Panel Data Estimation of the Intergenerational Correlation of Incomes[1]

by

Ramses Abul Naga[2] and Jaya Krishnakumar[3]

July 1999

Abstract

We consider the problem of estimating the intergenerational correlation of incomes in the context of a panel data framework with measurement errors. We present single equation estimation methods as well as system methods under various assumptions regarding the serial correlation of the error term and taking into account possible correlations among children of the same family. Application to a sample of US parents and children leads to estimates of the order of 0.42 to 0.60 for the coefficient of income transmission.

Keywords: Intergenerational mobility, panel data, errors in variables

JEL Classification Codes: I3, J6, C3

---

[1]We would like to thank Pietro Balestra, Christian Schluter and Dirk Van de gaer for their careful reading and valuable comments on an earlier version of the paper. Suggestions from the participants of the Economics Seminar at the University of St.Gallen are also acknowledged.

[2]DEEP and IEMS, University of Lausanne, Switzerland

[3]Department of Econometrics, University of Geneva, Switzerland

1 Introduction

Ever since President Johnson's declaration of the "War on Poverty" in the United States in the year 1964, economists have increasingly turned their attention and effort to understanding the mechanisms of income transmission. There are several important equity issues underlying the general question of the intensity of intergenerational inheritance of income status.

Equity concerns can be formulated along several lines. One question that often arises concerns the extent to which children raised in poor families are themselves likely to end up in the lower end of the distribution of income during their adult lives (Atkinson *et al.*(1983), ch. 1). The continuity problem in low incomes is often also formulated by contrasting the likelihoods that children raised in different social environments face, of ending up in poverty. Economists and social scientists also seek to assess the separate contributions of environment-related factors and genetically transmitted traits in the overall transmission of income status (the famous 'nature versus nurture' debate).

In immigration countries such as the United States, another question arises about the rate at which newcomers to the host country are likely to be assimilated in terms of economic and educational attainment (Stokey (1996)). To answer such questions, one needs to make progress both at the theory level, i.e. in the formulation of models of income transmission and the dynamics of wealth distribution (e.g. Galor and Zeira (1993)), as well as in the direction of the empirical analysis of longitudinal income data.

The estimation of the intergenerational correlation of incomes is a particularly complex task, since the variables of interests, namely the permanent incomes of parents and children, are typically unobservable quantities. Instead, the researcher usually possesses a short time-series (in the order of 3 or 4 observations) on some indicator of economic attainment such as hourly earnings, or a family income measure. The resulting measurement error problem has been shown by previous researchers in the area to render the OLS estimator biased towards zero (see e.g. Bowles (1972), Atkinson *et al.* (1983), Behrman and Taubman (1990)).

Of the various approaches proposed in order to deal with this measurement error problem, two methods have received considerable attention.

These include (i) averaging the income of parents (i.e. the error-ridden explanatory variable) over the several years of available data in order to reduce the ratio of transitory to total income variance, and (ii) the use of out of equation instruments (typically the father's education) for the income pertaining to the parents' generation (cf. for instance Solon (1992)).

The purpose of the present study is to estimate the intergenerational correlation of incomes within a panel data framework. It is rather surprising that no explicit treatment of this problem has been considered in the panel data perspective[4], since most recent studies (at least the US ones) have been based on samples extracted from longitudinal earnings surveys[5].

There are indeed several good reasons for considering this estimation problem within a panel data framework. Firstly, as shown by Griliches and Hausman (1986), Hsiao and Taylor (1991) and others, a wide range of errors in variables models may be identified without having to appeal to out of equation instruments. Secondly, by controlling for unobserved population heterogeneity, panel data estimators will usually have more explanatory power than their cross-section counterparts. Finally, because we have repeated measurements on the incomes of parents and children, we are able to estimate our various models under alternative scenarios regarding the serial correlation in the transitory components of income and test these scenarios. This latter point extends beyond efficiency considerations, since the choice of instruments itself is dictated by the assumptions regarding the serial correlation in errors of measurements.

Our study provides a discussion of two further issues: the pooling of instrumental variable regressions from the various cross-sections into a system (of equations) estimator, and the consequences of allowing for multiple children from the same parent family (entailing cross-sectional correlations across observations) for efficient estimation. When applied to a US sample of parents and children, the various models and estimators considered here provide estimates in the range of 0.42 to 0.60.

The structure of the paper is the following: section 2 discusses the data, section 3 deals with inference issues, section 4 contains the empirical appli-

[4] Though some insights into this problem are provided in Zimmerman (1992).

[5] See for instance Behrman and Taubman (1990), Altonji and Dunn (1991), Solon (1992), Zimmerman (1992) and Bjrklund and Jantti (1997)

cations, and section 5 concludes with a summary and directions for further research.

2 Data

The empirical applications contained in this study draw evidence from the panel study of income dynamics (PSID). The PSID is a US longitudinal survey of 5000 families which has been running annually since 1968. A detailed description of the history of the panel, and its main features, can be found in Hill (1993). Our sample was extracted from the merged family-individual tape of wave XXI (1968-1988) of the PSID. A summary of the characteristics of our sample can be found in Abul Naga (1998). Here we discuss three sampling aspects of our data which entail direct consequences for the estimation strategies to be pursued in section 3 below. These are (i) the extent to which our data meet the standard definition of a panel, (ii) the problem of measuring long-run economic status, or permanent income, and (iii) the presence of multiple children from the same parent family. We take up each of these points in turn.

We have a sample of 1426 parent-child pairs. Incomes of parents and children are observed respectively over the five year intervals 1967-71 and 1983-87. On the one hand data on dependent and explanatory variables are not contemporaneously observed, a feature which is not typical of the standard framework adopted in panel data analysis (for instance Hsiao (1986), Chap.3). On the other hand, the income data have a natural time ordering, which is essential for modelling and testing for serial correlation in the disturbances, and also for selecting valid instruments for explanatory variables (see below). As will become apparent in the next section, the estimators we adopt do not require incomes of parents and children to be contemporaneously observed. The estimators will be shown to have obvious panel data interpretations, but more generally, they may also be considered within the framework of multiple (simultaneous) equations systems.

Our main income concept is the total annual family income normalized by the family's needs, and deflated to 1967 dollars (the first year of the panel). The needs standard is the official Orshansky scale which is used in the US in order to quantify the extent of poverty. The price deflator used is the consumer price index. Such income measures may be treated as proxies for a household's long-run economic status, or following Griliches (1986), as random draws from a variable with a population mean equal to the household's permanent income. Thus, in section 3, we propose several instrumental variables procedures in order to deal with the measurement error problem inherent in our data.

As the income concept we have chosen to use is based on the total family income, we did not find it necessary to restrict our sample to father and son pairs. Thus, for a daughter from an initial sample family, the income of her household is matched to that of the parents' whether or not she happens to be the main income earner. We have also allowed for multiple children from the same parent family. We note that because of unobserved family background components, observations on siblings are expected to be correlated. Under such circumstances, efficient estimation will require the adoption of generalized least squares methods, which are also discussed in some length in the following section.

Table 1 : Summary Statistics

year	Parents		Children	
	mean	coef.variation	mean	coef.variation
1	1.73	0.80	2.86	0.70
2	1.88	0.85	3.00	0.71
3	2.00	0.84	3.05	0.76
4	2.15	0.83	3.15	0.77
5	2.31	0.83	3.25	0.83

1. The resource concept is the family income normalized by the Orshansky needs standard.
2. The first observation year (t=1) pertains to 1967 for parents, and to 1983 for children.

Table 1 provides some summary statistics for the incomes of parents and children. Parents appear to be on the whole poorer than the US population in the late 1960s. This may be explained by noting that approximately a half of the parents originate from the SEO file (a sample of families whose incomes fell short of twice their needs) of the PSID. Bicketti *et al.* (1988) have found that disadvantaged groups were more likely to exit the panel and that on the whole, the PSID has evolved towards a sample of middle class families. This finding may help in accounting for the seemingly lower level of income dispersion experienced by the sample of children, as well as the growth in their average income.

3 The Model and the Estimation Methods

We are interested in estimating a relationship of the form

$$\eta_{c,i} = \beta\eta_{p,i} + \alpha_i \tag{1}$$

where $\eta_{c,i}$ represents the logarithm of the permanent income of the child, $\eta_{p,i}$ the logarithm of the the permanent income of the parents and α_i a random disturbance term specific to the individual. If correct measurements of both the variables are available, then a consistent estimation of β would be given by OLS. However the variables in question are not directly observable and the only information one generally has are the actual income of the parent and the child at different points in time. Obviously these snapshots do contain a component equal to the permanent income but also include other transitory components. Hence these observations can be thought of as measurements of the permanent income subject to error. Denoting the observed incomes as y_{it} and x_{it}, we can write:

$$y_{it} = \eta_{c,i} + u_{it} \tag{2}$$
$$x_{it} = \eta_{p,i} + v_{it} \tag{3}$$

where u_{it} and v_{it} are uncorrelated random error terms with zero means. Thus the relationship between the observables is

$$y_{it} = \beta x_{it} + \alpha_i + (u_{it} - \beta v_{it}) \tag{4}$$

or

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it} \tag{5}$$

with

$$\varepsilon_{it} = u_{it} - \beta v_{it} \tag{6}$$

and

$$E(x_{it}\alpha_i) = E(\eta_c u_{it}) = E(\eta_p v_{it}) = E(u_{it}v_{it}) = 0 \tag{7}$$

Let us note that if only the dependent variable is measured with error ($\eta_{c,i}$ in our case), OLS would still consistently estimate β, whereas if the *explanatory* variable has a measurement error then OLS produces inconsistent estimates (cf. Greene (1993) e.g.). Though a remedy to this problem is given by the instrumental variable (IV) technique, in the case of pure cross section data, one needs to resort to extraneous instruments and this may pose a real practical difficulty. However there is an easier way out in the case of panel data and Griliches and Hausman (1986) were amongst the first ones to point out that the model can be identified without the use of external information. See also Hsiao and Taylor (1991), Wansbeek and Koning (1991), Biorn (1998) in this regard. Indeed, the observations on the explanatory variable relating to different time periods provide appropriate instruments for the current period, the gap to be chosen depending on the nature of serial correlation assumed for the errors.

In general, two approaches have been proposed in the literature for panel data models with measurement errors. Either the equation can be first differenced (to eliminate the specific effect α_i) and the instruments used in levels, or the equation is kept in levels and the instruments differenced (see Biorn (1998)).

For our model, special care needs to be taken while choosing the instruments as the true unobserved variable is time invariant and hence any differencing operation over time will leave us with only the difference of the corresponding v_{it}'s and any "instrument" correlated with this difference will necessarily be correlated with the same difference appearing in the error term. For instance, upon first differencing we get

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + \varepsilon_{it} - \varepsilon_{i,t-1} \tag{8}$$

But from (3) we have

$$x_{it} - x_{i,t-1} = v_{it} - v_{i,t-1} \tag{9}$$

which also appears in $\varepsilon_{it} - \varepsilon_{i,t-1}$. Thus any variable correlated with $(x_{it} - x_{i,t-1})$ (condition to be satisfied by a valid instrument) will always be correlated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$.

The above argument thus excludes the possibility of differencing in our case and we will therefore only consider estimating the equation in levels. For the same reasons, differencing of instruments is also inappropriate.

The set of valid instruments varies according to the type of serial correlation assumed for ε_{it}. Let us distinguish two cases, both assuming stationarity: (i) iid errors and (ii) MA(1) errors. More general MA processes will also be considered in the application section. The case of AR(1) errors offers no immediate solution and hence will not be pursued here.

Case 1: iid ε_{it}'s

Assumption A1:
$\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2); \quad \alpha_i \sim iid(0, \sigma_\alpha^2); \quad E(\alpha_j \varepsilon_{it}) = 0 \quad \forall i, j, t$

In terms of the individual components the assumptions would be

$E(u_{it}u_{js}) = \delta_{ij}\delta_{st}\sigma_u^2 \quad \forall i, j, s, t \quad E(v_{it}v_{js}) = \delta_{ij}\delta_{st}\sigma_v^2 \quad \forall i, j, s, t$
$E(\alpha_j u_{it}) = 0 \quad \forall i, j, t \quad E(\alpha_j v_{it}) = 0 \quad \forall i, j, t$
where δ_{ij} is the Kronecker delta. However it should be noted that only the overall variance $\sigma_\varepsilon^2 = \sigma_u^2 + \beta^2\sigma_v^2$ can be identified.

In this case, for any period t, all the observations at least one period apart are valid instruments. Let y_t, x_t, α and ε_t denote $(N \times 1)$ vectors obtained by piling all the individuals for period t, and define $r_t = \alpha + \varepsilon_t$. So for the period 1 equation written as

$$y_1 = x_1\beta + \alpha + \varepsilon_1 \equiv x_1\beta + r_1 \tag{10}$$

$z_1 = [x_2 \quad x_3 \quad ... \quad x_T]$ is a valid instrument matrix, assuming that we have observations for T time periods. Furthermore,

$$V(r_1) = V(\alpha) + V(\varepsilon_1) = \sigma_\alpha^2 I_N + \sigma_\varepsilon^2 I_N \equiv \sigma_r^2 I_N$$

Given this, two methods are possible: applying the IV technique equation by equation or combining all periods into a system and then applying IV with the whole set of instruments for all the periods together. Letting z_t denote the instrument set for the t-th period equation, the first method leads to

$$\hat{\beta}_{(t)} = [x_t' z_t (z_t' z_t)^{-1} z_t' x_t]^{-1} x_t' z_t (z_t' z_t)^{-1} z_t' y_t \tag{11}$$

as $V(r_t) = \sigma_r^2 I_N$.

The system estimator will be obtained as follows combining all t's:

$$\begin{array}{lcl} y_1 & = & x_1\beta + \alpha + \varepsilon_1 \\ . & & \\ . & & \\ . & & \\ y_T & = & x_T\beta + \alpha + \varepsilon_T \end{array}$$

or using appropriate notations

$$y = X\beta + (\iota_T \otimes I_N)\alpha + \varepsilon \equiv X\beta + r \tag{12}$$

where ι_T is a vector of ones, and y, X and r are $(NT \times 1)$ vectors. The resulting system variance covariance matrix is given by

$$V(r) \equiv \Sigma = \sigma_\alpha^2(\iota_T\iota_T' \otimes I_N) + \sigma_\varepsilon^2 I_{NT}$$

Writing the whole set of instruments as

$$Z = \begin{pmatrix} z_1 & 0 & 0 \\ 0 & z_2 & 0 \\ . & . & . \\ 0 & 0 & z_T \end{pmatrix} \tag{13}$$

we will have

$$\hat{\beta}_{(gls)} = [X'Z(Z'\Sigma Z)^{-1}Z'X]^{-1}X'Z(Z'\Sigma Z)^{-1}Z'y \tag{14}$$

This estimator can be shown to be a weighted average of the "individual" $\hat{\beta}_{(t)}$'s obtained by (11) for each period, with the weights being inversely proportional to the corresponding variance matrix. Needless to say that the same estimator can be derived using a Generalised Method of Moments (GMM) approach (cf. Greene (1993) e.g.) based on the orthogonality conditions between the instruments and the combined error term[6]. Note that we could also write a system OLS estimator as follows:

[6] An alternative estimator is the so-called Generalised Instrumental Variables Estimator (GIVE) which is obtained by using the instrument matrix $\Sigma^{-\frac{1}{2}}Z$ and then applying GLS on the model transformed by the instrument matrix. The relative efficiency of this estimator with respect to the one given in (14) cannot be established in general. We have not considered the GIVE in our empirical analysis due to the computational complications involved in its implementation.

$$\hat{\beta}_{(ols)} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y \tag{15}$$

So far we have assumed independence across different individuals in the sample. However our sample of parents and children contains siblings and it will be more realistic to assume that the individual effects of children of the same parents are correlated. This would imply non zero correlation between the specific effects α_i and α_j if i and j belong to the same family. Let us take this into account by means of a variable d_{ij} such that

$$d_{ij} = \begin{cases} 1 & \text{if } i,j \text{ are siblings} \\ 0 & \text{if not} \end{cases} \tag{16}$$

Let us further assume that

$E(\alpha_i \alpha_j) = d_{ij} c_\alpha$

This modifies our variance matrix of α as follows:

$$V(\alpha) = \begin{pmatrix} \sigma_\alpha^2 & d_{12}c_\alpha & ... & d_{1N}c_\alpha \\ d_{21}c_\alpha & \sigma_\alpha^2 & ... & d_{2N}c_\alpha \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ d_{N1}c_\alpha & .. & ... & \sigma_\alpha^2 \end{pmatrix} \equiv \Sigma_\alpha \tag{17}$$

Then

$$V(r_t) = \Sigma_\alpha + \sigma_\varepsilon^2 I_N \equiv \Sigma_0^* \tag{18}$$

and

$$V(r) \equiv \Sigma^* = (\iota_T \iota_T' \otimes \Sigma_\alpha) + \sigma_\varepsilon^2 I_{NT} = \begin{pmatrix} \Sigma_0^* & \Sigma_\alpha & ... & \Sigma_\alpha \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \Sigma_\alpha & \Sigma_\alpha & ... & \Sigma_0^* \end{pmatrix} \tag{19}$$

Thus the single equation estimator of β accounting for correlations across siblings would be given by

$$\hat{\beta}^*_{(t)} = [x_t' z_t (z_t' \Sigma_0^* z_t)^{-1} z_t' x_t]^{-1} x_t' z_t (z_t' \Sigma_0^* z_t)^{-1} z_t' y_t \tag{20}$$

and the system estimator by

$$\hat{\beta}^*_{(gls)} = [X'Z(Z'\Sigma^* Z)^{-1} Z'X]^{-1} X'Z(Z'\Sigma^* Z)^{-1} Z'y \qquad (21)$$

Now let us proceed to the MA(1) case.

<u>Case 2</u>: MA(1) errors

Assumption A2: $\varepsilon_{it} = w_{it} + \theta_\varepsilon w_{i,t-1}$

Here the underlying assumptions on the components are:

$$u_{it} = \rho_{it} + \theta_u \rho_{i,t-1} \qquad (22)$$

$$v_{it} = \lambda_{it} + \theta_v \lambda_{i,t-1} \qquad (23)$$

but once again only the parameters describing the overall error term can be identified. Let us denote
$\gamma_1 \equiv E(\varepsilon_{it}\varepsilon_{i,t+1})$

and we know for an MA(1) process

$E(\varepsilon_{it}\varepsilon_{is}) = 0 \quad \text{for} \quad |s-t| \geq 2$

The instrument set has to be modified according to the new assumptions as observations one period apart are no longer valid instruments. They have to be at least two periods away. Thus we would have (taking T=5 as is the case in our study):

$$\begin{aligned}
\tilde{z}_1 &= [x_3 \quad x_4 \quad x_5] \\
\tilde{z}_2 &= [x_4 \quad x_5] \\
\tilde{z}_3 &= [x_1 \quad x_5] \\
\tilde{z}_4 &= [x_1 \quad x_2] \\
\tilde{z}_5 &= [x_1 \quad x_2 \quad x_3]
\end{aligned}$$

Having made these changes and assuming for the moment that there are no siblings, the expression of the single equation estimator $\tilde{\beta}_{(t)}$ remains the same (equation(11)) where the new instrument $\tilde{z}_t$ is to be used instead of z_t. In the system estimation procedure, the variance of the vector r changes as we now have non zero correlation between r_t and r_{t-1} or r_{t+1}. Again assuming no siblings, we have

$$E(r_t r_t') = \sigma^2_\alpha I_N + \sigma^2_\varepsilon I_N \equiv \Sigma_0 \qquad (24)$$

$$E(r_t r'_{t+1}) = \sigma_\alpha^2 I_N + \gamma_1 I_N \equiv \Sigma_1 \tag{25}$$

and

$$E(r_t r'_s) = \sigma_\alpha^2 I_N \equiv \Sigma_2 \quad \text{for} \quad |s-t| \geq 2 \tag{26}$$

Thus

$$E(rr') = \begin{pmatrix} \Sigma_0 & \Sigma_1 & \Sigma_2 & \ldots & \Sigma_2 \\ \Sigma_1 & \Sigma_0 & \Sigma_1 & \ldots & \Sigma_2 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \Sigma_2 & \Sigma_2 & \ldots & \Sigma_1 & \Sigma_0 \end{pmatrix} \equiv \tilde{\Sigma} \tag{27}$$

Thus the modified system estimator would be written as

$$\tilde{\beta}_{gls} = [X'\tilde{Z}(\tilde{Z}'\tilde{\Sigma}\tilde{Z})^{-1}\tilde{Z}'X]^{-1}X'\tilde{Z}(\tilde{Z}'\tilde{\Sigma}\tilde{Z})^{-1}\tilde{Z}'y \tag{28}$$

where $\tilde{Z}$ now contains the new set of instruments.

Let us now introduce correlations among siblings. We get

$$\begin{aligned} E(r_t r'_t) &= \Sigma_\alpha + \sigma_\varepsilon^2 I_N \equiv \Sigma_0^* && (29) \\ E(r_t r'_{t+1}) &= \Sigma_\alpha + \gamma_1 I_N \equiv \Sigma_1^* && (30) \\ E(r_t r'_{t+2}) &= \Sigma_\alpha \equiv \Sigma_2^* && (31) \end{aligned}$$

Thus

$$E(rr') = \begin{pmatrix} \Sigma_0^* & \Sigma_1^* & \Sigma_2^* & \ldots & \Sigma_2^* \\ \Sigma_1^* & \Sigma_0^* & \Sigma_1^* & \ldots & \Sigma_2^* \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \Sigma_2^* & \Sigma_2^* & \ldots & \Sigma_1^* & \Sigma_0^* \end{pmatrix} \equiv \tilde{\Sigma}^* \tag{32}$$

Now, the single equation and system estimators taking account of such correlations are given respectively by

$$\tilde{\beta}^*_{(t)} = [x'_t\tilde{z}_t(\tilde{z}'_t\Sigma_0^*\tilde{z}_t)^{-1}\tilde{z}'_t x_t]^{-1}x'_t\tilde{z}_t(\tilde{z}'_t\Sigma_0^*\tilde{z}_t)^{-1}\tilde{z}'_t y_t \tag{33}$$

and

$$\tilde{\beta}^*_{(gls)} = [X'\tilde{Z}(\tilde{Z}'\tilde{\Sigma}^*\tilde{Z})^{-1}\tilde{Z}'X]^{-1}X'\tilde{Z}(\tilde{Z}'\tilde{\Sigma}^*\tilde{Z})^{-1}\tilde{Z}'y \tag{34}$$

Variance - covariance estimators

In order to implement all the above methods in practice, we need prior estimates of the various variance-covariance parameters appearing in the different blocks of $V(r)$. We give below the various sample moments that consistently estimate the relevant parameters.

As

$$r_{it} - r_{i.} = \varepsilon_{it} - \varepsilon_{i.} \tag{35}$$

and

$$V(\varepsilon_{it} - \varepsilon_{i.}) = \frac{(T-1)}{T}\sigma_\varepsilon^2 \tag{36}$$

where a dot in the place of a subscript indicates the average taken over it, σ_ε^2 can be estimated by

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)}\sum_i \sum_t (\hat{r}_{it} - \hat{r}_{i.})^2 \tag{37}$$

where $\hat{r}_{it} = y_{it} - x_{it}\hat{\beta}$ and $\hat{\beta}$ can be either the corresponding estimator for that time period (given by equation (11)) or a first stage system OLS estimator (given by (15)).
Using

$$V(r_{i.}) = V(\alpha_i) + V(\varepsilon_{i.}) = \sigma_\alpha^2 + \sigma_\varepsilon^2/T \tag{38}$$

we can estimate σ_α^2 as follows:

$$\hat{\sigma}_\alpha^2 = \frac{1}{N}\sum_i \hat{r}_{i.}^2 - \frac{1}{T}\hat{\sigma}_\varepsilon^2 \tag{39}$$

Finally, an estimator of c_α is given by

$$\hat{c}_\alpha = \frac{1}{N_J}\sum_{i,j\in J}(\hat{r}_{i.} - \tilde{r}_{..})(\hat{r}_{j.} - \tilde{r}_{..}) \tag{40}$$

where J is the group of individuals who are siblings, N_J is the number of persons in J and $\tilde{r}_{..}$ is the mean over this group.

In the MA(1) case, we need γ_1 in addition to the above. Knowing
$Cov(r_{it}, r_{i,t+1}) = \sigma_\alpha^2 + \gamma_1$
and
$Cov(r_{it} - r_{i.}, r_{i,t+1} - r_{i.}) = \frac{(T-1)}{T}\gamma_1$
we can obtain

$$\begin{aligned}\hat{\gamma}_1 &= \frac{T}{(T-1)}\frac{1}{(NT)}\sum_i\sum_t(\hat{r}_{it} - \hat{r}_{i.})(\hat{r}_{i,t+1} - \hat{r}_{i.}) \\ &= \frac{1}{(T-1)}\frac{1}{N}\sum_i\sum_t(\hat{r}_{it} - \hat{r}_{i.})(\hat{r}_{i,t+1} - \hat{r}_{i.})\end{aligned}$$

The above procedures (single equation and system) can be generalised to MA(2), MA(3),... errors in an exactly analogous way from iid to MA(1) without much difficulty. In fact these possibilities will also be explored empirically.

Specification tests

Before finishing with the methodology, let us look at the different specification tests that can be carried out in this context. The first type of tests that one can think of are those that concern the nature of serial correlation assumed for the errors r_t. For instance, in order to test the assumption of serial independence, denoted as MA(0), against the alternative MA(1), a Hausman test statistic can be constructed based on the difference between $\hat{\beta}$ and $\tilde{\beta}$ denoting the estimator under the null (time independence) as $\hat{\beta}$ and the one under the alternative (MA(1)) as $\tilde{\beta}$. Under the null, both are consistent with $\hat{\beta}$ being more efficient and under the alternative only $\tilde{\beta}$ is consistent. Let us consider the single equation case say for $t = 1$. The instrument will be $z_1 = [x_2 \quad x_3 \quad x_4 \quad x_5]$ under H_0 and $\tilde{z}_1 = [x_3 \quad x_4 \quad x_5]$ under H_a. The two estimators of β are given by:

$$\hat{\beta} = [x_1' z_1 (z_1' \hat{\Sigma}_0^* z_1)^{-1} z_1' x_1]^{-1} x_1' z_1 (z_1' \hat{\Sigma}_0^* z_1)^{-1} z_1' y_1 \qquad (41)$$

and

$$\tilde{\beta} = [x_1' \tilde{z}_1 (\tilde{z}_1' \hat{\Sigma}_0^* \tilde{z}_1)^{-1} \tilde{z}_1' x_1]^{-1} x_1' \tilde{z}_1 (\tilde{z}_1' \hat{\Sigma}_0^* \tilde{z}_1)^{-1} \tilde{z}_1' y_1 \tag{42}$$

where $\hat{\Sigma}_0^*$ is the estimator of the variance of r_1 under MA(1). It is easily shown that

$$V(\hat{\beta}) = [x_1' z_1 (z_1' \hat{\Sigma}_0^* z_1)^{-1} z_1' x_1]^{-1} \tag{43}$$

$$V(\tilde{\beta}) = [x_1' \tilde{z}_1 (\tilde{z}_1' \hat{\Sigma}_0^* \tilde{z}_1)^{-1} \tilde{z}_1' x_1]^{-1} \tag{44}$$

with $V(\hat{\beta}) < V(\tilde{\beta})$

and

$$Cov(\hat{\beta}, \tilde{\beta}) = V(\hat{\beta}) \tag{45}$$

Hence

$$V(\tilde{\beta} - \hat{\beta}) = V(\tilde{\beta}) - V(\hat{\beta}) \tag{46}$$

The test statistic is therefore

$$(\tilde{\beta} - \hat{\beta})' [V(\tilde{\beta}) - V(\hat{\beta})]^{-1} (\tilde{\beta} - \hat{\beta}) \sim \chi_1^2$$

with large values rejecting independence.

Similarly, MA(1) can be tested against MA(2) by constructing the test statistic based on the difference between the two corresponding estimators. In this case, the instrument sets for the first period equation are $z_1 = [x_3 \quad x_4 \quad x_5]$ under H_0 and $\tilde{z}_1 = [x_4 \quad x_5]$ under H_a. Other tests such as MA(2) against MA(3) or MA(1) against MA(3), MA(0) against MA(2) or MA(0) against MA(3) can all be devised similarly.

The same results hold for the system estimators too, the basic difference being the dimension of the variables involved. Let us illustrate the case iid against MA(1).

We have the system (see (12))

$$y = X\beta + r \tag{47}$$

and

$$Z = \begin{pmatrix} z_1 & 0 & 0 \\ 0 & z_2 & 0 \\ . & . & . \\ 0 & 0 & z_T \end{pmatrix} \tag{48}$$

for the null and

$$\tilde{Z} = \begin{pmatrix} \tilde{z}_1 & 0 & 0 \\ 0 & \tilde{z}_2 & 0 \\ . & . & . \\ 0 & 0 & \tilde{z}_T \end{pmatrix} \tag{49}$$

for the alternative. The expressions of the estimators will be

$$\hat{\beta} = [X'Z(Z'\Sigma^* Z)^{-1}Z'X]^{-1}X'Z(Z'\Sigma^* Z)^{-1}Z'y \tag{50}$$

and

$$\tilde{\beta} = [X'\tilde{Z}(\tilde{Z}'\Sigma^* \tilde{Z})^{-1}\tilde{Z}'X]^{-1}X'\tilde{Z}(\tilde{Z}'\Sigma^* \tilde{Z})^{-1}\tilde{Z}'y \tag{51}$$

Finally the test statistic will be given by

$$(\tilde{\beta} - \hat{\beta})'[V(\tilde{\beta}) - V(\hat{\beta})]^{-1}(\tilde{\beta} - \hat{\beta}) \sim \chi^2_K$$

Another important category of tests is the so-called overidentification tests of Sargan (1976). These tests use covariances between IV residuals and a set of instruments that need not have been used in the estimation to detect any model misspecification regarding the validity of the instruments. Under the null of no misspecification, the covariance between the residuals and the instrument set in question would be close to zero. More explicitly, let us say we have a model written as

$$y = X\beta + \varepsilon \tag{52}$$

and estimated by IV using W as instruments. Now suppose we have another potential instrument set $\tilde{W}$ of order m ($\tilde{W}$ may or may not overlap with W). Then the test statistic using the covariance matrix $\hat{\Delta}$ of $\hat{\varepsilon}'\tilde{W}$ is given by (see Godfrey (1988))

$$\hat{\varepsilon}'\tilde{W}\hat{\Delta}^{-1}\tilde{W}\hat{\varepsilon} \sim \chi^2_m$$

Now if $\tilde{W}$ is a subset of W (Pagan-Hall (1983) suggestion) then the above test is equivalent to testing $\delta = 0$ in the augmented regression

$$y = X\beta + \tilde{W}\delta + \varepsilon \tag{53}$$

as the above test statistic can be shown to be the same as that of Gallant and Jorgenson (1979) for testing $\delta = 0$ in the above model. The latter is based on the difference between the optimisation criterion in the unrestricted model (53) and the restricted model (with $\delta = 0$). It is given by

$$\tilde{S} - \hat{S} \sim \chi^2_m$$

where

$$\tilde{S} = \tilde{\varepsilon}' W (W \hat{\Sigma} W)^{-1} W' \tilde{\varepsilon}, \quad \tilde{\varepsilon} = y - X\tilde{\beta} - \tilde{W}\tilde{\delta}$$

$$\hat{S} = \hat{\varepsilon}' W (W \hat{\Sigma} W)^{-1} W' \hat{\varepsilon}, \quad \hat{\varepsilon} = y - X\hat{\beta}$$

and both models are estimated using W as instruments and $\hat{\Sigma}$ calculated using $\tilde{\varepsilon}$.

Coming to our model say for the single equation $t = 1$, in the iid case, one can test the validity of $\tilde{W} = [x_2]$ as instruments by estimating the augmented model

$$y_1 = x_1\beta + x_2\delta + r_1 \tag{54}$$

and testing $\delta = 0$. If the test is rejected then x_2 is not a valid instrument and only $[x_3 \quad x_4 \quad x_5]$ should be used as the instrument matrix. This can be repeated for $\tilde{W} = [x_2 \quad x_3]$ or $\tilde{W} = [x_2 \quad x_3 \quad x_4]$. In the MA(1) case, $W = [x_3 \quad x_4 \quad x_5]$ and $\tilde{W}$ can be either $[x_3]$ or $[x_3 \quad x_4]$.

Once again the procedure is easily extended for testing other MA structures and to the case of system estimators.

Let us now go on to look at the empirical results on the various estimators and tests discussed above.

4 Results

This empirical section is divided into three parts. Firstly we examine single equation estimation of the income transmission equation. Then, we consider a system analysis where all the equations are pooled to arrive at a unique estimator. Finally, in the third sub-section we carry out specification tests on the system estimates in an effort to reconcile the various findings and to arrive at general conclusions pertaining to the serial correlation in the error term, and hence the appropriate choice of instruments.

All the regressions results reported below include as explanatory variables a measurement on the parent family's annual income, age controls (age and 'age-squared' of the parent and child family heads) together with a constant. All estimations take account of correlations across siblings, and where appropriate, are robust to serial correlation up to three lags (or leads).

4.1 Single equation results

For the single equation analysis, we can report estimates for up to 5 years (year one pertains to observations on the child's 1983 income and her/his parents' 1967 income, while year 5 pertains to the child's 1987 income and the parents' 1971 resources). Single equation analysis is very popular in the empirical literature on intergenerational mobility (see for instance tables 2 and 3 of Solon (1992)) and table 2 of Bjrklund and Jantti (1997)). We note that potentially we could estimate a total of 25 equations, not just 5. This is so, because, in practice, nothing precludes us from using each of x_1 to x_5 to explain y_1 (i.e. five separate regressions), and to use the same variables to explain y_2 etc. Our choice here to regress y_t on x_t is purely expositional, and has been motivated by our interest in formulating our estimation problem in the context of the panel data framework. Below we report results for the first and last year of the panel only, since these will suffice to motivate the need for proceeding in the direction of system level estimation.

Table 2: IV results for period one of the panel

$\hat{\beta}$	instr. set : x_5	x_4	x_3	x_2	Sargan test
0.550(0.035)	×				–
0.544(0.033)	×	×			0.511
0.539(0.032)	×	×	×		0.525
0.532(0.031)	×	×	×	×	1.680

1. The dependent variable is y_1 (the child's 1983 income). The set of explanatory variables contains x_1 (the parents' 1967 income) together with age controls for the parent and child family heads.

2. The single equation estimator of β is as defined in equation (20). A × mark indicates that the corresponding variable is used as an instrument. The set of instruments (x_2 to x_5) contains leads on the parent family's income from 1968 to 1971.

3. Standard errors appear inside parentheses.

In table 2 we report results for period one of the panel. When x_5, parental income in 1971, is used to instrument x_1, (the corresponding 1967 variable) the coefficient of income inheritance β is estimated as 0.55. As a benchmark, the OLS estimate for this parameter is 0.485 with a standard error of 0.0245. As this latter estimate however is downwardly biased (and its standard error is incorrect), it is expected that IV estimates of β would be higher; this is indeed the case for all the figures reported in table 2.

All four estimates of table 2 are within the narrow bracket of 0.53 to 0.55 (less than one standard error from one another). With a sample size of 1426 observations, adding extra instruments can generally be recommended on efficiency grounds, since the resulting finite sample bias (which increases with the number of instruments included in the regression) is small. The results of table 2 confirm this pattern: standard errors fall from 0.035 to 0.031 as the set of instruments is expanded from x_5 to include x_5 to x_2. Finally, the Sargan tests (last column of the table) do not reject any of the specifications at the 5% level when x_4 to x_2 are included in the set of instruments.

While the findings of table 2 provide no indication of serial correlation in the disturbances, the story somewhat changes when we turn to period 5 results, reported in table 3.

Table 3: IV results for period five of the panel

$\hat{\beta}$	instr. set : x_1	x_2	x_3	x_4	Sargan test
0.599(0.034)	×				–
0.577(0.033)	×	×			1.796
0.564(0.032)	×	×	×		2.166
0.545(0.031)	×	×	×	×	14.040*

1. The dependent variable is y_5 (the child's 1987 income). The set of explanatory variables contains x_5 (the parents' 1971 income) together with age controls.
2. A * mark indicates that the Sargan test of over-identifying restrictions is rejected at the 5% level.

The estimates of table 3 use lags of family income as instruments, while those of table 2 used leads of parental income. Note first that the estimates of β in table 3 are generally larger (ranging between 0.55 and 0.60) than those of table 2, though these differences are not statistically significant. On the basis of the Sargan tests statistics we may also conclude that the findings reject the validity of a one period lagged income as an instrument at the 5% (also at the 1%) significance level. Conditional upon x_1 (income lagged four periods) being a valid instrument, x_2, and $[x_2, x_3]$ are not rejected as over-identifying instruments, while the inclusion of x_4, leads to a rejection of the orthogonality conditions. Thus, period 5 results suggest an MA(1) specification for the regression error term, while those of period 1 give no indication of any potential serial correlation in the disturbances. It is for this reason that we will also subject our system-level estimators to various specification tests, given that the set of instruments may be called to doubt in the light of our findings.

We close this sub-section by reporting some related findings in the literature. Using education of the household head as an instrument, Solon (1992) estimates β at 0.56 using a sub-sample of the PSID. Abul Naga (1998) estimates this same parameter at 0.63 (also using education as an instrumental variable for a sub-sample of the PSID). Some UK results provided by Dearden et al. (1997) estimate β to be in the order of 0.55 to 0.6 when education is used to instrument parental income. They also report an interesting finding that the incomes of daughters tend to be more correlated with their parents' economic outcomes than in the case of sons.

4.2 System estimation

Moving from the single equation framework to the system estimator is analogous to the transition from two-stage to three stage least squares methods in the general simultaneous equations framework. This allows for potential efficiency gains provided the specifications considered are validated using appropriate diagnostics tests.

Table 4 : System estimators

Model	System − OLS	System − GLS
$MA(0)$	0.498(0.026)	0.417(0.022)
$MA(1)$	0.520(0.027)	0.475(0.024)
$MA(2)$	0.538(0.027)	0.489(0.025)
$MA(3)$	0.565(0.030)	0.522(0.027)

1. MA(0) denotes a model with serially uncorrelated residuals, while MA(q) denotes a model with a moving average disturbance term of order q.
2. The system-OLS and system-GLS estimators are as defined in equations (14) and (21) or (34) respectively.

In comparison to the single equation results of tables 2 and 3, we may note that the system GLS estimate is substantially smaller under the MA(0) assumption - no serial correlation in the errors - (a value of 0.42) but that this estimate increases to 0.52 with a standard error of 0.027 under the MA(3) specification The system GLS estimator appears to be somewhat more sensitive to the time-series structure of the error term than the system OLS estimator, though with the OLS estimator we also find that the estimate of β increases as we move from the MA(0) specification to the MA(3) model, with values ranging from 0.50 to 0.57.

As discussed in section 3, the existence of serial correlation in the residuals reduces the set of potential instruments for parental income. Thus, while under the MA(0) and MA(1) assumptions all equations can be used to arrive at an estimate of β, this is no longer the case under the MA(2) and MA(3) specifications. Under the MA(2) specification, x_3 cannot be instrumented using lags and leads of x (when the panel spans a period of 5 years). Likewise, under the MA(3) structure, x_2, x_3 and x_4 will not possess within-equation instruments. Accordingly, for the MA(2) specification the period 3 equation has been deleted from the system estimator. Likewise, the MA(3) model only pools period one and period 5 data to estimate β. The resulting pattern for the standard errors of the system estimator is to take on increasing values as we move from the MA(0) to the MA(3) models (for both OLS and GLS).

Finally, note that the variance of the individual effect accounts for approximately 70% of the total variance in the error term: σ_α^2 is estimated at 0.354, while σ_ε^2 is estimated at 0.166 (results not shown in the table). Viewed under this perspective, it is clear that panel data models will provide a better fit for microdata with a substantial component of unobserved heterogeneity than pure cross-section regressions.

4.3 Hausman specification tests

Our final task in this empirical section of the paper consists in testing various scenarios regarding the time-series structure of the disturbances of the system estimator. Such tests are required in order to provide a sound basis for discriminating between the various estimates of β in table 4, with values ranging from 0.42 to 0.57.

Table 5 : Hausman specification tests

Specification	System − OLS	System − GLS
$MA(2)$ *vs.* $MA(3)$	11.525	16.539∗
$MA(1)$ *vs.* $MA(3)$	16.942 ∗∗	19.897 ∗∗
$MA(0)$ *vs.* $MA(3)$	25.232 ∗∗	52.449 ∗∗
$MA(1)$ *vs.* $MA(2)$	8.649	7.199
$MA(0)$ *vs.* $MA(2)$	18.661 ∗∗	46.828 ∗∗
$MA(0)$ *vs.* $MA(1)$	0.0127	49.246 ∗∗

A * (**) indicates that the null hypothesis is rejected at a 5% (1%) probability of type-I error.

Our benchmark assumption is that of an MA(3) process: the resulting system estimator will be consistent (but inefficient) under the MA(2), MA(1), or MA(0) specifications, while, the other way around, the MA(q) estimator is inconsistent under the MA(r) scenario, for $r > q$.

Specification tests are carried out using both the system-OLS and system-GLS estimators, as, in practice, a sub-optimal weighting scheme for the GLS estimator need not result is efficiency gains over OLS (Szroeter (1994)). Comparing our test statistics to a χ^2 variable with 6 degrees of freedom, our results (table 5) can be summarized as follows: the MA(1) and MA(0) specifications are rejected at the 1% level against the null hypothesis of an MA(3) process using both the system OLS and GLS estimators. The MA(2) hypothesis (first line of table 5) is not rejected using the system OLS estimator, though it is rejected at the 5% level (but not at 1%) using GLS. In this sense, the MA(1) and MA(0) assumptions appear to be inadequate, while, provided the MA(3) hypothesis is a valid one, the MA(2) model appears to be more plausible.

The last three lines of table 5 provide further results when the null specification is modified. A test of the MA(1) specification against the MA(2) model does not lead to a rejection (using both estimators) while the MA(0) assumption is rejected against MA(2) at the 1% level. Finally, the GLS estimator rejects the assumption of no serial correlation in the error term, against the MA(1) model, while OLS fails to reject the MA(0) model against MA(1). The overall pattern, then, is that both estimators reject the MA(q-2) model against MA(q), while system-OLS never rejects the MA(q-1) scenario against MA(q). Adopting a more global perspective, the system-GLS estimator is perhaps to be recommended over system-OLS on such grounds.

5 Conclusions

In this study, we have carried out the estimation of the intergenerational correlation of permanent incomes within the framework of a panel data model with measurement errors. This has enabled us to implement new estimators as well as specification tests, as the problem has not been explicitly considered from this perspective so far.

The availability of time-series on a cross-section of families has allowed us to exploit within-sample information to deal with the errors in variables problem as well as to develop system estimators for the parameters of interest. The panel data context has also been useful in testing various scenarios regarding the time-series structure of the error term and its implications for the choice of instruments. Our estimators have also taken into account possible correlations among children belonging to the same family.

For the various methods used in this study, we have estimated the coefficient of income inheritance to be in the range of 0.42 to 0.60. Single equation estimates vary between 0.53 and 0.60. Results vary depending on the choice of instruments for a given time period, but also from one period to another. System estimates were found to be between 0.42 and 0.57, with the system OLS yielding higher values than the GLS counterpart. On the basis of our specification tests we find that MA(0) (no serial correlation) is definitely rejected compared to MA(3), MA(2) and MA(1) (at least in the system-GLS case). Turning to MA(1), it is rejected compared to MA(3) but not compared to MA(2). Finally MA(2) is accepted against MA(3). Given that MA(3) is the least restricted model that our data permit and hence using this as a reference for comparing the other scenarios, we can say that there is some supporting evidence for an MA(2) structure of the errors.

These results are taken to imply a substantial amount of income continuity from one generation to the next. A slope value in the order of 0.5 to 0.55, while indicative of a process of regression towards the mean, is nonetheless far remote from an idealized equal opportunities situation.

In general, panel data models allow the researcher to relax the independence assumption between the specific effect and the explanatory variable. However, this was not possible in our study because the variable of interest subject to errors (namely permanent income of the parent family) is time-invariant. Furthermore, we have not considered other dynamic structures for the error term such AR or ARMA processes, which could provide a useful direction for further work.

References

Abul Naga, R. (1998), " Estimating the Intergenerational Correlation of Incomes: an Errors in Variables Framework," Discussion paper No.9812, DEEP, University of Lausanne.

Altonji, J. and T. Dunn (1991), "Relationships Among the Family Incomes and Labor Market Outcomes of Relatives," *Research in Labor Economics*, 12, 269-310.

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

Behrman, J. and P. Taubman (1990), "The Intergenerational Correlation between Children's Adult Earnings and their Parents Income: Results from the Michigan Panel of Income Dynamics," *Review of Income and Wealth*, 36, 115-127.

Bicketti, S., W. Gould, L. Lillard and F. Welch (1988), "The Panel Study of Income Dynamics after Fourteen Years: an Evaluation," *Journal of Labor Economics*, 6, 472-92.

Biorn, E. (1998) "Panel Data with Measurement Errors: Instrumental Variables and GMM Procedures Combining Levels and Differences", Invited Paper at the Eighth International Conference on Panel Data, Goteberg.

Bjorklund, A. and M. Jantti (1997), "Intergenerational Income Mobility in Sweden Compared to the United States," *American Economic Review* 87, 1009-1018.

Bowles, S. (1972), "Schooling and Inequality from Generation to Generation," *Journal of Political Economy*, 80, S219-S251.

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

Gallant, A. and D. Jorgenson (1979), " Statistical Inference for a System of Simultaneous, Non-linear, Implicit Equations in the Context of Instrumental Variable Estimation," *Journal of Econometrics*, 11, 275-302.

Galor, O. and J. Zeira (1993), " Income Distribution and Macroeconomics, " *Review of Economic Studies*, 60, 35-52.

Godfrey, L. (1988), *Misspecification Tests in Econometrics*, Cambridge, Cambridge University Press.

Greene, W.H. (1993), *Econometric Analysis*, Second Edition, Macmillan Publishing Company, New York.

Griliches, Z. (1986), "Economic Data Issues" in Z. Griliches and M. Intriligator (eds): *Handbook of Econometrics*, Vol. 3, North Holland, Amsterdam.

Griliches, Z. and J. Hausman (1986), "Errors in Variables in Panel Data," *Journal of Econometrics*, 31, 93-118.

Hill, M. (1993), *The Panel Study of Income Studies: a User's Guide*, New York, Sage Publications.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Hsiao, C. and G. Taylor (1991), "Some Remarks on Measurement Error and the Identification of Panel Data Models," *Statistica Neerlandica* 45, 187-194.

Pagan, A. and A. Hall (1983), "Diagnostic Tests as Residual Analysis," *Econometric Reviews*, 2, 159-218.

Sargan, D. (1976), "Testing for Misspecification after Estimating using Instrumental Variables", mimeo, London School of Economics.

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

Stokey, N. (1996), "Shirtsleeves to Shirtsleeves: The Economics of Social Mobility", reprinted in D. Jacobs, E. Kalai and M.Kamian (eds) (1998): *Frontiers of Research in Economic Theory: Nancy L. Schwartz Memorial Lectures 1983-1997*, Cambridge University Press, Cambridge .

Szroeter, J. (1994), "Exact Finite-Sample Relative Efficiency of Suboptimally Weighted Least Squares Estimators in Models with Ordered Heteroscedasticity", *Journal of Econometrics*, 64, 29-43.

Wansbeek, T. and R. Koning (1991), "Measurement Error and Panel Data," *Statistica Neerlandica*, 45, 85-92.

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