**Supplementary Materials for:**

**Beyond the Feinstein chart: Investigating differential achievement trajectories in a US cohort**

by

Elizabeth Washbrook (University of Bristol)

RaeHyuck Lee (Columbia University)

**A1. Preliminaries**

As described in our main paper, a flexible model for investigating the extent to which trajectories diverge by SES is

Where is the achievement of child on measurement occasion ; is a dummy variable indicating membership of the low (relative to high) SES group, which can easily be extended to a vector distinguishing multiple SES categories; and is an independently distributed mean zero residual term. The conditional SES gap at occasion 2 – that is, the difference in test scores predicted to open up between low and high SES children with an identical achievement level at time 1 – is given by

In this presentation we proceed with the more parsimonious model:

The model in equation (3) cannot be estimated directly because and are not observed. Instead we observe

where is a mean zero independently distributed measurement error.

Substituting for in (3) gives the estimating equation

It is clear from (4) that least squares estimates of the parameters in (5) will be biased due to the correlation of with the element of the error term in brackets.

**A2. The instrumental variables (IV) approach**

One method for correcting for this bias is to use an instrument for , a variable . For the model in (5) this yields two first stage equations

The key requirements for two-stage least squares estimates of (6) and (5) to yield consistent estimates of the parameters of interest are that

Instruments that are correlated with measurement errors because, for example, they are tests taken on the same occasion and under the same conditions as , will violate assumption (7a), although commonly it is assumed that the measurement errors are independent white noise processes in which case this assumption is automatically satisfied. Perhaps the more problematic requirement is the exclusion restriction (7b), i.e. that the instrument contains no information about once true achievement (plus its square and ) are conditioned on.

**A3. Categorizing by an auxiliary variable as an IV estimator[[1]](#footnote-1)**

Jerrim and Vignoles (2013) propose the use of an auxiliary test score to categorize children as “high” or “low” ability on the first measurement occasion as a way to correct for measurement error. Here we show how this auxiliary variable performs the same function as an instrumental variable, and apply the logic to the estimates shown in Feinstein’s original 2003 chart.

We take the model to be estimated in this example to be partway between models (1) and (3) in terms of generality.

where, similar to section A1, we can show that . Equation (8) allows for an interaction between and in predicting but, as can be seen in the first-stage equations below, the use of binary variables as instruments does not provide sufficient information to identify coefficients on the quadratic terms in (1).

Define as a binary variable equal to 1 if , child ’s score on the auxiliary variable, is above the 75th percentile, and zero if it is below the 25th percentile. For ease of comparison with the estimates presented in Feinstein (2003) and Jerrim and Vignoles (2013), children with scores on between the 25th and 75th percentiles are assumed to be discarded from the analysis.

We can define one first-stage equation for :

The standard 2SLS procedure takes the linear predictions from (9) and substitutes them into (8) in place of . Provided the IV assumptions are satisfied, the and parameters of interest are identified from

where the first term in square brackets is the first-stage linear prediction and the second term in square brackets is a residual that is uncorrelated with by construction.

The auxiliary variable approach presents the reduced form of (10) instead of the 2SLS estimates of (8). The relationship between the two can be seen by re-arranging (10) and collecting terms. Writing the reduced form as

equation (10) implies that

As shown in Table A1 below, the conditional group means used in Feinstein’s and Jerrim and Vigoles’ plots, , are estimates of different combinations of the first-stage () and reduced-form () parameters in (9) and (11) respectively. Given estimates of these parameters, the expressions (12a) to (12d) can be solved to give the IV estimators from (10).

To illustrate, we apply equations (13a) to (13d) to numbers directly taken from Feinstein’s original chart. In this case, is a dummy variable equal to 1 if the child is high ability at 22 months, equal to zero if the child is low ability at 22 months, and the observation is omitted from the analysis otherwise[[2]](#footnote-2). is the child’s percentile rank at 42 months, and is their percentile rank at 10 years[[3]](#footnote-3). It is important to note that in this analysis baseline ability is defined at 42 rather than 22 months, so the trajectories estimated here have a different interpretation from those in Feinstein’s original chart. The use of the 22 month outcome as an instrument means that we can only ask how trajectories diverged from an identical starting point at *42* months – the correction for measurement error in the 42 month score comes at the price of making statements about trajectories from the earliest observation in the data. The IV assumptions require uncorrelated measurement errors, and also that the 22 month score be redundant in the prediction of the 10 year outcome, once the 42 month score is held constant.

Table A1 provides estimates of the relevant group mean percentile scores at 42 months and 10 years read directly from Feinstein’s chart, and shows how they correspond to the parameters in equations (9) and (11). Table A2 then manipulates the group mean scores to isolate the individual model parameters (in the first two columns), and finally applies equations (13a) to (13d) to derive the IV estimators of (8) (the final column).

**Table A1. Group mean scores at 42 months and 10 years from Feinstein’s chart, and their interpretation as first-stage and reduced form IV equation parameters**

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Calculation** | **Model parameters** | **Observed value from Feinstein’s chart** |
| A. 42 month score () |  |  |  |
| High SES, high ability |  |  | 70 |
| Low SES, high ability |  |  | 56 |
| High SES, low ability |  |  | 42 |
| Low SES, low ability |  |  | 28 |
| B. Age 10 score () |  |  |  |
| High SES, high ability |  |  | 70 |
| Low SES, high ability |  |  | 40 |
| High SES, low ability |  |  | 58 |
| Low SES, low ability |  |  | 26 |

Notes: is a dummy variable denoting a high (top quartile) vs low (bottom quartile) score at 22 months. The model parameters show how the group means can be written in terms of a first-stage (eq 9) and reduced form (eq 11) IV equation.

**Table A2. First-stage, reduced form and second-stage IV estimates implied by Feinstein’s chart**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **First-stage (42m)** | | **Reduced form (10 years)** | | **Second-stage IV (10 years)** | |
|  | 42 |  | 58 |  | 40 |
|  | -14 |  | -32 |  | -31 |
|  | 28 |  | 12 |  | 0.4 |
|  | 0 |  | 2 |  | 0.2 |

Notes: Estimates in the first two columns are derived from the numbers in Table A1. They relate the outcome to SES and ability grouping at 22 months. Estimates in the final column are derived from the application of equations (13a) to (13d) and relate the outcome to SES and percentile rank at 42 months.

The results in Table A3 give an expression for the SES gap at 10 years for children with identical ability at 42 months but from different SES groups:

The gap varies with baseline ability and is slightly narrower for children with higher, rather than lower, ability at 42 months, as is shown in Figure A1. The estimates imply, for example, that among children at the 50th percentile of the ability distribution at 42 months, low SES children are predicted to fall to the 39th percentile by age 10 on average, while high SES children are predicted to rise to the 61st percentile, a gap of 22 ranks. Figure A2 translates the numbers into a format similar to Feinstein’s original chart, and plots the predicted trajectories associated with ability at the 12.5th and 87.5th percentiles at 42 months.

**Figure A1. SES differences in expected position in the ability distribution at 10 years, by position at 42 months**

**Figure A2. Examples of measurement error-corrected trajectories from 42 months derived from the estimates in Feinstein’s chart**

There are clearly many caveats associated with the estimates from these rough calculations: standard errors cannot be calculated without access to the underlying data; children from the medium SES group and with medium ability at 22 months contribute no information to the estimates; trajectories begin from a later time point than in Feinstein’s original analysis and so are not comparable. The purpose of the analysis is mainly to demonstrate the equivalence between the use of an auxiliary variable (here the 22 month ability grouping) and the concept of an instrumental variable to correct for measurement error. But these rough calculations are suggestive of very large differences in the progress of children from different SES groups from a common baseline at 3 years of age, even after some plausible correction is made for measurement error bias. They reiterate the point that we should not be quick to dismiss the substance of Feinstein’s findings on technical grounds even in his original BCS sample, let alone in other datasets.

**A4. Moment-based estimators to correct for measurement error**

The instrumental variables strategy relies on an untestable exclusion restriction, and also requires that an appropriate auxiliary variable be available in the dataset. An alternative approach to adjusting for measurement error imposes stronger distributional assumptions on the residual and measurement error terms, and draws on assumptions about the reliability of the initial test score . To illustrate we return to the model specification used in our analysis of the ECLS-K data, with equation (3) again the model to be estimated.

Now suppose that

where is simply the mean achievement score in the population on occasion 1 and is an individual-specific residual. Next assume that the residuals from (3) and (14), and the measurement errors from (4) are all independently normally distributed with constant variances (the classical errors-in-variables assumption employed in Jerrim and Vignoles, 2013).

We define the reliability of the test score , , as the ratio of the variances of the true to the observed residuals.

Finally, we need to allow for systematic differences in the mean achievement of the different SES groups at occasion 1

Overall mean achievement is simply the weighted sum of the means among the low and high SES groups respectively, and , with the weights given by the proportion of the population that are low- () and high- () SES respectively. The relationship between the residual and total variance of the observed test score is therefore

Under these assumptions, explicit expressions can be derived for the bias in least squares estimates for a range of simple models, including the fully interacted specification in equation (1)[[4]](#footnote-4). The derivation uses the formula for the conditional distribution of a sub-vector of variables from a multivariate normal distribution.

In this case, we use the assumptions given in equations 3, 4, 14, 15 and 17, and write

where

It follows from the properties of the normal distribution that

Using a hat above a parameter to denote its OLS estimator, we have that

And

The expressions for the expectations and covariances of , and , conditional on , in terms of the underlying structural parameters can be derived from the same set of five equations noted above. There are then inserted into the relevant matrices, and after some manipulation we derive expressions for the parameters estimated by least squares regression of (5):

These four equations can be rearranged to solve for the unknown parameters in the equation of interest (3), in terms of the least squares estimates and a number of population quantities.

Standardization of the observed test scores means that the values of and are fixed at 0 and 1 respectively. and can be estimated consistently as the SES sub-group means of the observed , and can be estimated by the sample proportion of low SES children. Together these values can be plugged into (18) to derive an estimate of the conditional variance . Equations (20a) to (20d) can hence be applied to calculate corrected estimates using a particular value of the reliability, with standard errors computed via the delta method (using the nlcom command in Stata 13)[[5]](#footnote-5).

The key requirement to implement the corrections in (20a)-(20d) is that we know the reliability of the baseline test, *r*. Test developers often provide estimates of reliability based on the internal consistency of the individual test items, but it seems unlikely that this will capture all potential sources of noise in all contexts. In cases where the reliability is uncertain or unknown, estimates for a range of values of can be computed to assess sensitivity. In addition, we might usefully ask how low the reliability would need to be for divergence by SES to be purely a statistical artefact. In terms of model (3), this can be calculated by setting (20b) to zero and solving for using the quadratic formula:

**References**

Cook, J.R. and Stefanski, L.A. (1994) Simulation-extrapolation estimation in parametric measurement error models*. Journal of American Statistical Association*, 89, 1314 – 1328

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, *70*(277), 73-97.

Goldstein, H. (1979). Some models for analysing longitudinal data on educational attainment. *Journal of the Royal Statistical Society. Series A (General)*, 407-442.

Jerrim, J., & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(4), 887-906.

1. This section was developed in collaboration with Bruce Bradbury. [↑](#footnote-ref-1)
2. We use the terms high and low ability in the section for consistency with Feinstein’s terminology. [↑](#footnote-ref-2)
3. The analogous treatment for Jerrim and Vignoles’ analysis of the MCS data defines if child was assessed as advanced or very advanced on the age 3 Bracken assessment, 0 if the child was classed delayed or very delayed, and undefined otherwise; is the age 3 BAS Naming Vocabulary percentile scores, and is the BAS Word Reading percentile score at age 7. [↑](#footnote-ref-3)
4. Calculations of this kind form the basis of the class of moment-based estimators that correct for measurement error bias. An example is provided in Goldstein (1979), and the Stata command eivreg provides a routine for deriving moment-based estimates for a simple linear model with no interaction terms. For more complex models where closed-form expressions of the estimated coefficients are too difficult or impossible to derive, simulation extrapolation (SIMEX) methods provide a convenient solution (Cook and Stefanski, 1994). [↑](#footnote-ref-4)
5. The standard errors reported in Table 2 of the main text are likely to be underestimated because sampling variability in the values of , and (as well as ) have not been accounted for. Given the magnitudes involved, however, it seems unlikely that this would affect any conclusions about statistical significance. [↑](#footnote-ref-5)