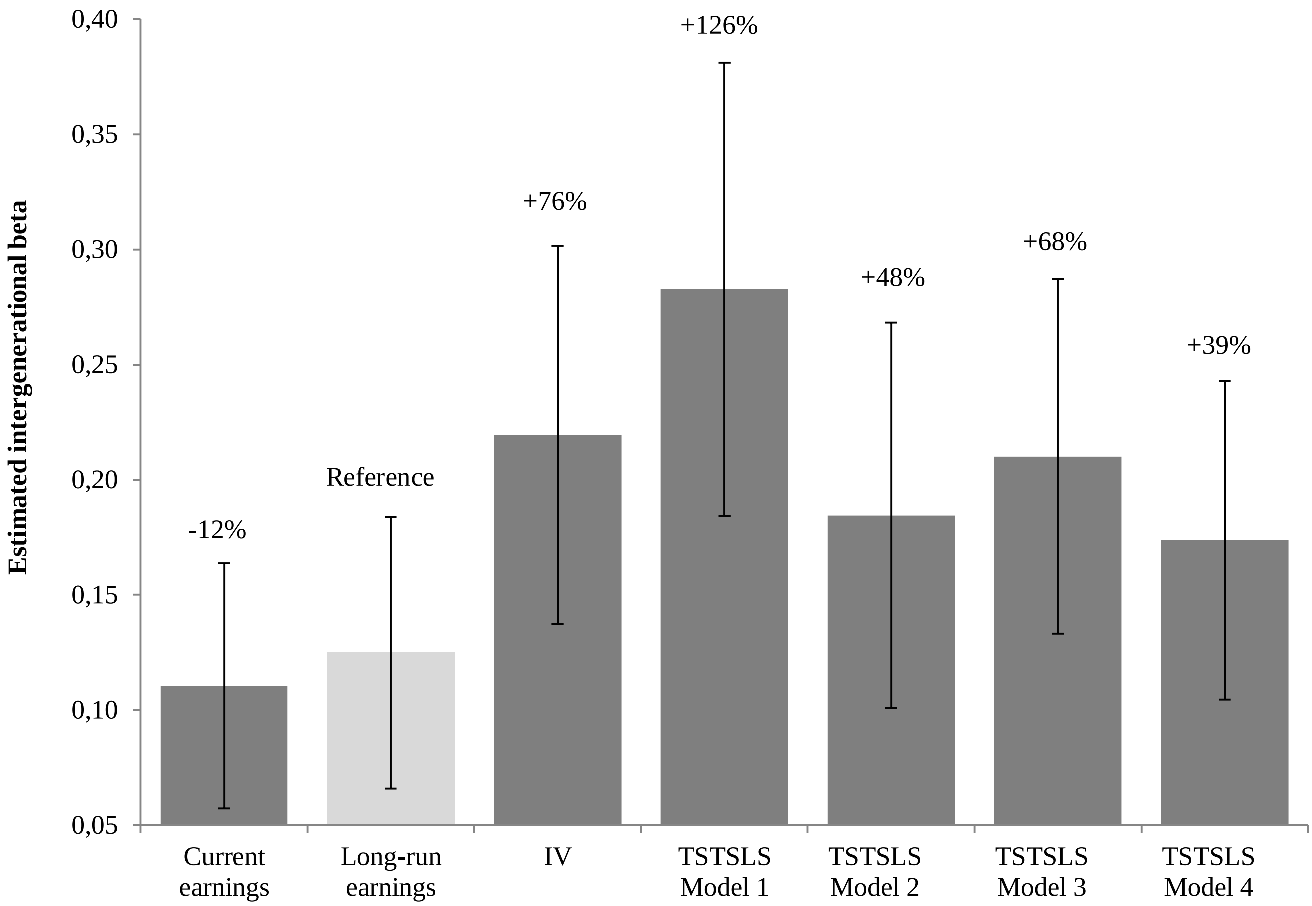
Figure 2. The correlation between imputed and observed long-run earnings



Notes:

i. Model 1 is where parental education is the only imputation variable used. Model 4 is where education and detailed occupational data are used.
ii. The 45^{o} line indicates where observed and imputed long-run earnings are in perfect agreement.
iii. The correlation equals 0.35 in the left hand panel and 0.53 in the right hand panel.

Figure 3. Estimates of the association between father's earnings and children's educational plans



Notes:

i. Estimates based upon linear probability model described in section 3. Response coded 1 if child plans to enter post-secondary education, 0 otherwise.

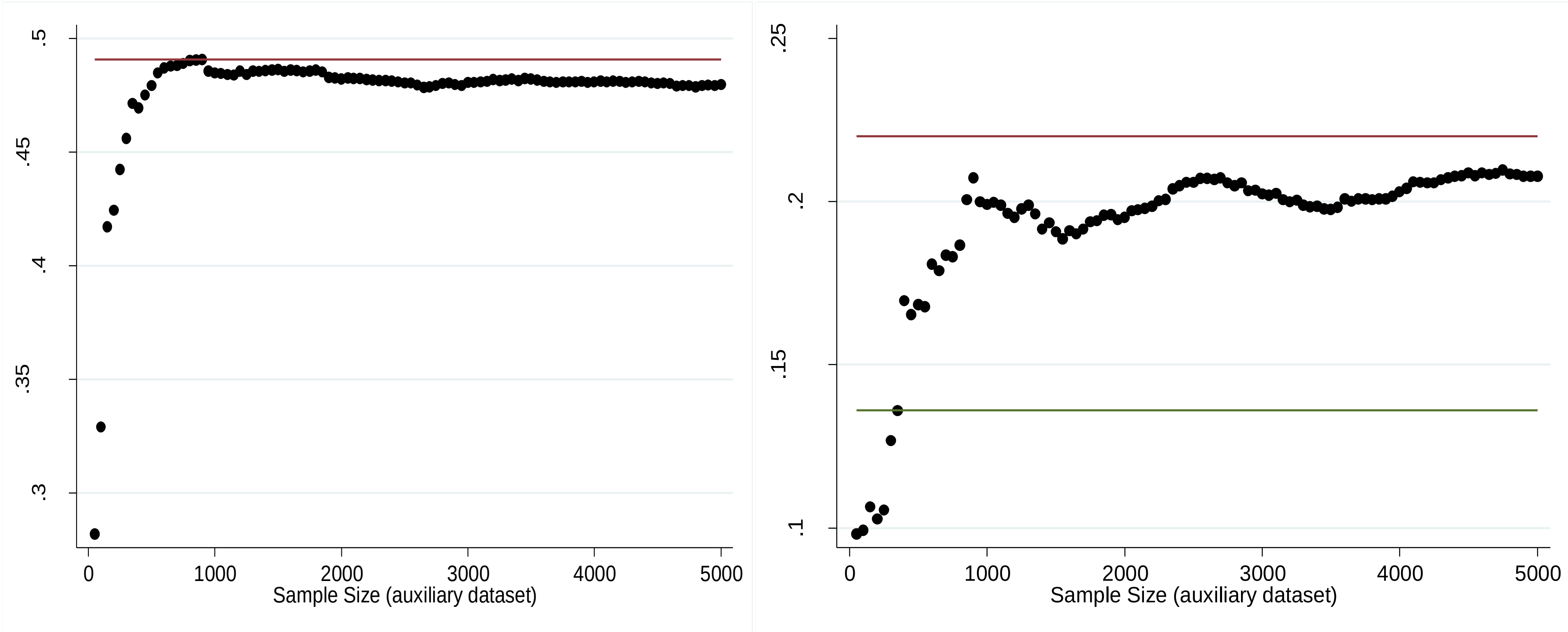
ii. Figures on the y-axis illustrate the percentage point change in the probability of a child expecting to enter post-16 education for a one log-unit change in father's hourly earnings.

iii. The four bars on the right are based upon TSTSLS predictions of long-run earnings (see section 3).

iv. Percentages above the bars refer to the percentage under or over estimation relative to the observed long-run earnings measure (reference group).

Figure 4. Correlation between imputed and observed long-run earnings using different auxiliary dataset sample sizes (imputation model 3)

(a) Correlation (imputed and observed) **(b) Regression estimates**



Notes:

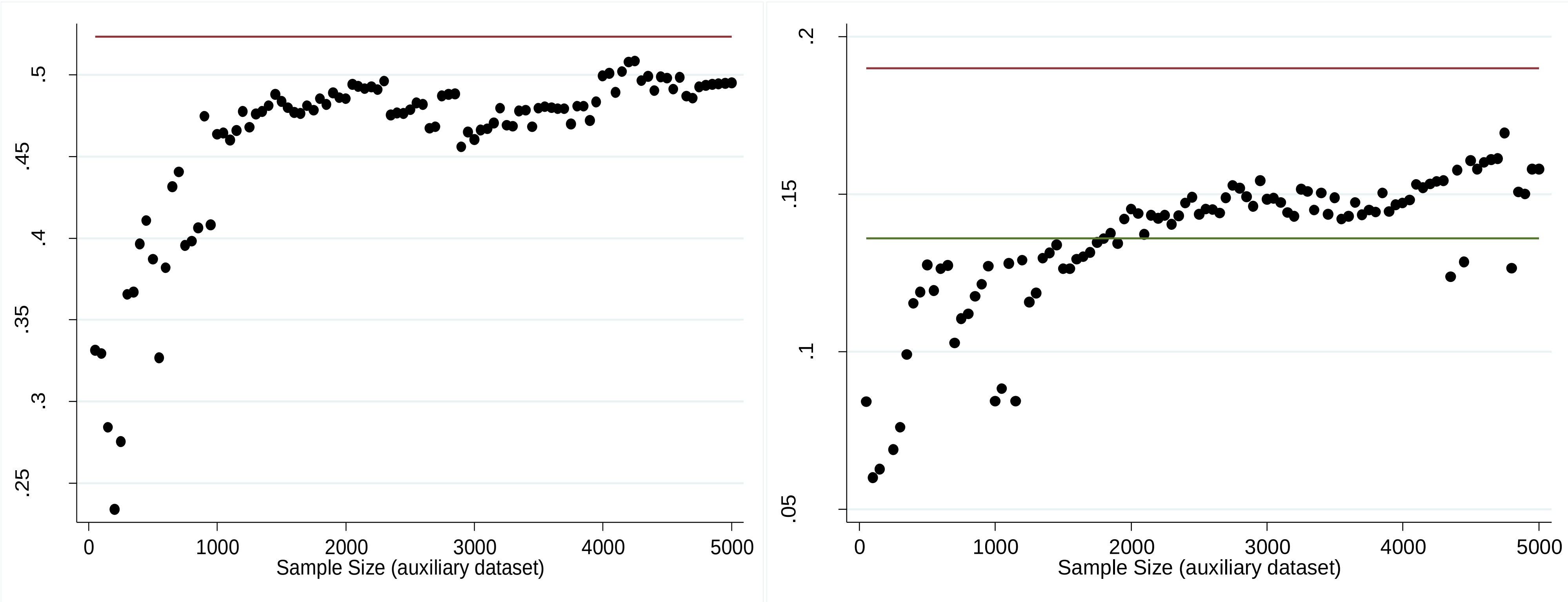
i. Panel (a) illustrates the association between the auxiliary dataset sample size and the association between imputed and observed earnings. The horizontal line at the top of the graph illustrates the estimated correlation coefficient when all 69,548 LFS observations have been used.

ii. Panel (b) refers to the association between imputed father's earnings and children's schooling intentions. The uppermost (red) line illustrates the estimate when all LFS observations were used. The lower (green) line is the estimate when observed time-average father's earnings have been used.

iii. Source: Authors' calculations using the BHPS dataset, applying TSTSLS imputation model 3

Figure 5. Correlation between predicted and actual long-run earnings using different auxiliary dataset sample sizes (imputation model 4)

(a) Correlation (imputed and observed) **(b) Regression estimates**

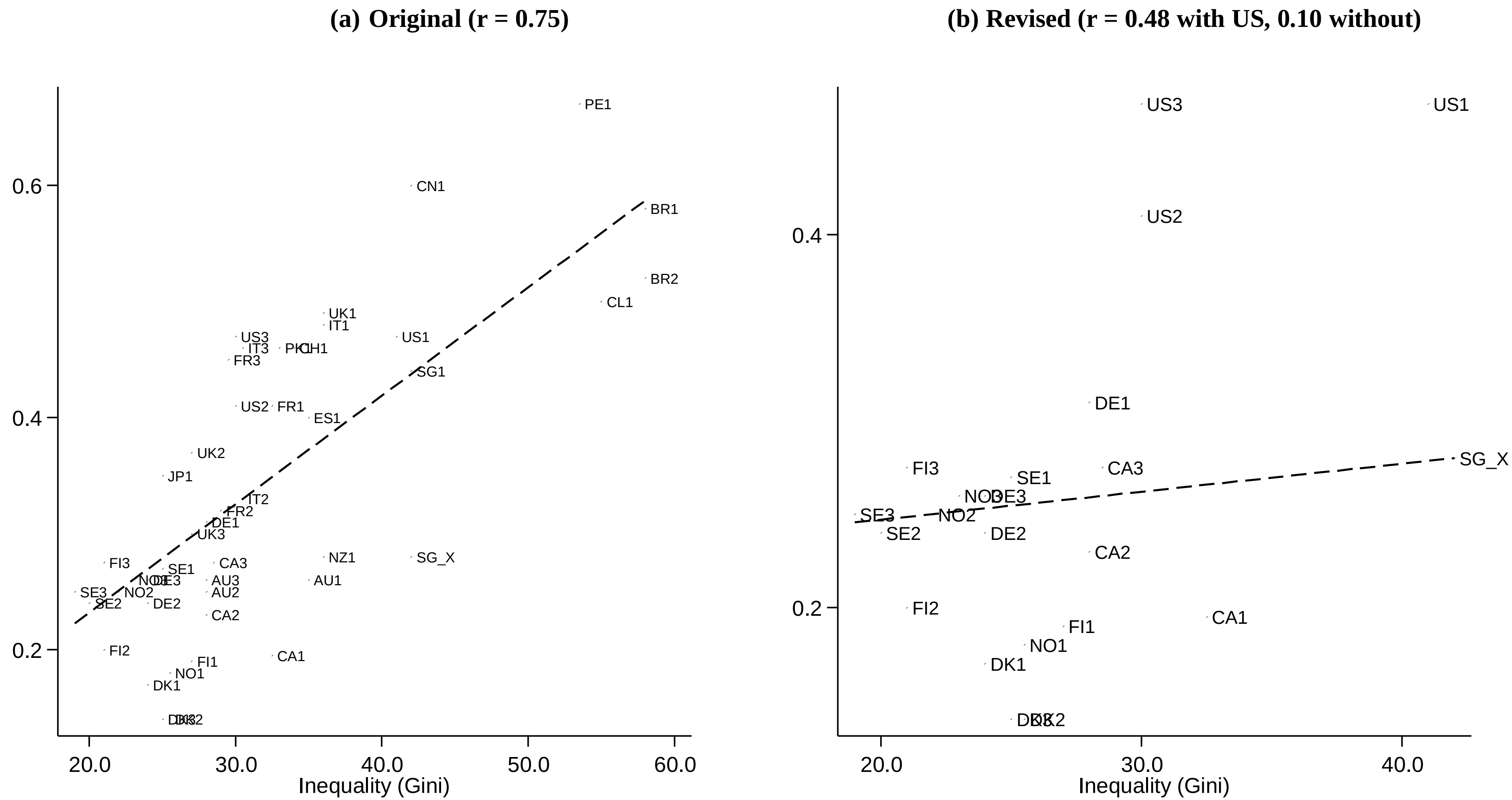


Notes

i. See notes to Figure 4 above

ii. Source: Authors' calculations using the BHPS dataset, applying TSTSLS imputation model 4

Figure 6. The 'Great Gatsby Curve': The cross-national link between earnings inequality and intergenerational earnings mobility



Note: Two letter country codes have been used (see http://www.worldatlas.com/aatlas/ctycodes.htm). The left hand figure includes all prefered income inequality and earnings elasticity estimates from Corak (2012), Björklund and Jantti (2009), Blanden (2013). A number has been appended after each country: 1 = Corak (2012), 2 = Björklund and Jantti (2009) and 3 = Blanden (2013). SG_X refers to Seng (2012), with the gini taken from Corak (2012). Dashed line refers to quantile (median) regression estimate.

Appendix A. Intergenerational mobility papers imputing father's earnings using TSTSLS

	Country	Database (Main data)	Sample size (Main data)	Offspring' income	Database (Auxiliary)	Sample size (Auxiliary)	Imputer variables and 1st stage R^2
Aaronson and Mazumder (2008)	United States	1940 to 2000 census data 1940-1970: 1% sample 1980-2000: 5% sample	Men, 25-54 years old, born btw 1921 and 1975.	Earnings	1940 to 2000 census data	1940-1970: 1% sample 1980-2000: 5% sample	State of birth R^2: Not reported
Andrews and Leigh (2009)	16 countries	1999 International Social Survey Program	Not reported	Son's log hourly wage.	1999 International Social Survey Program	Not reported	192 Occupation dummies (off-spring reported) R^2: Not reported
Bidisha (2013)	United Kingdom	1991-2005 British Household Panel Survey	3.823	Average log wages of full time workers and earnings of self-employees over the panel	1991-2005 British Household Panel Survey	935	Education (3 dummies), occupation (3 dummies); immigrant status; ethnic group; professional level (4 dummies); cohort (2 dummies); Hope-Goldthrope score; R2=0.323
Björklund and Jantti (1997)	Sweden and USA	1991 Swedish Level of Living Survey; Panel Survey of Income Dynamics	Sweden: 327 US: Not reported	Annual log earnings and capital market income	1968 Swedish Level of Living Survey	Sweden: 540 US: Not reported	Education (2 dummies); Occupation (8 dummies); Living in Stockholm Note: Children reports R2: Not reported
Cervini-Pla (2012)	Spain	2005 Encuesta de Condiciones de Vida	2,836 sons 1,696 daughters	Annual log earnings of sons. For daughters: log family income.	1980-81 Encuesta de Presupuestos Familiares	5, 929	Education (6 dummies) Occupation (9 dummies). R^2: 0.40

	Country	Database (Main data)	Sample size (Main data)	Offspring' income	Database (Auxiliary)	Sample size (Auxiliary)	Imputer variables and 1st stage R^2
Dunn (2007)	Brazil	1996 Pesquisa Nacional por Amostra de Domicilios	14,872	Annual log "earnings from all jobs".	PNAD 1976	37,396	Father's education (10 categories) R^2: Not reported.
Ferreira and Veloso (2006)	Brazil	1996 Pesquisa Nacional por Amostra de Domicilios	25,927	Log wages.	1976, 1981, 1986 and 1990 PNAD	59,340	Father's education (7 dummies) Father's occupation (6 dummies) R^2: Not reported
Fortin and Lefebvre (1998)	Canada	General Social Surveys 1986 and 1994	Father – son: 3,400 (1986) 2,459 (1994) Father-daughter: 2,474 (1986) 2.308 (1994)	Annual income	General Social Surveys 1986 and 1994	Circa 500,000 each year	Father's occupation (15 groups) R^2: Not reported
Gong et al. (2012)	China	2004 Chinese Urban Household Education and Employment Survey	5,475	Annual log income.	1987 to 2004 Urban Household Income and Expenditure Survey	Varies depending on UHIES sample.	Father's education; Father's occupation; Industry. R^2: Not reported
Lefranc et al. (2010)	France and Japan	1985,1995,2005 Social Stratification Survey for Japan. 1985, 1993, 2003 Formation, Quailification, Profession for France	Japan: 987 France 13,487	Japan: Individual primary income (labor + assets) before tax or transfer. France: Annual earnings from labor.	Japan:Social Stratification Survey France: Formation, Quailification, Profession for France	Fathers btw 25 and 54, in Japan. Fathers btw 24 and 60 in France.	Linking variables: Japan: year of birth; 3 educational levels and occupation. R2: N.R. France: year of birth; 6 levels of education. R2: N.R.
Lefranc (2011)	France	1970, 1977, 1985, 1993 and 2003 Formation, Quailification, Profession	29,415	Annual wages	1964, 1970, 1977, 1985, 1993 and 2003 Formation, Quailification, Profession	48,245	Father's education (6 groups). Note: Offspring reports R^2: Not reported

	Country	Database (Main data)	Sample size (Main data)	Offspring' income	Database (Auxiliary)	Sample size (Auxiliary)	Imputer variables and 1st stage R^2
Lefranc and Trannoy (2005)	France and USA	French Education-Training-Employment	1977: 2,023 1985: 2,114 1993: 771	Wages		2,364 – 6,488 depending on the year.	Father's education (8 groups) Father's occupation (7 groups) Note: Offspring reported. R^2: 0.49 - 0.54
Lefranc et al. (2011)	Japan	Japanese Social Stratification and Mobility Survey	2,273	Gross individual income	Japanese Social Stratification and Mobility Survey	7,170	Father education (3 groups) Father occupation (8 groups) Firm size (2 groups) Self-employment; Residential area (3 groups). R^2: 0.46
Leigh (2007)	Australia	4 different surveys: 1965, 1973, 1987 and 2004.	1965: 946 1973: 1871 1987: 243 2004: 2115	Hourly wages	4 different surveys: 1965, 1973, 1987 and 2004.	1965: 946 1973: 1871 1987: 243 2004: 2115	Father's occupations (78 to 241 groups depending on survey). Offspring reported. R^2: Not reported
Mocetti (2007)	Italy	Survey of Household Income and Wealth	3,200	Gross income from all sources but financial assets.	Survey of Household Income and Wealth	4,903	Father's education (5 groups; Work status (5 groups); employment sector (4 groups); geographical area (3 groups). R^2: 0.30
Nicoletti and Ermisch (2008)	UK	British Household Panel Survey	8,832	31-45 years old sons, with positive income (employed or self-employed) in at least one wave of the panel	BHPS	896	Father's occupation (4 groups) Father's education (5 groups). R^2: 0.31
Nuñez and Miranda (2010)	Chile	Caracterización Socioeconómica - 2006	11,186	25 to 40 years old log earnings of sons working at least 30hs x week	Caracterización Socioeconómica - 1987 and 1990	1987: 19,192 1990: 20,378	Father's occupation (4 groups) Father's education (5 groups). R^2: 0.29 - 0.37.

	Country	Database (Main data)	Sample size (Main data)	Offspring' income	Database (Auxiliary)	Sample size (Auxiliary)	Imputer variables and 1st stage R^2
Nuñez and Miranda (2011)	Chile (Greater Santiago)	2004 Employment and Unemployment Survey for the Greater Santiago	649	Log income	Employment and Unemployment Survey for the Greater Santiago	1,736 - 2,700 (depending on the year)	Father's education (3 groups) Father's occupation (5 groups) R^2: 0.48 – 0.66
Piraino (2007)	Italy	Survey of Household Income and Wealth	1,956	Gross income from all sources bar financial assets.	Survey of Household Income and Wealth	953	Father's education (5 groups); work status (4 groups); employment sector (4 groups); geographical area (2 groups) $R^2 = 0.33$.
Ueda (2009)	Japan	Japanese Panel Survey of Consumers	1,114 married sons; 906 single daughters; 1,390 married daughters	Gross annual earnings and income from all sources.	Japanese Panel Survey of Consumers		Father's years of education; Father's occupation and firm size (7 groups). R^2: Not reported.
Ueda (2012)	Korea and Japan	Korea: 1998 Labor Income Panel Japan: 1993-2006 Panel Survey of Consumers	Both countries: size varies depending on civil status of the sons and daughters	Annual earnings	Korea: 1998 Labor Income Panel Japan: 1993-2006 Panel Survey of Consumers	Korea: Fathers btw 25 and 54 Japan:	Korea: education and occupation Japan: parental income . R^2:Not reported
Ueda and Sun (2013)	Taiwan	2004-2006 Panel Study of Family Dynamics	745	Annual income	1983 Survey of Family Income and Expenditure in Taiwan Area	745?	Father's education (6 groups); Father's occupation (11 groups). R^2: Not reported.

Appendix B. Supplementary analysis using US data

Our empirical analysis is supplemented using two other large, high quality US datasets: The National Longitudinal Study of Youth 1979 cohort (NLSY79) and the Current Population Survey (CPS). The former acts as the 'main' sample and the latter as the 'auxiliary' sample. We have chosen these datasets as they meet the same criteria as the UK datasets (large sample size, detailed information available, widespread use, public accessibility, and the availability of child supplement data).

The National Longitudinal Survey of Youth 1979 (Main dataset)

The NLSY79 is a nationally representative American dataset that began by sampling 12,686 15 – 22 year olds in 1979. Cohort members were interviewed annually up to 1994, and bi-annually thereafter. The latest wave was conducted in 2010 when cohort members were between 46 and 53 years old. Throughout our analysis we include only the 7,544 individuals who took part in the latest survey wave, and apply the 2010 sampling weight to adjust for non-random non-response.

Table B1

Table B1 illustrates the number of earnings observations available for male and female cohort members after the age of 25[6]. Approximately 99% of male and female cohort members have five or more earnings observations, with the majority having ten or more. The 1% of observations with less than five earnings measures available are dropped from the analysis. This leaves a total of 7,475 observations (3,624 for males and 3,851 for females). A 'permanent' measure is then created for the remaining cohort members by averaging across all annual earnings reports after age 25. We call this X_{NLSY}. All earnings data has been adjusted to 2010 prices. Again it is important to note that the variable we have derived refers to long-run average *earnings* (labour market income only) for the reasons explained in the main text.

As part of the NLSY, respondents have also been asked detailed questions about their current occupation and educational attainment. These are the key imputation variables that will be used in our application of the TSTSLS technique. The former has been coded according to the detailed three-digit census occupational classification

[6] Note that we restrict earnings observations to post age 25 as there are likely to be non-trivial random fluctuations (i.e. a large transitory component in earnings) before this point.

system. We re-code respondents' occupation and highest education level held in 2010 into the following groups (consistent with the literature):

Education: (i) Less than high school; (ii) High school; (iii) Some college no degree; (iv) Associate degree; (v) Bachelor degree; (vi) Beyond bachelor degree

Social class: (i) Operatives and labourers; (ii) Production, crafts and repairs; (iii) Farming, forestry and fishing; (iv) Service; (v) Technical, sales and administrative; (vi) Managerial and professional; (vi) Non-occupational responses.

The NLSY79 also includes 'child supplement data'; the survey organisers have attempted to collect information about the children of all *female* cohort members. A battery of child assessments has been administered biennially since 1986. These surveys assessed the child's development using nationally normed tests. This includes the Peabody Individual Achievement Test (PIAT) in Math, which is a wide-ranging measure of achievement in mathematics for children aged five and over. It consists of 84 multiple-choice items of increasing difficulty. The norming sample has a mean of 100 and a standard deviation of 15. In order to test the robustness of intergenerational associations we use the PIAT Math scores (from now on math test scores) of children under age 15. We are able to link a total of 3,030 children to mothers who have at least five labour market earnings observations available and who took part in the final NLSY79 wave. The NLSY79 child sampling weight is applied during this part of the analysis.

Current Population Survey (auxiliary dataset)

We use numerous rounds of the CPS March annual supplement as our auxiliary dataset. This is cross-sectional data, collected by the United States Census Bureau, and has been designed to provide a nationally representative snapshot of the US labour force once every year. We pool information across all CPS waves between 2000 and 2010 to ensure a large sample size. The sample is then restricted to respondents who were between the ages of 18 and 65. This leaves a total of 1,366,340 observations (658,194 observations for males and 708,146 for females) in our analysis. The person weight, which helps to compensates for non-response and grosses the sample up to population estimates, is applied throughout.

As part of the CPS, respondents were asked a series of questions about their earnings from work and other sources of income (e.g. social security benefits, interest and dividends, business income). The survey organisers have recoded this information into two variables: 'total earnings from work' and 'total household income'. It is important to note that the CPS earnings / income data is not entirely free of error (Bollinger 1998). However, this is not a major concern in this paper; cross-sectional labour force datasets with self-reported earnings have been widely used in the TSTSLS applications we are trying to mimic. Hence this simply reflects one of the actual empirical difficulties researchers face when applying this methodology. As with the NLSY, all earnings data is adjusted to real 2010 prices. The CPS also contains detailed information on respondents' highest level of education and their current occupation. We convert this into the same broad education and social class groups as described for the NLSY.

The CPS is used to impute long-run earnings into the NLSY following the TSTSLS approach as we did with the UK datasets in the main text (the estimates from the former models can be found in Appendix Table B2). The only differences with respect to the UK analysis are: (i) the sample is now composed of males and females, so the results obtained are presented by gender; (ii) the long-run earnings and imputed earnings variables refer to annual rather than hourly earnings; (iii) in order to test the robustness of intergenerational associations, we investigate the link between children's math test scores (included within the NLSY child supplement data) and mother's earnings using OLS regression.

The quality of the TSTSLS earnings imputations

In Appendix Table B3 the comparison of imputed and observed long-run earnings are presented for males and females. As with the UK results, we present the R^2 values from our first-stage prediction equations in the top row of Table B3. Again these values typically fall between 0.30 and 0.40, highlighting a weak level of statistical 'fit'.

Appendix Table B3

Information on the variance of imputed and observed long-run earnings is presented in the second row. Regardless of the first-stage imputation model used, the variance of long-run earnings is significantly underestimated. The variance of observed (time – average) long-run earnings is approximately 0.60 log-points for males and females. This value falls between 0.12 and 0.23 log-points when using the various different TSTSLS imputation models. Underestimation of the long-run earnings variance is once again in the region of 40 to 80 percent.

Turning to the strength of the association between imputed and observed measures of long-run earnings, estimated correlation coefficients can be found in the third row of Table B3. The correlation between observed and predicted long-run earnings is modest, falling somewhere between 0.4 and 0.5. As with the UK data, when we focus on the TSTSLS imputation model 3, the estimated correlation coefficient is just 0.5.

The extent of agreement between time-average (observed) and TSTSLS (imputed) income quartile is summarised in Table B3 via Cohen's Kappa (fourth row) and the percentage agreement (fifth row). The Kappa statistics are in the range 0.15 to 0.28 – suggesting that there is evidence of only 'slight' to 'fair' agreement between observed and imputed earnings quartiles (see the results section in the main text for further details about the interpretation of these statistics). Furthermore, only between 37 and 47 percent of NLSY sample members are placed in the same earnings quartiles using the two techniques. This provides further evidence that the TSTSLS imputation procedure generates weak measures of long-run earnings.

Transition matrices are presented in Appendix Table B4, where we cross-tabulate TSTSLS imputed income quartile (imputation model 3) against the time-average income quartile. This confirms that the agreement between the two measures is rather low, and is consistent with our analysis using UK data.

Appendix Table B4

In order to establish whether the discrepancy between observed and imputed long-run earnings is associated with a set of observables characteristics, we consider the 'error' in the TSTSLS earnings imputations in more detail. As in the UK analysis, we create a new variable (D) which captures the difference between X_{AVG} and $\hat{X}_{TSTSLS}$, and estimate a series of bivariate OLS regression models to investigate whether there are observable factors associated with this difference. Appendix Table B5 shows the results for males and females. These results reinforce our hypothesis that the difference between TSTSLS imputed earnings ($\hat{X}_{TSTSLS}$) and observed time-averaged earnings (X_{AVG}) cannot simply be considered random 'noise'; the prediction error is clearly associated with a number of observable characteristics (including social class, parental education and children's test scores).

Appendix Table B5

The impact upon intergenerational associations

In the USA analysis we focus upon the relationship between mother's annual earnings and their children's math test scores. Appendix Figure B1 presents estimates from the OLS regression model, where children's math test scores (dependent variable) are regressed upon various measures of mothers' earnings (independent variable).

Appendix Figure B1

When using earnings data from a single year (X_{Single}), the parameter estimate of interest equals 1.7 (left-hand most bar). This suggests that a one log-unit increase in mother's annual earnings leads to an increase of 1.7 points in their offspring's math scores. The second bar from the left is when long-run earnings ($X_{AVG} = X_{BHPS}$) are used. The estimates coefficient is now 5.2, illustrating that use of current earnings leads to attenuation in intergenerational associations. The third column refers to the instrumental variable (X_{IV}) results. The estimated intergenerational association is now 13.9 – more than double the time-average estimate. Similar to the UK analysis, all the TSTSLS models lead to an overestimation of intergenerational associations. TSTSLS model 1, 2 and 3 all lead to substantial overestimation of the intergenerational association relative to the time-average estimate – usually by somewhere between 100 and 150 percent. Indeed, it is only when a very detailed imputation model is used (rarely found in the existing literature) that the upward bias is reduced to less than 50 percent.

According to the UK analysis we can argue that overestimation is likely, however when the auxiliary dataset is composed of a small sample size intergenerational associations may be *underestimated*. As in the main paper, Appendix Figure B2 and B3 illustrate this point. Both figures show the relationship between the auxiliary dataset sample size and the correlation between imputed and observed long-run earnings (left-hand panel) and the estimated association between imputed earnings and children's math test scores (right hand panel). Figure B2 refers to when TSTSLS imputation model 3 has been used and Figure B3 refers to imputation model 4. (Note that the intergenerational associations are only estimated for females).

Appendix Figure B2

Appendix Figure B3

The results are generally consistent with the UK analysis. The correlation between observed and imputed long-run earnings is typically between 0.45 and 0.5 when the auxiliary sample is composed of more than 1,000 observations. However, when the sample size is reduced the correlation starts to decrease dramatically. The results for females also illustrate how the estimated intergenerational association can become erratic when the sample size is small – particularly when there are less than 1,000 observations in the auxiliary dataset and / or the imputation model is particularly detailed.

Appendix B. References

Bollinger, Christopher. 1998. "Measurement Error in the Current Population Survey: A. Nonparametric Look." Journal of Labor Economics 16(3): 576-94.

Appendix Table B1. Number of earnings observations for the NLSY79 cohort members

Number of earnings observations	Males %	Females %
5	0.2	0.3
6	0.4	0.3
7	0.6	0.6
8	1.2	0.8
9	2.0	1.2
10	2.6	1.8
11	3.7	3.2
12	5.6	5.8
13	11.7	11.2
14	13.4	13.3
15	12.6	13.1
16	12.7	13.4
17	10.4	11.3
18	8.9	9.5
19	7.5	7.8
20	5.3	5.6
21	1.1	0.9
n	3,624	3,851

Notes:

Source: Author calculations using the NLSY79 dataset

Appendix Table B2. 'First-stage' regression estimates for USA data

	Model 1		Model 2		Model 3		Model 4	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Ethnicity (Ref: White)								
Black	-0.243	0.005	-0.209	0,005	-0.192	0.005	-0.138	0.005
Other	-0.135	0.006	-0.068	0,006	-0.113	0.006	-0.086	0.005
Age	0.180	0.001	0.174	0,001	0.161	0.001	0.133	0.001
Age - Squared	-0.002	0.000	-0.002	0,000	-0.001	0.000	-0.001	0.000
Education (Ref: High School)								
Less than high school	-0.396	0.005	-	-	-0.342	0.005	-0.261	0.005
Some college, non degree	0.055	0.004	-	-	0.038	0.004	0.005	0.004
Associate degree	0.225	0.005	-	-	0.165	0.005	0.094	0.005
Bachelor degree	0.523	0.004	-	-	0.369	0.005	0.279	0.005
Beyond bachelor degree	0.801	0.006	-	-	0.565	0.006	0.465	0.007
Occupation (Ref: Precision Production/ Craft /Repairers)								
Managerial and Professional	-	-	0.517	0.004	0.197	0.005		
Technical, Sales and Administrative	-	-	0.105	0.005	-0.046	0.005		
Service	-	-	-0.307	0.006	-0.340	0.005	3 digit SOC categories used	
Farming, Forestry and Fishing	-	-	-0.501	0.010	-0.432	0.010		
Operatives and Laborers	-	-	-0.211	0.005	-0.186	0.004		
Non-occupational responses	-	-	-0.800	0.011	-0.881	0.011		
Constant	13.726	0.020	13.682	0.019	13.461	0.019	11.735	0.294
R-squared	0.3197		0.3314		0.3677		0.4349	
Observations	529,414		529,414		529,414		529,414	

Appendix Table B3. Comparison of observed and imputed long-run earnings

(c) Males

	Observed	Model 1	Model 2	Model 3	Model 4
R-Squared	-	0.32	0.33	0.37	0.43
Variance	0.62	0.12	0.11	0.15	0.23
Correlation between imputed and observed long-run earnings	-	0.48	0.41	0.51	0.54
Kappa statistic	-	0.15	0.12	0.23	0.28
Percentage correct	-	38	35	43	47
Sample size (NLSY79)	3,624	3,624	3,624	3,624	3,624
Sample size (CPS)	-	529,414	529,414	529,414	529,414

(b) Females

	Observed	Model 1	Model 2	Model 3	Model 4
R-Squared	-	0.22	0.27	0.31	0.37
Variance	0.66	0.13	0.11	0.15	0.25
Correlation between imputed and observed long-run earnings	-	0.44	0.41	0.50	0.56
Kappa statistic	-	0.15	0.10	0.19	0.26
Percentage correct	-	37	33	40	45
Sample size (NLSY79)	3,851	3,851	3,851	3,851	3,851
Sample size (CPS)	-	501,216	501,216	501,216	501,216

Notes:

i. Source: Authors’ calculations using NLSY79 and CPS datasets

ii. R-squared is in reference to the first-stage prediction equation

iii. Model 1 – 4 indicates which TSTSLS imputation specification has been used. See section 3 for further details.

Appendix Table B4. Cross-tabulation of observed and predicted earnings quartile

(b) Males

		Predicted quartile				
		Bottom	**2nd**	**3rd**	**Top**	**n**
Observed quartile	**Bottom Quartile**	53	24	17	6	1,115
	2nd Quartile	30	31	27	12	872
	3rd Quartile	23	22	33	22	745
	Top Quartile	7	13	29	51	668

(b) Females

		Predicted quartile				
		Bottom	**2nd**	**3rd**	**Top**	**n**
Observed quartile	**Bottom Quartile**	62	19	13	6	949
	2nd Quartile	49	18	20	12	862
	3rd Quartile	30	20	24	26	850
	Top Quartile	11	12	27	50	791

Notes:

i. Figures refer to row percentages.
ii. The final column (n) refers to unweighted sample sizes
iii. The associated kappa statistic is 0.23 for males and 0.19 for females. See Table 2.
iv. Source: Authors' calculations using TSTSLS prediction model 3 (see section 3 for further details).

Appendix Table B5. Relationship between prediction error and selected characteristics

Panel A. Social class

(i) Males

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Social class (Ref: Precision production, craft, and repairers)								
Managerial and Professional	0.194*	0.049	0.048	0.053	0.075	0.053	-0.001	0.052
Technical, sales, and administrative	0.057	0.061	0.235*	0.067	0.232*	0.067	0.256*	0.065
Service	-0.371*	0.069	0.207*	0.076	0.176*	0.076	0.163*	0.074
Farming, forestry, and fishing	-0.468*	0.156	0.451*	0.197	0.242	0.175	0.449	0.277
Operatives and laborers	-0.062	0.060	0.259*	0.055	0.232*	0.056	0.215*	0.059
Non-occupational responses	-0.241	0.175	1.322*	0.203	1.270*	0.196	1.639*	0.250

(ii) Females

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Social class (Ref: Precision production, craft, and repairers)								
Managerial and Professional	0.054	0.122	0.158	0.128	0.102	0.136	-0.129	0.124
Technical, sales, and administrative	-0.048	0.122	0.238*	0.126	0.188	0.135	-0.032	0.124
Service	-0.554*	0.128	0.153	0.135	0.112	0.142	0.048	0.133
Farming, forestry, and fishing	-0.647*	0.248	0.399	0.331	0.286	0.289	0.344	0.574
Operatives and laborers	-0.052	0.145	0.415*	0.156	0.309*	0.162	0.144	0.150
Non-occupational responses	0.211	0.339	2.161*	0.365	2.270*	0.384	1.141	0.744

Panel B. Education

(i) Males

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Education (Ref: High school)								
Less than high school	-0.033	0.062	-0.384*	0.059	0.072	0.061	0.002	0.060
Some college, no degree	0.155*	0.083	0.191*	0.081	0.184*	0.084	0.158*	0.087
Associate degree	-0.002	0.055	0.171*	0.059	-0.007	0.059	0.039	0.059
Bachelor degree	0.097*	0.057	0.458*	0.059	0.116*	0.061	0.146*	0.063
Beyond bachelor degree	-0.088	0.079	0.505*	0.084	-0.068	0.087	0.010	0.083

(ii) Females

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Education (Ref: High school)								
Less than high school	-0.057	0.078	-0.508*	0.072	-0.087	0.075	-0.132*	0.075
Some college, no degree	0.129*	0.077	0.086	0.084	0.060	0.087	0.029	0.086
Associate degree	-0.040	0.052	0.103*	0.056	-0.095*	0.058	-0.034	0.058
Bachelor degree	0.074	0.058	0.407*	0.062	0.032	0.064	0.063	0.060
Beyond bachelor degree	-0.153*	0.057	0.479*	0.058	-0.152*	0.060	-0.180*	0.062

Panel C. Children's math test scores (only females)

	Model 1		Model 2		Model 3		Model 4	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Math test score	0.007*	0.001	0.010*	0.002	0.005*	0.002	0.004*	0.002

Panel D. Industry

(i) Males

	Model 1		Model 2		Model 3		Model 4	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Industry (Ref: Wholesale and retail)								
Agriculture/Forestry/Fishing/Hunting	-0.249*	0.139	-0.331*	0.169	-0.265*	0.159	-0.333*	0.178
Mining	0.356*	0.170	0.350*	0.133	0.353*	0.169	0.179	0.149
Construction	-0.048	0.073	-0.215*	0.075	-0.171*	0.078	-0.049	0.077
Manufacturing	0.059	0.069	0.141*	0.071	0.073	0.073	0.035	0.072
Transport/Utilities	0.148*	0.074	0.242*	0.079	0.191*	0.080	0.315*	0.086
Financial Activities	0.233*	0.104	0.361*	0.120	0.233*	0.115	0.141	0.111
Professional/Business Services	-0.063	0.084	0.029	0.090	-0.050	0.090	-0.046	0.088
Education/Health Services	-0.292*	0.093	-0.088	0.098	-0.297*	0.100	0.074	0.099
Leisure and Hospitality	-0.548*	0.111	-0.344*	0.121	-0.426*	0.121	-0.074	0.126
Public administration	-0.065	0.082	0.294*	0.095	0.127	0.096	-0.072	0.096
Other services	-0.334*	0.108	-0.381*	0.121	-0.396*	0.120	-0.221*	0.114

(ii) Females

	Model 1		Model 2		Model 3		Model 4	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Industry (Ref: Wholesale and retail)								
Agriculture/Forestry/Fishing/Hunting	0.139	0.219	-0.258	0.221	0.015	0.206	-0.476*	0.203
Mining	0.296	0.275	0.403	0.445	0.363	0.369	-0.033	0.268
Construction	0.319*	0.176	0.157	0.176	0.228	0.194	0.226	0.180
Manufacturing	0.518*	0.079	0.495*	0.087	0.536*	0.089	0.406*	0.091
Transport/Utilities	0.347*	0.096	0.521*	0.109	0.409*	0.108	0.223*	0.107
Financial Activities	0.458*	0.083	0.474*	0.091	0.465*	0.091	0.125	0.092
Professional/Business Services	0.222*	0.084	0.306*	0.092	0.282*	0.091	0.060	0.092
Education/Health Services	0.016	0.065	0.124*	0.071	0.021	0.072	0.018	0.073
Leisure and Hospitality	-0.304	0.084	-0.131	0.093	-0.144	0.093	-0.012	0.098
Public administration	0.341*	0.077	0.517*	0.089	0.437*	0.089	0.064	0.092
Other services	-0.281*	0.109	-0.042	0.119	-0.127	0.123	-0.080	0.118

Notes:

i. Results from a series of bivariate regressions.

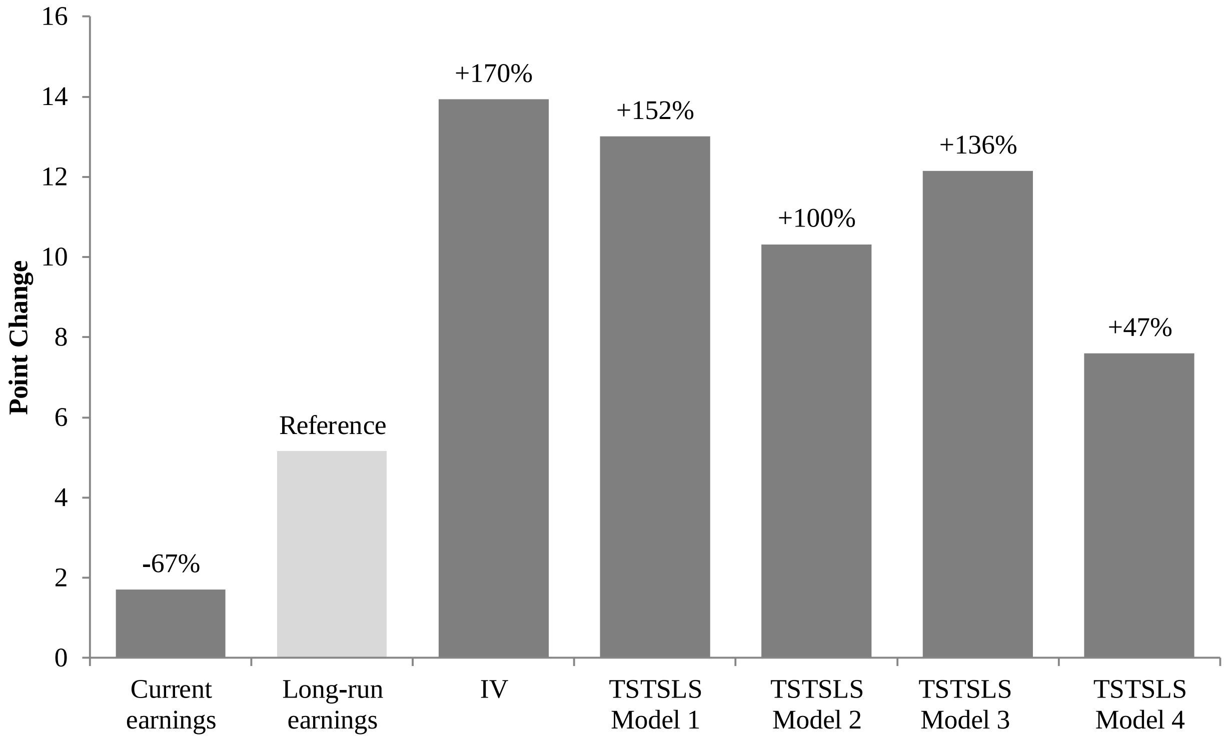
ii. * indicates statistical significance at the ten percent level.

iii. All figures refer to standard deviation differences in relation to the reference group.

iv. Model 1 – model 4 refer to the different TSTSLS imputation model used.

v. Source: Authors' calculations using the NLSY79 dataset.

Appendix Figure B1. Estimates of the association between mothers' earnings and children's math test scores



Notes:

i. Estimates based upon OLS model.

ii. Figures on the y-axis illustrate the point change in the children's math test scores a one log-unit change in mothers' annual earnings.

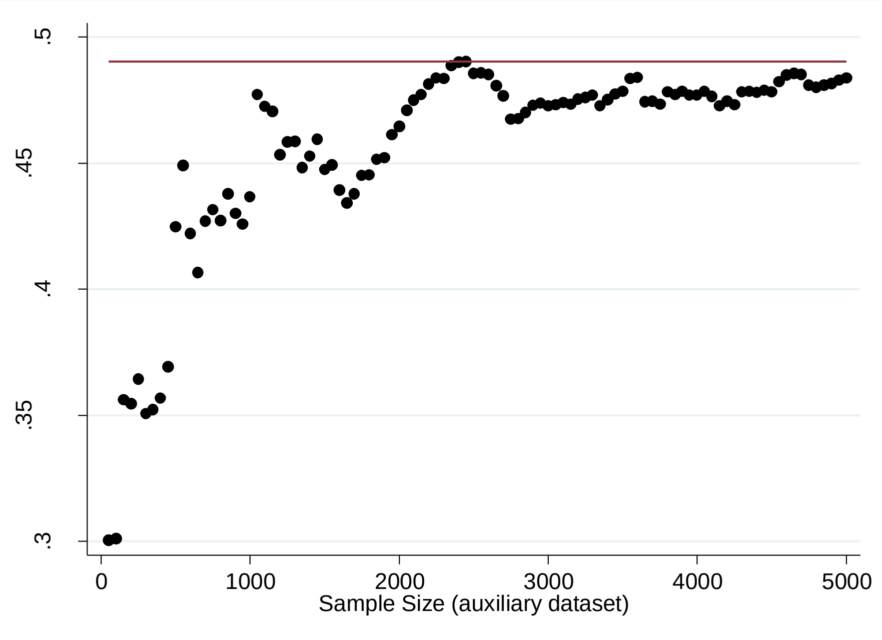
iii. The four bars on the right are based upon TSTSLS predictions of long-run earnings.

iv. Percentages above the bars refer to the percentage under or over estimation relative to the observed long-run earnings measure (reference group).

Appendix Figure B2. Correlation between predicted and observed long-run earnings using different auxiliary dataset sample sizes (imputation model 3)

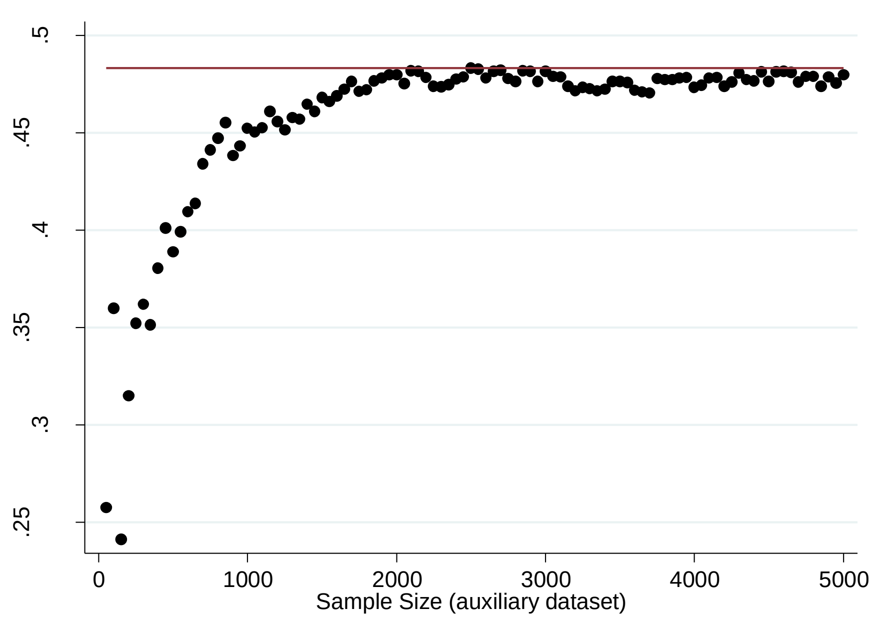
(i) Males

(a) Correlation (imputed and observed)

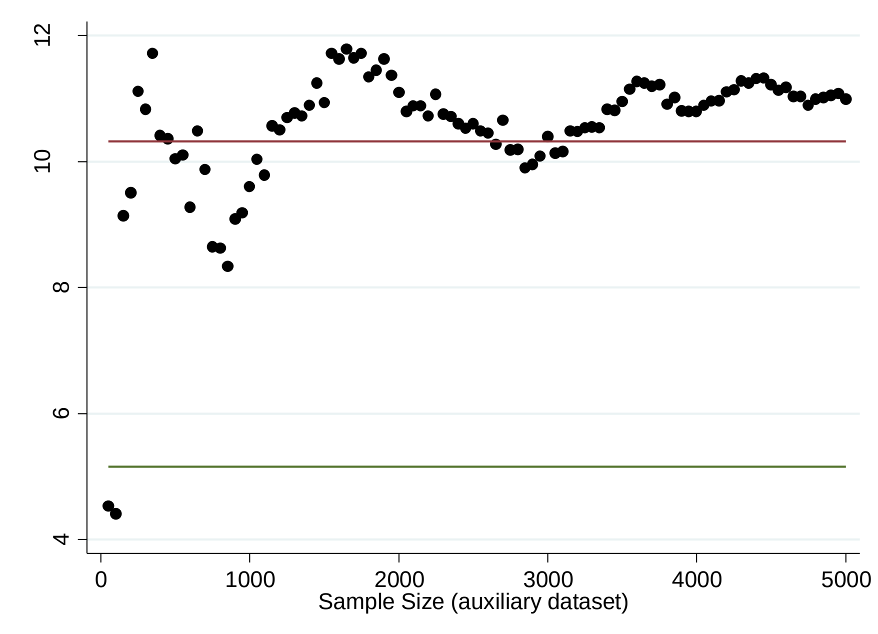


(ii) Females

(a) Correlation (imputed and observed)



(b) Regression estimates



i. Panel (a) illustrates the association between the auxiliary dataset sample size and the association between imputed and observed earnings. The horizontal line at the top of the graph illustrates the estimated correlation coefficient when all 529,414 CPS observations have been used.

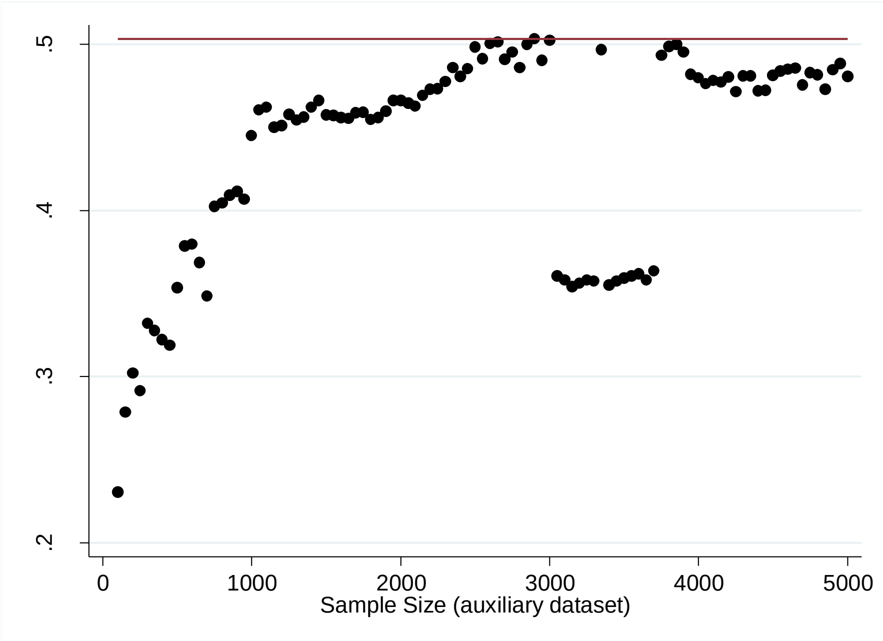
ii. Panel (b) refers to the association between imputed mother's earnings and children's math scores. The uppermost (red) line illustrates the estimate when all CPS observations were used. The lower (green) line is the estimate when observed time-average mother's earnings have been used.

iii. Source: Authors' calculations using the NLSY79 dataset, applying TSTSLS imputation model 3

Appendix Figure B3. Correlation between predicted and actual long-run earnings using different auxiliary dataset sample sizes (imputation model 4)

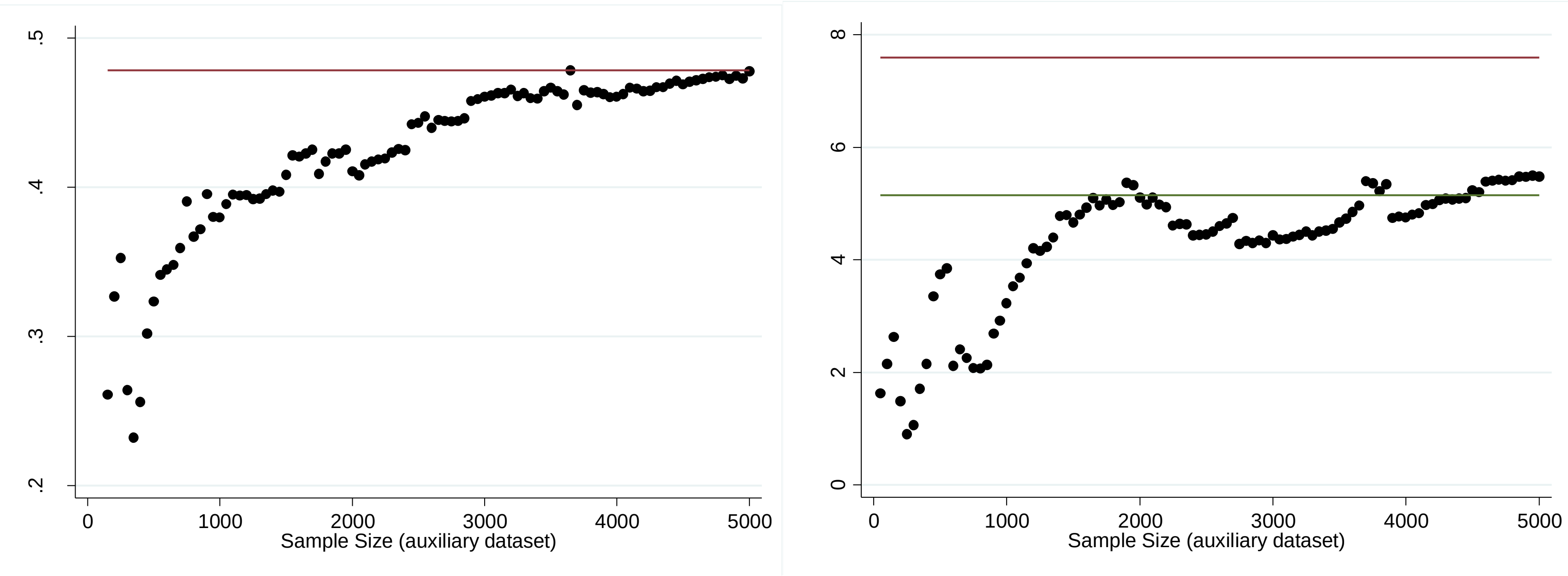
(i) Males

(a) Correlation (imputed and observed)



(ii) Females

(a) Correlation (imputed and observed) **(b) Regression estimates**



Notes

i. See notes to Figure 4 above

ii. Source: Authors' calculations using the NLSY79 dataset, applying TSTSLS imputation model 4

Appendix C. 'First-stage' regression estimates

	Model 1		Model 2		Model 3		Model 4	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
Ethnicity (Ref: White)								
Black	-0.201	0.016	-0.099	0.015	-0.112	0.015	-0.089	0.014
Chinese	-0.046	0.039	-0.002	0.034	-0.049	0.034	-0.020	0.031
Other	-0.101	0.010	-0.069	0.009	-0.075	0.009	-0.075	0.008
Age	0.013	0.000	0.010	0.000	0.010	0.000	0.010	0.000
Age - Squared	-0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
Education (Ref: Degree)								
Other higher education	-0.263	0.007	-	-	-0.166	0.006	-0.118	0.006
A-Level	-0.438	0.005	-	-	-0.227	0.006	-0.179	0.006
O-Level	-0.473	0.006	-	-	-0.250	0.006	-0.200	0.006
CSE	-0.674	0.010	-	-	-0.350	0.010	-0.276	0.009
None	-0.614	0.007	-	-	-0.308	0.008	-0.233	0.007
Occupation (Ref: Senior managers / officials)								
Professionals	-	-	0.040	0.007	-0.036	0.007	4 digit SOC categories used	
Associate professional and technical	-	-	-0.192	0.007	-0.182	0.007		
Administration	-	-	-0.473	0.008	-0.434	0.008		
Skilled Trade	-	-	-0.497	0.007	-0.407	0.007		
Service	-	-	-0.704	0.011	-0.629	0.011		
Sales and customer service	-	-	-0.636	0.010	-0.565	0.010		
Plant and machine operative	-	-	-0.629	0.007	-0.515	0.007		
Elementary occupations	-	-	-0.740	0.007	-0.635	0.007		
Constant	3.017	0.004	2.948	0.005	3.094	0.006		
R-squared	0.299		0.390		0.416		0.488	
Observations	69,548		69,548		69,548		69,548	

Appendix D. 'Split sample' robustness test

In this Appendix we illustrate that the problems highlighted with the TSTSLS imputations of father's earnings is not simply due to differences between the main and auxiliary datasets that we analyse (e.g. that they represent different populations or measure the key variables in different ways). To do so, we perform what we call a 'split – sample' robustness test. Specifically, in the main text we used the BHPS as our 'main' sample and the LFS as our 'auxiliary' sample. In this Appendix, we use just the BHPS data – splitting it into two random parts[7]. One half of this split BHPS dataset is defined as the auxiliary sample and the other half is defined as the main sample. We then follow exactly the same modelling strategy as outlined in section 2 of the paper. The advantage of the analysis in this appendix is that we can be sure that the main and auxiliary samples are (i) drawn from and represent the same population and (ii) that the imputer (Z) variables are defined and measured in exactly the same way. If our results are consistent with those presented in the main text, then we can rule out the possibility that our findings are simply being driven by such differences between the main and auxiliary datasets.

In Appendix Table D1 we present our key findings. These are analogous to those presented for the United Kingdom in Table 2 in the main text. There is little change to our results or substantive conclusions. In particular, note that the variance of imputed earnings is typically well below that when using the time-average approach. Moreover, the correlation between imputed and observed long-runs earnings never exceeds 0.50. All Kappa and percentage correct statistics are very low – well below rules of thumb often used to define minimum acceptable quality thresholds. In additional analysis, not presented for brevity, we also confirm that there are observable characteristics that are strongly and significantly associated with the prediction error (i.e. the difference between observed and imputed values). This provides support for our finding that differences between the two cannot simply be thought of as random noise.

In conclusion, the results presented in this appendix are in close agreement with those presented in the main text. This demonstrates that our substantive findings are robust to any possible differences between the main and auxiliary samples – including target population and measurement of the imputer variables.

[7] For each observation we take a random draw from a normal distribution with mean 0 and standard deviation 1. If this random draw is negative, the respondent is defined as part of the 'main sample'. If the random draw is positive, they are defined as part of the 'auxiliary' sample.

Appendix Table D1. 'Split sample' summary results

	Observed	Model 1	Model 2	Model 3	Model 4
R-Squared	-	0.15	0.25	0.28	0.54
Variance	0.24	0.05	0.07	0.08	0.16
Correlation between imputed and observed long-run earnings	-	0.34	0.46	0.49	0.40
Kappa statistic	-	0.13	0.15	0.21	0.22
Percentage correct	-	35	37	41	41
Sample size (Main)	1,219	1,210	1,205	1,196	1,197
Sample size (Auxiliary)	-	1,295	1,291	1,291	1,295