Intergenerational Mobility in Britain: Revisiting the Prediction Approach of Dearden, Machin and Reed*

by

Ramses Abul Naga
Université de Lausanne

and

Frank A. Cowell
London School of Economics and Political Science

Discussion Paper
No. DARP 62
September 2002

Distributional Analysis Research Programme
The Toyota Centre
Suntory and Toyota International
Centres for Economics and
Related Disciplines
London School of Economics
Houghton Street
London WC2A 2AE

---

* This research was partially supported by grant number SGS/00403/G from the Nuffield Foundation. We wish to thank Christophe Kolodziejczyk and Ceema Namazie for discussions. We are grateful to Stephen Machin for providing us with the data used in this paper and for helpful comments on an earlier draft. Responsibility for any errors remains ours.

Distributional Analysis Research Programme

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge and most papers are downloadable from the website. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Sue Coles on:

Telephone:	UK+20 7955 6678
Fax:	UK+20 7955 6951
Email:	s.coles@lse.ac.uk
Web site:	http://sticerd.lse.ac.uk/DARP

ABSTRACT

The *prediction approach* proposed by Dearden, Machin and Reed (DMR) consists in (1) regressing the observed incomes of the child and parent families on separate sets of predetermined variables, and (2) regressing the child's predicted income on that of the parents. Conceptually, this estimator must relate to the 2SLS/IV estimator. We re-derive the prediction estimator in matrix form, and reconsider its consistency requirements. The measurement model of DMR is then embedded within a simultaneous equations framework, for which an alternative 2SLS/IV estimator is proposed. The latter produces larger estimates for the intergenerational correlation. The policy relevance of the two sets of findings is then discussed.

Keywords: Intergenerational mobility, measurement error, prediction approach, simultaneous equations.

JEL classification: I3, J62

Contact address: Dr Ramses ABUL NAGA, DEEP, Université de Lausanne, CH-1015 Lausanne, Switzerland.
Professor Frank Cowell, STICERD, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, email: f.cowell@lse.ac.uk

1 Introduction

Decently Dearden, Machin and Reed (1997) used a new data set, the National Child Development Study (NCDS) to examine income continuities in Britain. Their study is of great value and topicality: since the work of Atkinson, Maynard and Trinder (1983), no major study of this sort had been undertaken on British data. Furthermore, the NCDS is a more representative sample of Britain than the data used by Atkinson et al., which only sampled the city of York. One problem with the NCDS in comparison to some US data sets used for the analysis of intergenerational income linkages – such as the Panel Study of Income Dynamics and the National Longitudinal Survey – is that the NCDS does not provide repeated measurements on the parent family's income. For this reason, the estimation techniques developed for panel data (e.g. Solon, 1992 and Zimmerman, 1992) could not be applied in the context of the study of Dearden et al. In the light of this Dearden et al. proposed a new estimator for the Galtonian model: the *prediction approach*. The suggested estimator consists of (1) predicting the permanent income of the child and parent families and (2) regressing the former variable on the latter. This innovative method is related to the two-stage least squares / instrumental variables (2SLS / IV) estimator of the simultaneous equations model. The rationale underlying this new estimator merits scrutiny since classical measurement error on the dependent variable does not bias the slope estimates of a regression model. The thrust of the present paper is threefold:

1. We re-examine the stochastic properties of Dearden et al.'s prediction estimator and show that the consistency conditions given by the authors must be modified to guarantee convergence of the prediction estimator to the parameter of interest, namely the intergenerational earnings correlation.

2. Reformulating the measurement model as a simultaneous equations system enables us to characterize an alternative, two-stage least squares / instrumental variables estimator. We also discuss the link between the two estimators, and under which conditions these two can be made to coincide.

3. A re-estimation of the intergenerational correlation using Dearden et al.'s dataset shows the empirical relevance of our findings. Estimates of the intergenerational earnings correlation between fathers and sons are

in the order of 0.40 to 0.45 using the prediction approach. However, our estimates are in the region of 0.60.

The plan of the paper is the following. In section 2 we re-examine the prediction-estimator method. In section 3 we turn to the problem of characterizing its consistency requirements. Further elaboration of this question is provided in an appendix to the paper. In section 4 we provide a simultaneous equations approach to the estimation problem considered by the authors in order to examine more closely the relation between the prediction approach and the 2SLS/IV estimator. In section 5 we re-run some of the estimations of Dearden et al. using our proposed estimator for the NCDS data: the policy relevance of both sets of results are discussed in relation to the respective estimation methods. Section 6 concludes the paper.

2 The prediction estimator re-examined

The story opens with the estimation of the following regression

$$y_i^{child} = \alpha + \beta y_i^{parent} + \varepsilon_i \tag{DMR1}$$

where y_i^{child} and y_i^{parent} are the unobserved permanent incomes of children and parents respectively, or, in matrix notation:

$$Y^c = D\boldsymbol{\beta} + \varepsilon \tag{1}$$

where Y^c is an $n \times 1$ vector with ith element y_i^{child}, D is a $n \times 2$ matrix with ith row $[1 \quad y_i^{parent}]$, $\boldsymbol{\beta} = [\alpha \quad \beta]'$ and ε is an $n \times 1$ vector with ε_i in the ith entry.[1] (Equation numbers *DMR** denote original numbers in Dearden et al.) For both parents and children it is proposed to run regressions of the form

$$\begin{aligned} y_{it} &= y_i + \gamma x_{it} + w_{it} \\ &= \delta q_i + \gamma x_{it} + f_i + w_{it} \end{aligned} \tag{DMR2}$$

where y_{it} is measured income, and permanent status y_i is defined as $y_i = \delta q_i + f_i$. x_{it} contains information on time-varying characteristics such as

[1]We take ε to refer to the population relation (DMR1); ε_i is used both in relation to (DMR1) and (DMR7) by Dearden et al.

the age of the family head and the region of residence. The variables in q_i on the other hand measure time-invariant determinants of permanent status such as education and social class; the terms f_i and w_{it} are disturbance terms, f_i being a time-invariant effect pertaining to permanent status. In matrix notation (DMR2) and the relationship for permanent status become, respectively:

$$Y_t = Q\delta + X\gamma + f + w_t \tag{2}$$

$$Y = Q\delta + f \tag{3}$$

where Y, Y_t, f, and w_t are column vectors, Q is an $n \times \ell$ matrix, X is an $n \times k$ matrix, δ is a $\ell \times 1$ vector and γ is a $k \times 1$ vector.

Dearden et al. describe their prediction approach as follows:

1. Use (DMR2) to obtain estimates of permanent status for parent and child using the decomposition

   $$\hat{y}_{pi} := \hat{\delta} q_i = y_i - f_i \tag{DMR7}$$

   or equivalently

   $$Q\hat{\delta} = Y - f. \tag{4}$$

2. Use predicted incomes to run the least squares regression

   $$\hat{y}_{pi}^{child} = \alpha + \beta \hat{y}_{pi}^{parent} + \varepsilon_i \tag{DMR8}$$

The authors use the decomposition (DMR7) to derive their consistency condition. Writing (DMR7) as $y_i = \hat{\delta} q_i + f_i$, we obtain a statement

$$data = estimate + disturbance.$$

However, in contrast to this one would expect to find a decomposition of the form $data = estimate + residual$, with the residual component being a function of f_i, w_{it}, and the observations on q and x. To resolve this discrepancy let us re-examine the relationship captured by (DMR7) . First, to obtain $\hat{\delta}$, premultiply (2) by the idempotent matrix $M = I - X(X'X)^{-1}X'$. Given that $MX = 0$ we have:

$$MY_t = MQ\delta + M(f + w_t) \tag{5}$$

From the Frisch-Lovell theorem (Davidson and MacKinnon, 1992 ch.1) the OLS estimator of δ in (2) is equal to the estimator $\hat{\delta}$ of the transformed model:

$$\hat{\delta} = (Q'MQ)^{-1}Q'MY_t \tag{6}$$

Substituting for Y_t from (2) we obtain

$$\hat{\delta} = \delta + (Q'MQ)^{-1}Q'M(f + w_t)$$

and, instead of (DMR7), we have

$$Q\hat{\delta} = Q\delta + A(f + w_t) \tag{7}$$

where

$$A := Q(Q'MQ)^{-1}Q'M. \tag{8}$$

Using (3), equation (7) becomes

$$Q\hat{\delta} = Y - f + A(f + w_t) \tag{9}$$

Contrasting this with (4), it is clear that (4) lacks the term $A(f + w_t)$. Rearranging (9) in the form

$$Y = Q\hat{\delta} + \{[I - A]f - Aw_t\},$$

we may treat the term inside the braces as a residual[2] and thus obtain a decomposition of the type

$$data = estimate + residual.$$

To derive the prediction estimator explicitly we use

$$\hat{Y}^j = A^j Y_t^j,\ j = p, c \tag{10}$$

where A^p, A^c are, respectively, the implementations of A in (8) in the case of parents and children. Define ι_n as a $n \times 1$ vector of ones. The prediction estimator is then found from an OLS regression of $\hat{Y}^c$ on $\hat{D} := [\iota_n \ \ \hat{Y}^p]$:

$$\hat{\boldsymbol{\beta}}_p = (\hat{D}'\hat{D})^{-1}\hat{D}'\hat{Y}^c \tag{11}$$

[2]Note also that the residual components pertaining to f and w_t are orthogonal since $[I - A]A = 0$.

3 Consistency conditions

In order to re-examine the conditions under which the prediction estimator is consistent, we propose writing $\hat{\boldsymbol{\beta}}_p$ as an OLS estimator of an errors-in-variables model where the predictors are noisy measurements on the permanent incomes of parents and children.[3] Using (9) we may write for parent and child the relation

$$\hat{Y}^j = Y^j - f^j + A^j(f^j + w_t^j), \ j = p, c \tag{12}$$

The population model is the Galtonian regression (1), whereas the measurement model takes the form

$$\hat{Y}^c = \iota_n \alpha + \hat{Y}^p \beta + \varepsilon - f^c + A^c(f^c + w_t^c) + f^p \beta - A^p(f^p + w_t^p)\beta \tag{13}$$

To focus attention on the slope estimator $\hat{\beta}_p$ write (13) in deviation from its mean. This is equivalent to pre-multiplying (13) by the $n \times n$ matrix $T := I - \frac{1}{n}\iota_n \iota_n'$:

$$T\hat{Y}^c = T\hat{Y}^p \beta + T\varepsilon - Tf^c + TA^c(f^c + w_t^c) + Tf^p \beta - TA^p(f^p + w_t^p)\beta \tag{14}$$

Using the properties $T^2 = T = T'$ we have that $\hat{\beta}_p = (\hat{Y}^{p\prime} T \hat{Y}^p)^{-1} \hat{Y}^{p\prime} T \hat{Y}^c$, i.e.,

$$\begin{aligned} \hat{\beta}_p &= \beta + (\hat{Y}^{p\prime} T \hat{Y}^p)^{-1} \{ \hat{Y}^{p\prime} T \varepsilon - \hat{Y}^{p\prime} T f^c + \hat{Y}^{p\prime} T A^c (f^c + w_t^c) \\ &\quad + \hat{Y}^{p\prime} T f^p \beta - \hat{Y}^{p\prime} T A^p (f^p + w_t^p) \beta \} \end{aligned} \tag{15}$$

Assume that Q is uncorrelated with ε, so that the first term inside the braces, $\hat{Y}^{p\prime} T \varepsilon$, converges in probability to zero. Furthermore, according to Dearden et al.:

> If we assume that the intergenerational transmission coefficient is identical for both observed and unobserved status, so that $f_i^{child} = \beta f_i^{parent} + \eta_i$, and that f_i^{parent} and the error term η_i are uncorrelated ($\sigma_{f^p \eta} = 0$) then a least squares regression of $\hat{y}_{pi}^{child}$ on $\hat{y}_{pi}^{parent}$ will produce an unbiased estimate of β.

[3]Note however that in (12) below signals and errors are not independent since Y^j is a function of f^j ($j = p, c$).

The condition proposed in Dearden et al.[4] ensures that $\text{plim}(\hat{Y}^{p\prime} T f^p \beta - \hat{Y}^{p\prime} T f^c) = 0$. However, the remaining term inside the braces of (15), namely $\hat{Y}^{p\prime} T A^c (f^c + w_t^c) + \hat{Y}^{p\prime} T A^p (f^p + w_t^p)\beta$, does not trivially vanish in probability since A^c appears in one component while A^p appears in the latter. This composite term comes from replacing (DMR7) by (9).

In the appendix we show that consistency of the prediction estimator may be established independently of the requirement that the intergenerational transmission coefficient is identical for both observed and unobserved status. Standard orthogonality conditions for (1) parental variables, namely $E(Q^{p\prime}\varepsilon) = E(Q^{p\prime} f^p) = E(X^{p\prime} f^p) = E(X^{p\prime} w_t^p) = 0$, (2) child variables $E(Q^{p\prime} f^c) = E(X^{c\prime} f^c) = E(X^{c\prime} w_t^c) = 0$ together with (3) a milder form of an intergenerational moment condition – namely $E(Q^{p\prime} f^c) = 0$ – are sufficient to guarantee consistency of the prediction approach.

4 A simultaneous equations approach

A further question in relation to the prediction approach arises at a conceptual level. It does not immediately follow why one must predict the dependent variable, namely the child's long-run status, when errors-in-variables biases only occur as a result of the mis-measurement of explanatory variables. So, let us consider an approach to the estimation of β within the linear simultaneous equations framework.

Using the relations (2) and (3), we may substitute $Y^c = Y_t^c - X_t^c \gamma^c - w_t^c$ and $Y^p = Y_t^p - X_t^p \gamma^p - w_t^p$ in the population regression model to obtain

$$Y_t^c = \iota_n \alpha + Y_t^p \beta + X_t^c \gamma^c + X_t^p \theta^p + \varepsilon + w_t^c - w_t^p \beta \tag{16}$$

where $\theta^p = -\gamma^p \beta$. In the above equation Y_t^p may effectively be taken as an endogenous variable since it correlates with the composite error term $\varepsilon + w_t^c - w_t^p \beta$, via its dependence on w_t^p. An equation where Y_t^p appears as the dependent variable is thus required. Using (2) once again, we may write for the parent family:

$$Y_t^p = Q^p \delta^p + X_t^p \gamma^p + f^p + w_t^p \tag{17}$$

Together (16) and (17) define a triangular simultaneous equations system. The classical two-stage least squares estimator obtains by (1) regressing Y_t^p

[4] $\hat{y}_{pi}^{child}$ and $\hat{y}_{pi}^{parent}$ are the predicted incomes obtained from the first stage regressions (see equation DMR8). These appear in vector forms as $\hat{Y}^c$ and $\hat{Y}^p$ in our equation (10).

on all the predetermined variables of (16) and (17), namely Q^p, X_t^p, and X_t^c to obtain a predictor $\tilde{Y}^p$, and (2) substituting $\tilde{Y}^p$ for the endogenous regressor Y_t^p in a least squares regression on (16). Equivalently, (16) may be directly estimated by the method of instrumental variables where Q^p is used to instrument Y_t^p (Johnston, 1984; pp. 472-478). In the above simultaneous equations framework there is thus no need to run prior residual regressions, or to make use of Stewart's (1983) Grouped Dependent Variable estimator on earnings equations as Dearden et al. have set out to solve their problem.

The two-stage least squares estimator is consistent under the usual condition that the predetermined variables be uncorrelated with the disturbance terms. For the identification of (16), it is thus required that Q^p, X_t^p, and X_t^c all be uncorrelated with ε, w_t^c and w_t^p.

One further point may be noted. It is possible within the framework of the two-stage least squares / instrumental variables estimator to define an approach similar in nature to the prediction estimator. Define in (16) the vector $\xi = [\alpha\ \beta\ \theta^p\ \gamma^c]$ and let $\chi = [\iota_n\ Y_t^p\ X^p\ X^c]$. Then (16) may be written more compactly as

$$Y_t^c = \chi\xi + \phi \tag{18}$$

where $\phi = \varepsilon + w_t^c - w_t^p\beta$. Now let $\Xi = [\iota_n\ Q^p\ X^p\ X^c]$ be the matrix of instruments for (18) and define $P_\Xi = \Xi(\Xi'\Xi)^{-1}\Xi'$. The 2SLS/IV estimator of ξ may then be written as

$$\hat{\xi} = (\chi' P_\Xi \chi)^{-1}\chi' P_\Xi Y_t^c \tag{19}$$

Now consider an estimation procedure whereby in a first stage χ is regressed on Ξ, and likewise Y_t^c is regressed on the same matrix of instruments Ξ. Define $\tilde{\chi} = P_\Xi\chi$ and $\tilde{Y}^c = P_\Xi Y_t^c$. Then, using again the property $P_\Xi^2 = P_\Xi = P_\Xi'$, it follows that $\hat{\xi}$ can be obtained from a regression of $\tilde{Y}^c$ on $\tilde{\chi}$. In other words, provided the same variables Ξ are used to predict the dependent and explanatory variables of a Galtonian model such as (16), one also obtains a two-stage least squares estimator. However, the estimator of Dearden et al. differs from the present one, since it utilizes Q^p to predict Y^p and Q^c to predict Y^c.

5 An application to the NCDS

To see whether the distinction between the prediction approach and the 2SLS/IV estimator of the triangular system (16-17) matters in practice consider a data application. A first clue to this question is provided by the instrumental variables regressions of the authors (table 3 of DMR) which produce larger estimates than $\hat{\boldsymbol{\beta}}_p$. There are some differences however between the IV estimator of DMR, which uses residuals from an initial Grouped Dependent Variable regression, and the one proposed here. For this reason, we re-estimate the intergenerational correlation β using the same fathers and sons sub-sample of Dearden et al., which is fully described in section 2 of their paper.

In table 1 we report selected estimations of the intergenerational correlation of earnings between fathers and sons using the prediction approach (see tables 3 and 4 of Dearden et al.), and using the instrumental variables regression on the triangular system (16) and (17). The selected estimates of Dearden et al. vary between 0.39 and 0.46, whereas ours are in the range of 0.57 to 0.60. Our standard errors are more than twice as large than theirs, resulting in less tight confidence intervals for the parameter of interest β. However, the variance estimator for the prediction approach is also likely to be mis-specified as a result of the omission of the above mentioned stochastic components of (9).

It should also be noted that our reported estimates of β vary less than those of Dearden et al., since the matrix Q_c does not enter our computations of the 2SLS / IV estimator of (16-17). For this same reason, our last two estimates of the intergenerational correlation do not change, while the inclusion of verbal and numerical test scores for children (Q_c variables) modifies the prediction approach estimate of β from 0.439 to 0.455.

As a whole then, the estimates of β obtained from our triangular system suggest that income inheritance in Britain is more intense than would be inferred from the prediction approach. Even though it may be possible to conceive of a scenario where β is negative, in general one would take 0 (the perfectly mobile society) and 1 (the perfectly rigid case) to be the limiting scenarios of interest in terms of policy concerns. With an estimate of 0.45 for β, and a generation taken as a time span of 25 years, a child whose father earned twice the average pay, would take 22 years to fall to 150% of the mean. With an estimate of 0.60, it would require an additional 12 years (i.e. a total of 34 years) for this same child to come down to the 150% level. Two generations down the line, the grandchild of the person earning double the average earnings level would be 20% above the mean with the 0.45 estimate,

Table 1: Selected estimates of the earnings correlation between fathers and sons

variables included	$\hat{\beta}_p$	$\hat{\beta}_{2SLS}$	n
X^p, Q^p=[edu ,DSC74]	0.425	0.594	1565
X^c, Q^c=[edu ,DSC91]	(0.027)	(0.063)	
X^p, Q^p=[edu ,DSC58]	0.389	0.576	1565
X^c, Q^c=[edu ,DSC91]	(0.028)	(0.067)	
X^p, Q^p=[edu ,DSC74,DSC58]	0.441	0.572	1565
X^c, Q^c=[edu ,DSC91]	(0.029)	(0.061)	
X^p, Q^p=[edu ,DSC74]	0.428	0.604	1441
X^c, Q^c=[edu ,DSC91]	(0.030)	(0.069)	
X^p, Q^p=[edu ,DSC58,DSC65,DSC74]	0.439	0.597	1441
X^c, Q^c=[edu ,DSCFJ,DSC81,DSC91]	(0.030)	(0.064)	
X^p, Q^p=[edu ,DSC58,DSC65,DSC74]	0.455	0.597	1441
X^c, Q^c=[edu ,DSCFJ,DSC81,DSC91,MATH,VERB]	(0.030)	(0.064)	

Notes:
1. n denotes sample size, standard errors inside parentheses.
2. X^c contains regional dummies for the son's place of residence, X^p contains age and age squared of the father and regional dummies.
3. DSCt is social class at year t and DSCFJ at the time of the first job.
4. edu denotes educational attainment, MATH numerical ability and VERB verbal ability.

but 36% above this same threshold with the 0.60 value. These two estimates therefore convey different pictures about the rate of convergence of incomes in Britain.

We do not claim that in practice our estimates are more plausible than the 0.39 to 0.46 range implied by the prediction approach. Solon (1992) for instance argues that the education of the individual may be an invalid instrument, resulting in an over-estimate of β by 2SLS/IV techniques:[5] but Abul Naga (2002), using panel data, finds only limited evidence in support of this claim. Overall our results go in the general direction of the conclusions reached by Dearden et al., namely "...that the β estimates derived from the prediction approach reported earlier may have been biased downwards. The estimates based on more detailed models move closer in the direction of the IV results" [pp.58-59].

Bound et al. (1994) provide interesting evidence that measurement error in reported earnings may take other forms than that the classical errors-in-variables model, where signal and noise are taken to be uncorrelated.[6] If, for instance, w_t^c is correlated with X_t^c, but not with Q_t^c, both the prediction approach and 2SLS/IV will produce inconsistent estimators, but for different reasons. To see this, assume (in obvious notation) a regression relationship

$$\eta = X\beta + u$$

satisfying $E(X'u) = 0$. The dependent variable is unobserved. Instead y is observed, such that $y = \eta + v$ and $E(X'v) \neq 0$. Assume Z is uncorrelated with u and v, and define

$$P_Z := Z(Z'Z)^{-1}Z',$$

$$M_Z := I - P_Z.$$

Now predict y to obtain $\hat{y} = P_Z y$. We then have

$$\hat{y} + M_Z\eta - P_Z v = \eta.$$

Thus, if we regress y on X we obtain a measurement model

$$y = X\beta + u + v$$

[5] As the father's earnings are also instrumented in the prediction approach, this claim on its own cannot readily explain why in practice the two estimators behave so differently.

[6] See also Solon (1999) for a discussion related to the estimation of Galtonian regressions.

while if $\hat{y}$ is regressed on X, we have

$$\hat{y} = X\beta + u - M_z\eta + P_Z v.$$

The least-squares estimators of those two models will be inconsistent, but will have different probability limits. In one case $\hat{\beta}$ is biased because X is correlated with v, while in the other it is the presence of η in the error term which causes problems. This would also provide some clues as to why the two estimators behave so differently.

6 Concluding comments

Dearden, Machin and Reed (1997) proposed a new estimator for the Galtonian model of income transmission which they have called the *prediction approach*. We have re-derived this estimator in matrix form and have reconsidered its consistency requirements, as its original derivation had omitted several stochastic components. We have re-written the model of Dearden et al. as a triangular simultaneous two-equation system, for which a two-stage least squares estimator is readily available. The relation between the latter and the prediction estimator was further discussed.

The 2SLS/IV estimator estimated the intergenerational correlation at 0.60 whereas the prediction approach estimates this same parameter at 0.40 to 0.45. If the duration of a generation can be taken as a 25 year period, a child raised in a family with double the average earnings level, would take 22 years to fall to 150% of the mean when the intergenerational correlation is estimated at 0.45. It would require an additional 12 years for this same transition to occur when the intergenerational correlation is equal to 0.60.

It is not immediately clear why in practice the two estimators behave so differently. As a means of reconciling these findings, we have suggested that measurement error in the child's earnings may be correlated with the right hand variables of the Galtonian regression. If this were the case, both estimators would be inconsistent, but for different reasons.

7 Appendix

In this appendix we elaborate further on the consistency requirements pertaining to the prediction estimator, as formulated in section 3.

Taking each of the five terms in (15) separately, we have

$$\text{plim}(\hat{Y}^{p\prime}T\varepsilon) = \text{plim}[Y_t^{p\prime}M^pQ^p(Q^{p\prime}M^pQ^p)^{-1}Q^{p\prime}T\varepsilon] \tag{A1}$$

$$\text{plim}(\hat{Y}^{p\prime}Tf^c) = \text{plim}[Y_t^{p\prime}M^pQ^p(Q^{p\prime}M^pQ^p)^{-1}Q^{p\prime}Tf^c] \tag{A2}$$

$$\begin{aligned}\text{plim}[\hat{Y}^{p\prime}TA^c(f^c + w_t^c)] &= \text{plim}[Y_t^{p\prime}M^pQ^p(Q^{p\prime}M^pQ^p)^{-1}Q^{p\prime}TQ^c \\ &\quad (Q^{c\prime}M^cQ^c)^{-1}Q^{c\prime}M^c(f^c + w_t^c)]\end{aligned} \tag{A3}$$

$$\text{plim}(\hat{Y}^{p\prime}Tf^p\beta) = \text{plim}[Y_t^{p\prime}M^pQ^p(Q^{p\prime}M^pQ^p)^{-1}Q^{p\prime}Tf^p\beta] \tag{A4}$$

$$\begin{aligned}\text{plim}[(\hat{Y}^{p\prime}TA^p(f^p + w_t^p)\beta)] &= \text{plim}[Y_t^{p\prime}M^pQ^p(Q^{p\prime}M^pQ^p)^{-1}Q^{p\prime}TQ^p \\ &\quad (Q^{p\prime}M^pQ^p)^{-1}Q^pM^p(f^p + w_t^p)\beta]\end{aligned} \tag{A5}$$

The terms (A1) and (A2) may be shown to converge in probability to zero provided the following conditions hold:

$$E(Q^{p\prime}\varepsilon) = 0 \tag{C1}$$

$$E(Q^{p\prime}f^c) = 0 \tag{C2}$$

While, recalling the definition of the M^j matrix, $M^j = I - X^j(X^{j\prime}X^j)^{-1}X^{j\prime}$ ($j = p,\ c$), (A3) converges in probability to zero if M^c is orthogonal to both f^c and w_t^c. That is, if

$$E(X^{c\prime}f^c) = 0 \tag{C3}$$

$$E(X^{c\prime}w_t^c) = 0 \tag{C4}$$

A sufficient condition for (A4) to vanish is that Q^p be uncorrelated with f^p:

$$E(Q^{p\prime}f^p) = 0 \tag{C5}$$

Equation (A5) is similar to (A3). Orthogonality conditions there may take the form

$$E(X^{p\prime}f^p) = 0 \tag{C6}$$

$$E(X^{p\prime}w_t^p) = 0 \tag{C7}$$

8 References

Abul Naga R. (2002): "Biases of the Ordinary Least Squares and Instrumental Variables Estimators of the Intergenerational Earnings Correlation: Revisited in the Light of Panel Data", working paper, DEEP, Université de Lausanne.

Atkinson A., A. Maynard and C. Trinder (1983): *Parents and Children: Incomes in Two Generations*, London, Heinemann.

Bound J., C. Brown, G. Duncan and W. Rodgers (1994): "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data", *Journal of Labor Economics* 12, 345-368.

Davidson R. and J. MacKinnon (1992): *Estimation and Inference in Econometrics*, Oxford, Oxford University Press.

Dearden L., S. Machin and H. Reed (1997): "Intergenerational Mobility in Britain", *Economic Journal* 107, 47-66.

Solon G. (1992): "Intergenerational Income Mobility in the United States", *American Economic Review* 82, 393-408.

Solon G. (1999): "Intergenerational Mobility in the Labor Market", in Ashenfelter O. and D. Card eds: *Handbook of Labor Economics* volume 3, 1761-1800, Amsterdam, Elsevier.

Stewart M. (1983): "On Least Squares Estimation when the Dependent Variable is Grouped", *Review of Economic Studies* 50, L,737-753.

Zimmerman D. (1992): "Regression Toward Mediocrity in Economic Stature", *American Economic Review* 82, 409-429.