Società italiana di economia pubblica

XXI CONFERENZA

PUBLIC CHOICE E POLITICAL ECONOMY
I fondamenti positivi della teoria di finanza pubblica

Pavia, Università, 24-25 settembre 2009

SECONDARY EFFECTS IN EDUCATION TRANSMISSION: A FINITE MIXTURE MODEL

VALENTINO DARDANONI, ANTONIO FORCINA E SALVATORE MODICA

società italiana di economia pubblica

dipartimento di economia pubblica e territoriale – università di Pavia

SECONDARY EFFECTS IN EDUCATION TRANSMISSION: A FINITE MIXTURE MODEL

VALENTINO DARDANONI, ANTONIO FORCINA, SALVATORE MODICA *

Abstract. Education is transmitted from parent to child 'primarily' through its effect on the child's endowment; but there is also an additional, 'secondary' effect *given* the child's endowment, which presumably operates through the social environment. Using data from the NCDS, we find that this secondary channel is gender dependent: fathers exert significant influence only on their sons, and mothers have a weaker influence, only on daughters.

These results are obtained in a finite-mixture model where educational achievements and auxiliary discrete multivariate responses are used to identify the underlying endowment. Generalizing existing models which assume independence of responses conditional on endowments, within our framework the effect of covariates and the temporal dependence among responses can be modelled in a flexible way.

JEL Classification Numbers: I21, C35

Keywords: Education Transmission, Finite Mixtures.

1. introduction

It is by now well appreciated that finding a positive association between parents' and children's education may simply reflect correlation between underlying unobserved heritable endowments: better endowed

Date: May 25, 2009.
* *Affiliations.* Dardanoni: Facoltà di Economia Università di Palermo, vdardano@unipa.it; Forcina: Facoltà di Economia Università di Perugia, forcina@stat.unipg.it Modica: Facoltà di Economia, Università di Palermo modica@unipa.it. We gratefully acknowledge financial support by Italian MIUR.

parents are more educated and have better endowed, hence more educated, children. Following the seminal contribution of Becker and Tomas on intergenerational income transmission, as adapted by Solon to the educational context (see also [21]), estimation of the causal effect of parents schooling on the educational achievements of their children is typically based on the simple linear reduced form equation

$$S^c = \alpha + \beta S^p + \gamma R^p + \delta U^p + \epsilon \tag{1}$$

where S^c and S^p denote child's and parent's schooling, R^p and U^p parent's child-rearing ability and unobservable endowments.

Estimation of β in equation (1) by controlling for U^p has been the object of recent research, starting from Behrman and Rosenzweig [4] who use differences on MZ twin parents and find that only father's education as an effect on childrens'. In two important follow-ups, Plug [27], by using adoptees, mostly confirms Behrman and Rosenzweig's findings, while Black–Devereux–Salvanes [6], by using reforms in municipal compulsory schooling laws as instruments, find almost no causal link between parents' and children's education.[1]

Sociologists (see e.g. Erikson *et al.* [15] and Jackson *et al.* [23]) have been actively investigating the effect of family background on children's schooling by following an early important distinction made by Boudon [8] between "primary" and "secondary" causal effects, where primary effects concern the influence of family background on children's scholastic attainment exterted through changes induced on children's

[1]On Behrman–Rosenzweig [4] see also the critical Comment [1] by Antonovics and Goldberger (and the authors' Reply [5]). The difficulties in estimating β are illustrated in the survey by Holmlund–Lindahl–Plug [21], where these three different methods (namely use of twins, adoptees and schooling laws instruments) are applied to a single data set, and it is shown that the three approaches produce results which are in conflict with each other.

academic ability, while secondary effects are those that relate family background to actual scholastic attainment given academic ability. This distinction is similar to the one between "indirect" and "direct" effects which is currently used in the causal inference literature (see for instance [26], [28], [30]).

In this simple linear setting, this distinction may be clarified by examining the following equations which jointly imply (1) above:

$$U^c = d + eS^p + fR^p + gU^p + \eta \tag{2}$$

$$S^c = a + bS^p + cU^c + \psi \tag{3}$$

where U^c denotes *child*'s unobservable endowment, equation (2) studies how parental background helps children to develop a given endowment, and equation (3) studies how parent's schooling helps actual scholastic attainment given that level of endowment. The primary effect of parent's schooling is thus measured by $c \cdot e$ while the secondary effect is measured by b; the causal effect β corresponds to $b + c \cdot e$.[2] If one aims to estimate β, the analysis can be based on equation (1) which requires some suitable device (say twins, adoptees, or instruments) that controls for the parents' unobservable endowment. Estimation of the secondary effect, the b parameter above, constitutes instead a less ambitious task since it requires to control for the child's endowment for which more reliable information may be available.

The present paper is an attempt to estimate the secondary effect of parents' education on children's educational achievement by exploiting recent advances in finite mixture models. The interpretation of a possible secondary effect emerges by asking why, among subjects with a

[2] In Appendix A we show that a similar decomposition also holds in the non-linear setting of the present work.

given level of schooling endowment, those with more educated parents should attain higher educational levels. Possible explanations include, for example, the intergenerational transmission of education-dependent labor market skills, better information on the value of education, instillation of more ambitious schooling preferences, or simply greater parental pressure reflecting social norms. Note that in any given social context, this kind of influence (which may be called a *role effect*) is likely to be gender dependent, in the sense that mothers and fathers may have different effects on daughters and sons.

Our methodology is applied to the English NCDS dataset which is discussed in detail in section 2. As main dependent variable we consider, rather than years of schooling, a binary indicator of schooling *attainment* of a significant educational certification, since this is more likely to reflect the value assigned to formal education by the students and their parents. The unobservable endowment is interpreted as schooling potential at the time of the exams (in our case 16 years of age).[3] We identify this latent variable with a finite mixture model which exploits the fact that the dataset contains information on a rich set of variables concerning different kinds of abilities measured from early age. Notice that, as argued above, we are not seeking identification of parents' unobsevable endowments. Omission of parents' endowment in (2) clearly makes any estimate of the primary effect biased, but estimation of the secondary effect is unaffected by this mispecification.

The paper has both methodological and empirical content. From the methodological point of view, the paper lays down the theoretical

[3]In particular, the analysis of *evolution* of endowments from early age is outside the scope of the paper. This important line of research is actively pursued by Heckman and associates, cfr. e.g. [11, 12, 19], who also arrive at evaluating dynamic complementarities of early and late interventions in the formation of skills.

framework for handling a complex mixture model with discrete multivariate responses where schooling attainment and the results of cognitive and non cognitive measurements are modelled jointly as dependent both on covariates and on a discrete unobserved endowment which may itself depend on covariates. Generalizing standard finite mixture models with multivariate discrete responses which assume independence of responses conditional on endowments we allow, more realistically, some dependence between early and late ability indicators.

The empirical findings confirm the presence of role effects in our sample: given the child's schooling ability, more educated parents bring their offsprings to a greater level of education. This influence, however, does not emerge if sons and daughters are pooled in the same sample. Indeed it is crucial to examine sons' and daughters' subsamples separately to uncover the effect we estimate, which then appears neatly: fathers' education affects *only* that of their sons; mothers' education has a weaker effect, and only on daughters. Of course, this gender-dependent link may well be confined to the social context to which our data refers.

The sequel is organized as follows. The data are described in section 2; we then derive the model to be estimated in section 3, which also contains a discussion of related literature. Results of estimation are reported in section 4. Section 5 contains some concluding remarks. In Appendices we elaborate on the decomposition of primary and secondary effects, provide details of the identification and estimation of our model, and collect auxiliary Tables.

2. The Data Structure

In the British educational system students at the age of 16 take the so called O-Level exams on a set of chosen topics. If a student has reached a minimum standard in terms of quantity of subjects taken and grades obtained, she is allowed, if she wishes, to access the next level of education (the so-called A-Level).

We use data from the National Child Development Survey (NCDS).[4] This data set is a UK cohort study targeting all the population born in the UK between the 3rd to the 9th of March 1958. Individuals were surveyed at different stages of their life and information on their schooling results and their background was collected. Our main dependent variable is the binary variable OL which takes value 1 iff the subject has passed at least five O-Level exams, which is the number of exams typically required for continuing the scholastic career. In our sample OL is equal to one for about fifty percent of the subjects. In the NCDS subjects are tested at the age of 7, 11 and 16 for mathematics, reading and general cognitive skills, and at the age of 7 and 11 information on non-cognitive skills is also collected; we use the results of these tests for identification of the unobservable endowment U^c. In particular, we first replace, at each age, the original maths and reading scores with the principal component (in all cases this explains no less than 90% of the total variance), thus creating three test score variables (7,11 and 16) for each math and reading. For non-cognitive skills (available at ages 7 and 11) analogous factor analysis yields two factors, such that the first factor may be interpreted as ability to relate to other individuals, while the second captures emotional problems. From these measurements of

[4]The same data set is extensively used, among others, by Blundell–Dearden–Sianesi [7], who study the effect of education on earnings.

cognitive and non cognitive skills we extract six variables by summing test scores at 7 and 11; we call them EM, LM, ER, LR, NP, NS, which are meant to capture early (7 and 11) and late (16) math and reading endowments, and (early) personal and social noncognitive skills. Since our finite mixture nodel requires use of discrete responses, we actually use discretized versions of the above variables which take value $k \in \{1, \ldots, K\}$ iff the subject is in the kth quantile of the corresponding continuous variable. In our application we use $K = 4$.

Since test scores take a large number of distinct values, one could alternatively have modelled them as continuous variables. Within finite mixture models, when a response variable is treated as continuous, a parametric form of the density conditional on covariates and the latent has typically to be specified; for instance one could assume a normal distribution where the mean (and perhaps the variance) depends on the latent and covariates. Thus, the price for using the true observed values of the test scores is that a specific conditional density is typically imposed. On the other hand, when the distribution is discretized, while some information is lost, the resulting density is multinomial and does not require any parametric restriction. When the number of discrete categories is not too small, this loss of information may have less serious drawbacks than imposing a parametric density.

Parents' schooling is defined as the age at which they left school;[5] fathers' and mothers' schooling are collected into the vector $\boldsymbol{s} = (S^f, S^m)$. Regarding other family background variables, the NCDS contains also information on parents' interest in their child's education, as reported by teachers separately for mothers and fathers; these can be considered

[5]Unfortunately NCDS does not have data on parents' scolastic attainment.

a proxies for the true child-rearing abilities. Data on parents' interest are originally classified into 5 distinct categories (overconcerned, very interested, shows some interest, little interest, can't say); we have created two parent's interest binary variables R^f and R^m which take value one if the parent is above the median after summing the original responses at the ages considered (7, 11 and 16). The two parents' interest variables are collected in the vector $\boldsymbol{r} = (R^f, R^m)$.

In the Becker-Tomes-Solon approach described by the reduced form equation (1) it has been argued that family income may possibly affect children's schooling attainments if families are credit constrained. A thorough and perspective review of the role of income in this model is contained in the survey [21] by Holmlund, Lindahl and Plug. Since the NCSD contains information on family incomes, we have created a family income variable for possible inclusion in the model discussed in the next section. However, using the income variable caused a drop of more than 50% of the sample size, and in all the configurations we estimated was never found significant. We thus decided not to use income as a covariate; the vector of all family background characteristics is then denoted by $\boldsymbol{x} = (\boldsymbol{s}, \boldsymbol{r})$.

From the NCDS we selected all subjects for whom we had information on OL, test scores and $\boldsymbol{x}$; the resulting sample is made of 5195 individuals, 2627 sons and 2568 daughters. A complete description of the data is available at

`http://www.esds.ac.uk/longitudinal/access/ncds`.

3. Model formulation

3.1. **The Equation System.** We wish to estimate the secondary effect of parents' schooling $\boldsymbol{s} = (S^f, S^m)$ on child's educational achievement OL given the child's unobservable endowment U^c by a finite mixture model. These models, essentially, decompose the probability distribution of an observed response vector $\boldsymbol{y}$ conditional on observed covariates $\boldsymbol{x}$ into a mixture of conditional probabilities involving an unobservable discrete random variable $\boldsymbol{v}$, subject to a suitable set of parametric restrictions which make the mixture identifiable. Formally one could write

$$P(\boldsymbol{y} \mid \boldsymbol{x}) = \sum\nolimits_{\boldsymbol{v}} P(\boldsymbol{v} \mid \boldsymbol{x})\, P(\boldsymbol{y} \mid \boldsymbol{v}, \boldsymbol{x});$$

in our context the response vector $\boldsymbol{y}$ includes the binary OL variable together with the set of cognitive and non cognitive test scores, so $\boldsymbol{y} = (OL, EM, LM, ER, LR, NP, NS) \equiv (OL, \boldsymbol{y}_{-ol})$. The covariates are $\boldsymbol{x} = (\boldsymbol{s}, \boldsymbol{r})$; and $\boldsymbol{v} = (U^c, U^p)$, which can be thought as a pair of discrete random variables which index distinct types of unobservable endowments for child and parents.

A crucial assumption of this paper is that parents' endowment U^p, while affecting child's endowment U^c, is irrelevant for predicting the response vector $\boldsymbol{y}$ once U^c is accounted for. Under this ignorability assumption, formally

$$P(\boldsymbol{y} \mid u^c, u^p, \boldsymbol{x}) = P(\boldsymbol{y} \mid u^c, \boldsymbol{x})\,, \tag{4}$$

marginalizing with respect to U^p in the above expression for $P(\boldsymbol{y} \mid \boldsymbol{x})$ and using (4) leads to

$$P(\boldsymbol{y} \mid \boldsymbol{x}) = \sum\nolimits_{u^c,u^p} P(u^c, u^p \mid \boldsymbol{x})\, P(\boldsymbol{y} \mid u^c, u^p, \boldsymbol{x}) = \sum\nolimits_{u^c} P(u^c \mid \boldsymbol{x})\, P(\boldsymbol{y} \mid u^c, \boldsymbol{x}). \tag{5}$$

Equation (5) implies that ignoring U^p affects the first component and thus prevents unbiased estimation of the effect of $\boldsymbol{x}$ on U^c (and of the primary effect), but does not affect the second component; in particular the effect of $\boldsymbol{x}$ on OL, which is the focus of this paper, can be estimated consistently.

We want U^c to identify *absolute* rather than residual (after adjusting for family background) latent abilities, for otherwise the secondary effect would be inflated. Thus we prevent test score responses to depend on $\boldsymbol{x}$ conditionally on U^c:

$$P(\boldsymbol{y}_{-OL} \mid U^c, \boldsymbol{x}) = P(\boldsymbol{y}_{-OL} \mid U^c)\,,$$

which should be seen as an assumption on the nature of the unobservable endowment jointly identified be the set of responses.

Notice that identifiability of mixture models with multivariate response variables is usually achieved by the restrictive assumption of independence of the responses conditionally on the latent (see Goodman [17] for the seminal paper on finite mixture model with multivariate binary responses, and Huang and Bandeen-Roche [22] for a recent general treatment under conditional independence). In this paper we allow dependence of late maths and reading scores on early ones and provide evidence that the model is still identifiable. This enhances flexibility of the model, since it seems realistic to assume that even conditioning on ability, early results help predicting later ones; our data strongly confirm this presumption.

Finally, to model the conditional probabilities of interest one has to choose an appropriate set of link functions for each response variable and the latent. A logit link is a natural candidate for the binary OL. All the remaining responses are ordinal and thus a natural choice is to use a set of *global logits*. The latent U^c instead is assumed to be purely qualitative with each category indexing a specific unobserved type; thus we use the so called *adjacent logits*, which provide the simplest link function for unstructured qualitative variables. Recall that for a variable X taking values in $1, \ldots, k$, global logits are defined as $\log\big(\Pr(X \geq x)/\Pr(X < x)\big)$, and adjacent logits are defined as $\log\big(\Pr(X = x)/\Pr(X = x-1)\big)$ for $x = 2, \ldots, k$.

In conclusion we estimate the following system, where for simplicity we write U instead of U^c, a discrete random variable taking values $1, 2, \ldots, m$:

$$P(OL = 1 \mid U = u, \boldsymbol{S} = \boldsymbol{s}) = F\big(\alpha_u^{OL} + \beta_f^{OL} s^f + \beta_m^{OL} s^m\big), \tag{6}$$

$$P(U = u \mid \boldsymbol{X} = \boldsymbol{x}) = G\big(\alpha_u^U + \beta_{u,f}^U s^f + \beta_{u,m}^U s^m + \gamma_{u,f}^U r^f + \gamma_{u,m}^U r^m\big), \quad u = 2, \ldots, m, \tag{7}$$

and, for $j = 2, 3, 4,\ em, er = 1, \ldots, 4$ and $u = 1, \ldots, m$

$$\begin{aligned}
&P(EM \geq j \mid U = u) = H\big(\delta_j^{EM} + \alpha_u^{EM}\big)\\
&P(LM \geq j \mid U = u, EM = em) = H\big(\delta_j^{LM} + \alpha_u^{LM} + \rho_{em}^{LM}\big)\\
&P(ER \geq j \mid U = u) = H\big(\delta_j^{ER} + \alpha_u^{ER}\big)\\
&P(LR \geq j \mid U = u, ER = er) = H\big(\delta_j^{LR} + \alpha_u^{LR} + \rho_{er}^{LR}\big)\\
&P(NP \geq j \mid U = u) = H\big(\delta_j^{NP} + \alpha_u^{NP}\big)\\
&P(NS \geq j \mid U = u) = H\big(\delta_j^{NS} + \alpha_u^{NS}\big),
\end{aligned} \tag{8}$$

where F, G and H are respectively the binary, adjacent and global logit link functions. In total there are $[(m-1)+7\times m]\,\alpha$'s, $[2+2\times(m-1)]\,\beta$'s, $(6\times 3)\,\delta$'s, $2\times(m-1)\,\gamma$'s and $2\times 4\,\rho's$. Notice however that, for each test score ρ_1 and δ_2 must be constrained to 0 to avoid a dummy variable trap and have each equation identified. Notice that child-rearing ability does not enter equation (6), in accordance with equation (3) in the introduction.

Summing up, our equation system is made, beside the OL equation of main interest, by an equation which specifies the conditional distribution of the unobservable heterogeneity, and a system of auxiliary equations which are instrumental for identifying it.

3.2. **Estimation.** The system of equations (6–8) can be written more compactly as

$$\boldsymbol{\lambda} = \boldsymbol{B}(\boldsymbol{x})\boldsymbol{\psi} \tag{9}$$

where

- $\boldsymbol{\lambda}$ is the vector which collects the m logits of OL, the $m-1$ logits of U, the $3\times m$ logits of each of the early math, early reading and non cognitive test scores, and the $3\times 4\times m$ logits for each of the late math and late reading test scores
- $\boldsymbol{B}(\boldsymbol{x})$ is a design matrix whose dependence on $\boldsymbol{x}$ reflects the effect of the covariates on the different elements of the joint distribution
- $\boldsymbol{\psi}$ is the vector which collects the model parameters α's, β's, γ's, δ's and ρ's.

The standard approach for maximum likelihood estimation of finite mixture models is the EM algorithm. Because it is natural to assume that the discrete responses of our model follow the multinomial distribution, the E-step is equivalent to compute, for each subject, the posterior probability of belonging to each latent class. The M-step requires maximization of a multinomial likelihood with individual covariates. Details on the implementation are described in the appendix. It has been shown (Dempster–Laird–Rubin [14]) that the algorithm converges to the maximum of the true likelihood. It is well known that the EM algorithm may converge even if the model is not identified, a crucial issue for finite mixture models. Since unfortunately there is no general result applicable to our context, we have used the numerical test described by Forcina [16]. This test implements the condition of parametric identification of Rothemberg [29], and consists in checking that the Jacobian of the transformation between the parameters of the saturated model for the observable responses and the mixture model parameters is of full rank for a wide range of parameter values.

3.3. **Discussion.** The finite-mixture approach is used in many branches of statistics such as biometrics and psychometrics and the resulting models are usually called latent class models (see e.g. [25, 31]). It is also becoming increasingly popular in econometrics (see e.g. Greene, ([18], Ch. 16) or Cameron and Trivedi, ([9], Ch. 18) for an introduction, and Heckman and Singer [20] for a seminal use in economics).

The assumption that U is discrete implies that no structure is imposed on child's unobservable endowments; in other words, each category corresponds to a distinct qualitative type. Thus unobservable endowments are allowed to be multidimensional, a feature often stressed, for instance, in the labor market literature (see e.g. [19]).

Unlike in the analyses of Heckman and collaborators [11, 12, 19], we do not attempt to model the dynamics of the unobservable individual heterogeneity U. This is so because our interest lies in capturing the unobservable schooling endowments at the time when OL exams are taken. U can be seen as a cross-classification of underlying abilities which accounts for different cognitive and non cognitive skills and for different evolutions over time. What matters here is that a sufficient number of types is used to capture the unobservable heterogeneity within a model which is identifiable.

4. Results

We start by estimating model (9) separately for sons and daughters under a different number of unobservable types. To ease interpretation of the estimated coefficients we first centered all covariates. Maximum likelihood estimation is performed by an EM algorithm as described in the appendix; computations are based on a set of Matlab functions available from the authors. The maximized log-likelihood $L(\hat{\boldsymbol{\psi}})$, Aikike Information Criterion $AIC(\hat{\boldsymbol{\psi}}) = -2L(\hat{\boldsymbol{\psi}}) + 2v$, and Schwartz's Bayesian Information Criterion $BIC(\hat{\boldsymbol{\psi}}) = -2L(\hat{\boldsymbol{\psi}}) + \ln(n)v$, where n denotes sample size and v is the number of parameters (which depends on the number of latent classes m), are given in Table 5 on page 30. In both samples $BIC(\hat{\boldsymbol{\psi}})$ is lowest with four latent classes and $AIC(\hat{\boldsymbol{\psi}})$ is lowest with five. We will report estimation results up to five classes, which seem sufficient to capture any unobserved heterogeneity contained in our samples.

To ease reading of the Tables we have reordered the types in a decreasing order of scholastic ability as measured by the probability of O-Level achievement (which can be done since finite mixture models are invariant to types' rearrangment). So $U = 1$ type has highest probability of achieving O-Level for given parental background.

4.1. **Schooling attainment.** Estimated parameters of the schooling attainment equation (6) page 11 are contained in Tables 1 and 2 below.

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{OL}	1.8364	0.0840	2.5611	0.1658	2.6280	0.2268	2.4268	0.2818
α_2^{OL}	-1.3569	0.0829	0.5987	0.0831	1.4577	0.1129	2.3957	0.2370
α_3^{OL}	-	-	-2.0429	0.1345	-0.8510	0.1071	0.5670	0.1419
α_4^{OL}	-	-	-	-	-3.5415	0.5486	-1.1811	0.1380
α_5^{OL}	-	-	-	-	-	-	-4.4320	1.3423
β_f^{OL}	0.0406	0.0441	-0.0075	0.0453	-0.0053	0.0452	-0.0079	0.0459
β_m^{OL}	0.1171	0.0514	0.0952	0.0523	0.0950	0.0529	0.1075	0.0528

TABLE 1. Daughters OL

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{OL}	1.5644	0.0789	2.2944	0.1376	2.4529	0.1807	2.6017	0.2024
α_2^{OL}	-1.7472	0.0884	-0.3446	0.0758	0.8778	0.1265	1.1766	0.1758
α_3^{OL}	-	-	-2.4201	0.1627	-1.5463	0.1613	-0.1178	0.2598
α_4^{OL}	-	-	-	-	-2.3635	0.1813	-1.5625	0.1594
α_5^{OL}	-	-	-	-	-	-	-2.4097	0.1881
β_f^{OL}	0.1860	0.0454	0.1360	0.0466	0.1431	0.0488	0.1102	0.0524
β_m^{OL}	-0.0439	0.0523	-0.0352	0.0535	-0.0828	0.0560	-0.0809	0.0599

TABLE 2. Sons OL

A glance at the two tables above confirms the gender-dependent secondary effects described in the introduction. Indeed, the coefficients of the parents' education variables in the OL equation (last two rows in the Tables) show that after controlling for the child's schooling endowments, the father's education significantly helps his son's chance of achieving OL certification, and not his daughter's; and that on the other hand, mothers' education helps daughters' schooling but not sons'. Notice that estimated β coefficients are similar in both samples under 3, 4 or 5 latent classes, and that father's effect on sons seems to be greater and more significant than mother's effect on daughters. The results above also show that the effect of being a high rather than low U type is overwhelming for O-Level achivement, as can be seen by looking at the α-coefficients. Since parents' schooling has been centered, each α indicates the logit of the probability of O-Level achivement for an individual with average parental schooling. Recall that in the logit scale a change of value, say, from -2 to +2 implies a change in the probability of success from about 12% to 88%.

The relevance of unobserved heterogeneity we find in this paper can be seen as a furtrher validation of the analyisis of Keane and Wolpin [24], who estimate a a dynamic structural model of schooling, work, and occupational choice decisions with four unobserved types. They find that unobserved endowment heterogeneity, as measured by innate talents and human capital accumulated up to the age of 16, accounts for 90 percent of the variance in lifetime utility. Understanding unobserved heterogeneity thus is of primary importance, and the work which is currently being done by Heckman and collaborators ([11, 12, 19]) seems a very promising step towards that direction.

4.1.1. **Secondary effects**. We translate the above estimates into a quantitative appraisal of the secondary effect. We consider the effect on OL attainment (conditional on unobserved endowment) of increasing a parent education by three years of schooling (which seems a reasonable measure of change of educational status), leaving unchanged the schooling of the other parent, starting from a situation where both parents have an average level of schooling (here denoted by μ^i, $i = f, m$).

The average effect of increasing a parent's schooling for a given level of child's endowment U can be calculated as:

$$\begin{aligned}\delta(u, S^i) &= \Pr(OL = 1 \mid S^j = \mu^j, S^i = \mu^i + 3, U = u) - \\ &\quad \Pr(OL = 1 \mid S^j = \mu^j, S^i = \mu^i, U = u) \\ &= \Lambda(a^{OL}(u) + 3\beta_i^{OL}) - \Lambda(a^{OL}(u)), \quad i, j = f, m, \quad u = 1, \ldots, m\end{aligned}$$

where $\Lambda(t) = \exp(t)/(1+\exp(t))$ denotes the logit link function, and the second equation follows since we have centered father's and mother's schooling. These effects can be consistently estimated using the coefficients in the OL-equation; furthermore, using the estimated variance matrix of $(a^{OL}(u), \beta_i^{OL})$, an asymptotic standard error can be derived by application of the delta method. Their numerical values are in the following table, for the case of 4 unobservable types:

Daughters

	$\delta(u, S^f)$	se	$\delta(u, S^m)$	se
U=1	-0.0010	0.0086	0.0158	0.0087
U=2	-0.0024	0.0210	0.0398	0.0202
U=3	-0.0033	0.0283	0.0629	0.0369
U=4	-0.0004	0.0036	0.0089	0.0074

Sons

	$\delta(u, S^f)$	se	$\delta(u, S^m)$	se
U=1	0.0262	0.0088	-0.0201	0.0148
U=2	0.0806	0.0247	-0.0540	0.0381
U=3	0.0709	0.0285	-0.0331	0.0205
U=4	0.0403	0.0175	-0.0176	0.0108

TABLE 3. Secondary Effects of Father's and Mothers' Schooling

The table shows that the secondary effect, measured as difference in probabilities of achievement, is non-negligible and statistically significant only for fathers on sons; notice also that direct effects are highly nonlinear in U. As an example, for fathers on sons we find that that for $U = 1$ the probability of getting OL certification passes from 0.9208 to 0.9470, and for $U = 2$ from 0.7064 to 0.7870.

4.1.2. **Pooled Sample**. For the sake of comparison we have estimated model (9) on the pooled sample of sons and daughters. Below for semplicity we only report the estimated results of the secondary effects β^{OL} in the case of 5 latent classes (in the pooled sample this corresponds to the lowest BIC value).

	coeff	se
β_f^{OL}	0.0418	0.0348
β_m^{OL}	0.0213	0.0386

Table 4. β^{OL} in pooled sample

As can be seen by comparing these estimated coefficients with the corresponding ones in the daugthers and sons subsamples in tables 1 and 2, the pooled sample estimates are approximately equal to the average of the corresponding estimates for daughters and sons. They suggest a somewhat stronger role of fathers, but the coefficients are not significant. Thus considering sons and daughters subsamples separately is crucial for clarifying the nature of this asymmetry.

4.2. **Endowments.** Estimates for the endowment equation (7) page 11 are contained in Tables 6 and 7 in the appendix. Both tables reveal a very strong positive association between child's endowment U and family background characteristics, since having more educated and/or more concerned parents decreases the odds of being a type $U = u + 1$ rather then $U = u$ for all u's in both samples (recall we have rearranged the types according to their probability of O-level achievement). This is seen from the negative significant values of most β's. However, as we discuss in section 3 above and in the introduction, since we do not control for parents' unobservable endowments, these estimates of the primary effect of parents schooling may be biased.

4.3. **Test scores.** Finally, parameters estimates of the auxiliary equation (8) page 11 are contained in Tables 8 to 13 in the appendix. The tables show that knowledge of the child's type helps overwhelmingly to predict both math and reading test scores, and to a lesser extent non cognitive test results. Notice also that while math and reading test scores are increasing in the latent types (that is, $U = u$ types have stochastically greater cognitive test scores than $U = u + 1$), this does not hold exactly in these samples for non cognitive scores. Notice also that in both samples even after conditioning on U, there is still in general a strong positive correlation between early and late test results, as shown by the generally significant values of the ρ parameters. This confirms that the increased generality of our model with regard to conditional dependence closely matches the complexity of the real processes under study. In practice, this dependence may be taken an an indication that certain subsets of responses require, in addition to the overall endowment captured by U, a certain amount of specific abilities.

5. Concluding Remarks

Within the broad issue of education transmission there are two important recent lines of research related to the present work. One analyzes the causal effect of parents' education on children's controlling for parents' unobserved endowment, by use of twin parents, adoptees or compulsory schooling law instruments ([1, 4, 5, 6, 21, 27]). The other, lead by Heckman ([11, 12, 19]) studies the effect of sequential interventions and their complementarity on the *evolution* of endowments and on labor market outcomes.

We follow an early insight of Boudon [8], who distinguished between 'primary' and 'secondary' causal effects, which is being actively investigated by sociologists ([15], [23]).[6] By applying recently developed finite mixture models to the NCDS dataset, the present paper controls for the child's own schooling endowments at the age of 16, and measures the secondary effect of parents' schooling on children's educational achievement at the same age. The effect of parents' schooling on children's educational attainment *given* the latter's potential mainly reflects parental pressure, and can thus be interpreted as a role effect. To allow for the possibility of its dependence on gender we consider sons and daughters subsamples separately. We find that fathers' education has a significant secondary effect on children's, but its impact is entirely confined to their sons; mothers' education has a slightly less significant secondary effect on daughters.

[6]([15] and [23] find that secondary effects are sizeable, accounting for roughly a quarter of the total effect. Their results are however not directly comparable with ours since estimation procedures, data employed and variable definitions are different.

This result may reflect social norms of Western families in the nineteen-seventies (the data domain), and if the women's role has changed a different picture may emerge from more recent data. But the message we get from our findings remains that children respond to family pressure on schooling attainment. From a policy –or rather 'cultural'– viewpoint this suggests that when parents' pressure is weak only the social environment, school primarily, can make up for this loss by helping the young to appreciate the value of education.

References

[1] Antonovics, K. L. and A. S. Goldberger (2005): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Comment", *American Economic Review* **95**, pp. 1738–1744

[2] Bartolucci, F., R. Colombi and A. Forcina (2007): "An Extended Class of Marginal Link Functions for Modelling Contingency Tables by Equality and Inequality Constraints", *Statistica Sinica*, **17**, pp. 691–711.

[3] Bartolucci, F. and A. Forcina (2006): "A Class of Latent Marginal Models for Capture-Recapture Data with Continuous Covariates", *Journal of the American Statistical Association* **101**, pp. 786–794

[4] Behrman, J. R. and M. R. Rosenzweig (2002): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?", *American Economic Review* **92**, pp. 323–334

[5] Behrman, J. R. and M. R. Rosenzweig (2005): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Reply", *American Economic Review* **95**, pp. 1745–1751

[6] Black, S. E., P. J. Devereux and K. G. Salvanes (2005): "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital", *American Economic Review* **95**, pp. 437–449

[7] Blundell, R., L. Dearden and B. Sianesi (2005): "Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National

Child Development Survey", *Journal of the Royal Statistical Society A*, **168**, pp. 473–512

[8] Boudon, R. (1974): *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society* Wiley, New York

[9] Cameron, C. A. and P. K. Trivedi, (2006): *Microconometrics: Methods and Applications*, CUP.

[10] Catchpole, E. A., and Morgan, B. J. T. (1997): "Detecting Parameter Redundancy," *Biometrika*, 84, pp. 187-196.

[11] Cunha, F. and J. J. Heckman (2007): "The Technology of Skill Formation", *American Economic Review* **97**, pp. 31-47

[12] Cunha, F., J. J. Heckman and S. Schennach (2007): "Estimating the Technology of Cognitive and Noncognitive Skill Formation", Working Paper

[13] Dardanoni, V. and A. Forcina (2008): "Multivariate ordered regression", *mimeo*

[14] Dempster, A. P., N. M. Laird and D. B. Rubin (1977): "Maximum Likelihood for Incomplete Data Via the EM Algorithm", *Journal of the Royal Statistical Society* Series B **39**, pp. 1–22

[15] Erikson, R., J. H. Goldthorpe, M. Jackson, M. Yaish, and D. R. Cox (2005): "On Class Differentials in Educational Attainment", *Proceedings of the National Academy of Science* **102**, pp. 9730–9733

[16] Forcina, A. (2008): "Identifiability of Extended Latent Class Models with Individual Covariates," *Computational Statistics and Data Analysis*, 52, pp. 5263-5268

[17] Goodman, L. (1974): "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models", *Biometrika* **61**, pp. 215–231

[18] Greene, W. (2008): *Econometric Analysis*, Prentice Hall.

[19] Heckman, J., J. Stixrud and S. Urzùa (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior", *Journal of Labor Economics* **24**, pp. 411-482

[20] Heckman, J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, **52**, pp. 271-320

[21] Holmlund, H., Lindahl, M. and E. Plug (2008): "The Causal Effect of Parent's Schooling on Children's Schooling: A Comparison of Estimation Methods, *IZA Discussion Paper*, **3630**.

[22] Huang G. and K. Bandeen-Roche (2004): "Building an Identifiable Latent Class Model, with Covariate Effects on Underlying and Measured Variables" *Psychometrika* **69**, pp. 5-32

[23] Jackson, M., Erickson, R., Goldthorpe, J.H. and M. Yaish (2007): "Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-Level in England and Wales" *Acta Sociologica* **50**, pp. 211-29

[24] Keane, M. P. and K. I. Wolpin (1997): "The Career Decisions of Young Men", *Journal of Political Economy* **105**, pp. 473-522

[25] Lindsay, B., C. Clogg, and J. Grego (1991): "Semiparametric Estimation of the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis", *Journal of the American Statistical Association* **86**, pp. 96-107

[26] Pearl, J. (2000): *Causality*, Cambridge University Press

[27] Plug, E. (2004): "Estimating the Effect of Mother's Schooling Using a Sample of Adoptees", *The American Economic Review* **94**, pp. 358-368

[28] Robins, J. M. (2004) "Testing and estimation of direct effects by reparameterizing directed acyclic graphs", XXX

[29] Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, **39**, pp. 577-591.

[30] Rubin, D. B. (2004), "Direct and indirect causal effects via potential outcomes", *Scandinavian Journal of Statististics*, **31**, 161-170

[31] Vermunt, J. K. and J. Magidson (2003), "Latent Class Analysis", in *Encyclopedia of Research Methods for the Social Sciences*, Michael S. Lewis-Beck, Alan Bryman and Tim Futing Liao editors, Sage Publications, NewBury Park.

Appendix A: Primary and secondary causal effect

In this appendix we derive an expression for the decomposition of total effect of parents' education on their children's into primary and secondary effects which extends to a general context the decomposition for the linear model given in the Introduction Consider the following causal diagram where Y is a binary outcome, X is a discrete input of interest, U and V are discrete unobservable variables and T is also a discrete response:

$$\begin{tikzcd} & V \arrow[r] \arrow[d] & X \arrow[dl] \arrow[d] \\ T & U \arrow[l] \arrow[r] & Y \end{tikzcd}$$

This is a simplified version of the probability model described in (8) where Y represents scholastic attainment and X parent's years of schooling while U and V are meant to capture, respectively, child's and parent's unobservable endowments and T represents the response to a set of ability tests. For conciseness and when no ambiguity arises, we use the convention that a conditioning variable will be denoted by its value, thus for example $\Pr(Y \mid x, v)$ means $\Pr(Y \mid X = x, V = v)$. Let also

$$\Delta_{S,s}(t_1, t_0 \mid w) = \Pr(S = s \mid t_1, w) - \Pr(S = s \mid t_0, w)$$

denote the causal effect of setting $T = t_1$ rather than $T = t_0$ on the probability of $S = s$ when controlling for $W = w$.

Proposition. *Under the conditional independencies encoded in the causal diagram above,*

$$\begin{aligned} \Delta_{Y,1}(x+1, x \mid v) &= \sum\nolimits_u \Delta_{Y,1}(x+1, x \mid u) \Pr(U = u \mid x, v) \\ &+ \sum\nolimits_{u > u_0} \Delta_{U,u}(x+1, x \mid v) \Delta_{Y,1}(u, u_0 \mid s+1), \end{aligned}$$

where $U \geq u_0$.

Proof. By using the conditional independence between Y and V given X, U,

$$\Pr(Y|x,v) = \sum\nolimits_u \Pr(Y \mid x,u) \Pr(U = u \mid x,v);$$

by substitution, the expression for $\Delta_{Y,1}(x+1, x \mid v)$ may be expanded as

$$\sum\nolimits_u \big[\Pr(Y = 1 \mid x+1, u) \Pr(U = u \mid x+1, v) \\ - \Pr(Y = 1 \mid x, u) \Pr(U = u \mid x, v) \big].$$

The result follows by adding and subtracting $\sum_u \Pr(Y = 1 \mid x + 1, u) \Pr(U = u \mid x, v)$, rearranging terms and noting that, if u_0 is the minimum of U, we may write $\Pr(U = u_0 \mid x, v) = 1 - \sum_{u>u_0} \Pr(U = u \mid x, v)$. $\square$

The first component may be interpreted as a weighted average of the secondary effect of X on Y while the second term is the sum of the products of the effect of X on U times the effect of U on Y and may be interpreted as a measure of the primary effect. Though the above decomposition is not unique, it is one way of generalizing the decomposition which holds in the system of linear equations given in the introduction.

Finally, notice that the causal diagram above implies that, if we do not control for V, there is an omitted variable bias in the calculation of the effect of X on U unless X and V are independent.

Appendix B: Estimation and Identifiability

The true log-likelihood. Let $\boldsymbol{n}(i)$ be the vector of size $t = 2 \times 4^6$ containing the frequency table of the response variables $\boldsymbol{Y}$ in lexicographic order for the subjects with covariate $\boldsymbol{x}_i$; if there is a single subject with such features, $\boldsymbol{n}(i)$ is a vector of 0s except for a 1 in the cell corresponding to the response pattern $\boldsymbol{y}(i)$. We will denote by s the number of strata, i.e. the number of different covariate configuration in the sample.

Let also $\boldsymbol{q}(i)$ denote the vector whose elements $q(i, \boldsymbol{y})$ equal to the probability of the response pattern $\boldsymbol{y}$ for subjects with covariate $\boldsymbol{x}_i$. The log-likelihood may be written as

$$L(\boldsymbol{\psi}) = \sum L_i(\boldsymbol{\psi}) = \sum \boldsymbol{n}(i)' \ln[\boldsymbol{q}(i)].$$

Here $\boldsymbol{\psi}$ is the full vector of parameters defined in equation (9); in the right hand side it is $\boldsymbol{q}$ that depends on it, we have omitted dependence to ease reading. Let $\boldsymbol{p}(i)$ denote the vector with elements $p(i, u, \boldsymbol{y})$ equal to the probability of the event $(u, \boldsymbol{y})$ (again in lexicographic order with the categories of U running slowest) for subjects with covariate $\boldsymbol{x}_i$ and $\boldsymbol{m}(i)$ be the vector containing the corresponding unobservable frequency table; finally let $\boldsymbol{L} = ({\boldsymbol{1}_{m+1}}' \otimes \boldsymbol{I}_t)$ denote the matrix which transforms the latent frequencies into the observed ones, so that $\boldsymbol{n}(i) = \boldsymbol{L}\boldsymbol{m}(i)$ and $\boldsymbol{q}(i) = \boldsymbol{L}\boldsymbol{p}(i)$. Because $\boldsymbol{\psi}$ is defined in terms of the latent probabilities, maximizing $L(\boldsymbol{\psi})$ may be seen as a missing data problem which may be tackled by the EM algorithm (Dempster–Laird–Rubin [14]). If the latent U could be observed, the corresponding log-likelihood would have the form

$$\Lambda(\boldsymbol{\psi}) = \sum \Lambda_i(\boldsymbol{\psi}) = \sum \boldsymbol{m}(i)' \ln[\boldsymbol{p}(i)].$$

The E step. Because the multinomial is a member of the exponential family, the conditional expectation involved in the E step is equivalent to computing the so called posterior probability of latent class U given the observed configuration $\boldsymbol{y}$

$$\Pr(U \mid \boldsymbol{y}, \boldsymbol{x}_i) = \frac{p(i, U, \boldsymbol{y})}{q(i, \boldsymbol{y})}$$

so that $m(i, u, \boldsymbol{y}) = n(i, \boldsymbol{y}) \Pr(u \mid \boldsymbol{y}, \boldsymbol{x}_i)$ follows from a simple expectation of a multinomial distribution for U.

The M step. Implementation of the Fisher scoring algorithm which maximizes $\Lambda(\boldsymbol{\psi})$ with respect to the model parameters $\boldsymbol{\psi}$ requires computation of the score vector (first derivative with respect to $\boldsymbol{\psi}$) and of the expected information matrix (minus the expected value of the second derivative). Since $\Lambda(\boldsymbol{\psi})$ is a multinomial log-likelihood, exponential family results make such calculations straightforward. In brief, we rewrite Λ in terms of the canonical parameters of the multinomial distribution, say $\boldsymbol{\theta}_i$, and exploit the fact that there are invertible and differentiable mappings from $\boldsymbol{\theta}_i$ to the vector of probabilities $\boldsymbol{p}(i)$ and from $\boldsymbol{p}(i)$ to $\boldsymbol{\lambda}_i$ (the latter mapping is described in Theorem 1 of [2]), while $\boldsymbol{\lambda}_i$ is linked to $\boldsymbol{\psi}$ by the linear regression model. Calculations of the score vector and the expected information matrix are described in the following section.

Asymptotic variances. Though the EM algorithm is very reliable, by itself, it does not provide a consistent estimate of the variance matrix of the model parameters. An estimate of the variance matrix may be derived from the expected information matrix of the true likelihood as follows. Write $L_i(\boldsymbol{\psi}) = \boldsymbol{n}(i)'\tilde{\boldsymbol{G}}\boldsymbol{\gamma}_i - n_i \ln[\mathbf{1}' \exp(\tilde{\boldsymbol{G}}\boldsymbol{\gamma}_i)]$ where $\boldsymbol{\gamma}_i$, the canonical parameter of the observed multinomial, is defined by $\tilde{\boldsymbol{H}} \ln[\boldsymbol{L} \exp(\boldsymbol{G}\boldsymbol{\theta}_i)/\mathbf{1}' \exp(\boldsymbol{G}\boldsymbol{\theta}_i)]$ where $\tilde{\boldsymbol{H}}$ is a $(t-1)\times t$ contrast matrix of full row rank, $\tilde{\boldsymbol{G}}$ is its right inverse and $\boldsymbol{G}$ is the design matrix which defines the latent canonical parameters $\boldsymbol{\theta}_i$ and has v columns of full rank. Let $\boldsymbol{\gamma}$ denote the vector obtained by stacking the $\boldsymbol{\gamma}_i$ one below the other, similarly for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. By the chain rule

$$\frac{\partial L(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'} = \frac{\partial L(\boldsymbol{\psi})}{\partial \boldsymbol{\gamma}'} \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\psi}'} = \boldsymbol{\delta}' \boldsymbol{D},$$

where $\boldsymbol{\delta}$ has blocks of the form $\tilde{\boldsymbol{G}}'(\boldsymbol{n}(i) - n_i \boldsymbol{q}_i)$ with $n_i = \mathbf{1}'\boldsymbol{n}(i)$. Because, obviously, $E(\boldsymbol{\delta}) = \mathbf{0}$, it follows that the information matrix takes the form

$$\boldsymbol{F} = \boldsymbol{D}'\, diag[n_1 \boldsymbol{\Omega}_1, \ldots, n_s \boldsymbol{\Omega}_s] \boldsymbol{D},$$

where $\boldsymbol{\Omega}_i = diag(\boldsymbol{q}_i) - \boldsymbol{q}_i \boldsymbol{q}_i'$ and s is the number of strata. An explicit expression for computing $\boldsymbol{D}$ may be derived by the chain rule as follows

$$\boldsymbol{D} = \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\theta}'} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\lambda}'} \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\psi}'} = \boldsymbol{Q}\boldsymbol{R}\boldsymbol{B};$$

because $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are made of s blocks, one for each stratum, $\boldsymbol{Q}$ and $\boldsymbol{R}$ are block diagonal with elements

$$\begin{aligned} \boldsymbol{Q}_i &= \frac{\partial \boldsymbol{\gamma}_i}{\partial \boldsymbol{q}_i'} \frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{p}_i'} \frac{\partial \boldsymbol{p}_i}{\partial \boldsymbol{\theta}_i'} = \bar{H}\, diag(\boldsymbol{q}_i)^{-1} \boldsymbol{L} \boldsymbol{\Omega}_i \boldsymbol{G} \\ \boldsymbol{R}_i &= \left[\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{p}_i'} \frac{\partial \boldsymbol{p}_i}{\partial \boldsymbol{\theta}_i'} \right]^{-1} = \left[\boldsymbol{C}\, diag(\boldsymbol{M}\boldsymbol{p}_i)^{-1} \boldsymbol{M} \boldsymbol{\Omega}_i \boldsymbol{G} \right]^{-1}. \end{aligned}$$

Model identifiability. A model is globally identified when the mapping from $\boldsymbol{q}$ (or equivalently $\boldsymbol{\gamma}$) and $\boldsymbol{\psi}$ is one to one. The weaker notion of local identifiability (see Catchpole and Morgan, [10]), which is easier to verify, suffices to ensure that asymptotic approximations hold. This requires that at any $\boldsymbol{\psi}_0$ the set of points such that $\|\boldsymbol{q}(\boldsymbol{\psi})-\boldsymbol{q}(\boldsymbol{\psi}_0)\| \leq \epsilon$ satisfy $\|\boldsymbol{\psi} - \boldsymbol{\psi}_0\| > \delta > 0$. The results of Catchpole and Morgan (1997, Theorem 4) imply that, the fact that $\boldsymbol{D}$ is of full rank for any admissible $\boldsymbol{\beta}$, is a necessary and sufficient condition for the model to be locally identifiable; a similar result was proved by Rothemberg, [29] who showed that the information matrix must be positive definite everywhere.

A necessary condition for local identifiability is that the size of $\boldsymbol{\gamma}$ is greater or equal to the size of $\boldsymbol{\psi}$; however, with non linear models, there are several counterexamples showing that this condition is not sufficient. Because general results on identifiability of the class of models proposed in this paper are not available, we have applied the method proposed by Forcina [16]. This provides a way of drawing a sufficiently large random sample of points from the parameter space and an efficient way of checking numerically whether, at each point, the jacobian is well away from rank deficiency. Because our model has passed such a text with a sample of 10,000 draws, we are confident that the model is indeed locally identified.

Appendix C: Log-Likelihood, BIC and AIC

		Daughters			Sons		
latent cl.	param.	$L(\hat{\boldsymbol{\psi}})$	$BIC(\hat{\boldsymbol{\psi}})$	$AIC(\hat{\boldsymbol{\psi}})$	$L(\hat{\boldsymbol{\psi}})$	$BIC(\hat{\boldsymbol{\psi}})$	$AIC(\hat{\boldsymbol{\psi}})$
2	39	-19419.04	39146.18	38916.09	-20492.98	41294.06	41063.96
3	51	-19076.66	38553.72	38255.32	-20084.53	40571.95	40271.06
4	63	-18968.75	38432.11	38063.51	-19977.58	40452.85	40081.16
5	75	-18925.14	38439.09	38000.28	-19938.27	40469.03	40026.54

Table 5. Maximized log-likelihood, BIC and AIC

Appendix D: Tables for Endowments

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_2^U	-0.3603	0.0549	0.4166	0.0761	0.9273	0.1229	1.2911	0.2410
α_3^U	-	-	-0.5558	0.0791	-0.1846	0.0761	0.1055	0.1322
α_4^U	-	-	-	-	-1.5904	0.1964	-0.1100	0.1118
α_5^U	-	-	-	-	-	-	-1.8189	0.2738
$\beta_{2,f}^U$	-0.1957	0.0453	-0.2029	0.0419	-0.1930	0.0445	-0.0782	0.0548
$\beta_{2,m}^U$	-0.1770	0.0507	-0.1656	0.0484	-0.1917	0.0528	-0.2582	0.0674
$\gamma_{2,f}^U$	-1.5466	0.1125	-0.8420	0.1486	-0.6708	0.2032	-0.7945	0.3389
$\gamma_{2,m}^U$	-1.0654	0.1120	-0.6054	0.1458	-0.4791	0.1940	-0.2199	0.2953
$\beta_{3,f}^U$	-	-	-0.1106	0.0611	-0.0569	0.0552	-0.2133	0.0652
$\beta_{3,m}^U$	-	-	-0.0921	0.0650	-0.0748	0.0631	0.0193	0.0736
$\gamma_{3,f}^U$	-	-	-1.2381	0.1470	-1.0717	0.1487	-0.5359	0.2050
$\gamma_{3,m}^U$	-	-	-0.8422	0.1408	-0.7352	0.1466	-0.5219	0.2018
$\beta_{4,f}^U$	-	-	-	-	-0.3281	0.1156	0.0373	0.0751
$\beta_{4,m}^U$	-	-	-	-	-0.0624	0.1025	-0.1384	0.0804
$\gamma_{4,f}^U$	-	-	-	-	-1.0747	0.2876	-0.9098	0.1905
$\gamma_{4,m}^U$	-	-	-	-	-0.7423	0.2429	-0.5787	0.1866
$\beta_{5,f}^U$	-	-	-	-	-	-	-0.3582	0.1422
$\beta_{5,m}^U$	-	-	-	-	-	-	-0.0392	0.1236
$\gamma_{5,f}^U$	-	-	-	-	-	-	-1.1035	0.3870
$\gamma_{5,m}^U$	-	-	-	-	-	-	-0.7362	0.3078

Table 6. Daughters U

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_2^U	-0.1297	0.0513	0.2780	0.0635	0.3230	0.0977	0.0468	0.1126
α_3^U	-	-	-0.7062	0.0817	-0.1351	0.1023	-1.0552	0.2608
α_4^U	-	-	-	-	-0.5905	0.1266	1.1924	0.2159
α_5^U	-	-	-	-	-	-	-0.6212	0.1308
$\beta_{2,f}^U$	-0.2457	0.0447	-0.2616	0.0430	-0.2112	0.0476	-0.1665	0.0498
$\beta_{2,m}^U$	-0.2796	0.0519	-0.1674	0.0504	-0.1636	0.0580	-0.1729	0.0624
$\gamma_{2,f}^U$	-1.3505	0.1109	-0.7676	0.1359	-0.4590	0.1819	-0.4728	0.1921
$\gamma_{2,m}^U$	-0.7953	0.1106	-0.5649	0.1332	-0.3389	0.1756	-0.3162	0.1863
$\beta_{3,f}^U$	-	-	-0.0725	0.0675	-0.1266	0.0721	-0.2455	0.1408
$\beta_{3,m}^U$	-	-	-0.1955	0.0716	-0.1564	0.0797	0.0220	0.1304
$\gamma_{3,f}^U$	-	-	-1.2015	0.1558	-0.7579	0.1782	-0.0028	0.3142
$\gamma_{3,m}^U$	-	-	-0.4712	0.1460	-0.4401	0.1757	-0.0910	0.3107
$\beta_{4,f}^U$	-	-	-	-	-0.0028	0.0897	0.0590	0.1393
$\beta_{4,m}^U$	-	-	-	-	-0.1212	0.0948	-0.1618	0.1243
$\gamma_{4,f}^U$	-	-	-	-	-0.9131	0.2113	-0.7393	0.2894
$\gamma_{4,m}^U$	-	-	-	-	-0.3900	0.1929	-0.3770	0.2866
$\beta_{5,f}^U$	-	-	-	-	-	-	0.0058	0.0916
$\beta_{5,m}^U$	-	-	-	-	-	-	-0.1313	0.0960
$\gamma_{5,f}^U$	-	-	-	-	-	-	-0.9297	0.2153
$\gamma_{5,m}^U$	-	-	-	-	-	-	-0.3969	0.1960

Table 7. Sons U

Appendix E: Tables for Auxiliary Responses

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{EM}	3.0125	0.0962	4.4651	0.1509	4.9269	0.1877	5.8042	0.3120
α_2^{EM}	-0.0885	0.0602	1.9942	0.1078	2.6795	0.1219	3.7201	0.1837
α_3^{EM}	-	-	-0.6688	0.0830	0.5600	0.0962	2.0592	0.1574
α_4^{EM}	-	-	-	-	-1.8420	0.2436	0.0983	0.1169
α_5^{EM}	-	-	-	-	-	-	-2.1136	0.3435
δ_3^{EM}	-1.8558	0.0723	-2.1502	0.0882	-2.0739	0.0817	-2.1675	0.0899
δ_4^{EM}	-3.4321	0.0931	-4.1088	0.1243	-4.0054	0.1178	-4.1757	0.1288
α_1^{LM}	1.8991	0.1524	4.4272	0.3106	4.9899	0.3277	6.4947	0.5699
α_2^{LM}	-0.7827	0.0845	0.9376	0.1558	2.0736	0.1946	4.0399	0.3801
α_3^{LM}	-	-	-0.9589	0.0927	-0.2271	0.1283	-1.0212	0.2091
α_4^{LM}	-	-	-	-	-1.4779	0.1713	-0.3871	0.1273
α_5^{LM}	-	-	-	-	-	-	1.6843	0.2322
ρ_2^{LM}	0.5219	0.1181	0.3541	0.1370	0.2088	0.1412	0.2046	0.1428
ρ_3^{LM}	1.1376	0.1391	0.5423	0.1673	0.5987	0.1666	0.4242	0.1824
ρ_4^{LM}	2.6546	0.1541	1.4526	0.2002	1.2820	0.2165	-0.9197	0.2404
δ_3^{LM}	-2.1085	0.0862	-2.0929	0.0803	-2.1555	0.0915	-2.2235	0.0934
δ_4^{LM}	-4.1554	0.1143	-5.0585	0.2167	-4.7837	0.1636	5.2271	0.2465

Table 8. Daughters Math

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{EM}	3.2306	0.0966	4.6058	0.1444	5.3432	0.1895	5.8617	0.3059
α_2^{EM}	0.1241	0.0558	1.8360	0.1006	2.6035	0.1348	2.2261	0.1518
α_3^{EM}	-	-	-0.8502	0.0939	1.2390	0.1219	5.0838	0.4649
α_4^{EM}	-	-	-	-	-1.4455	0.1633	1.1886	0.1243
α_5^{EM}	-	-	-	-	-	-	-1.4929	0.1725
δ_3^{EM}	-1.6050	0.0617	-1.9688	0.0828	-1.9951	0.0831	-2.0646	0.0866
δ_4^{EM}	-3.3342	0.0896	-4.0718	0.1216	-4.1923	0.1277	-4.7819	0.2590
α_1^{LM}	2.6603	0.1638	4.6509	0.2641	5.9619	0.3868	5.6605	0.3868
α_2^{LM}	-0.5548	0.0812	1.1428	0.1555	3.0557	0.2534	3.2672	0.2811
α_3^{LM}	-	-	-0.8976	0.0944	0.1138	0.1735	1.6861	0.4079
α_4^{LM}	-	-	-	-	-0.9049	0.1073	0.0581	0.1713
α_5^{LM}	-	-	-	-	-	-	-0.9253	0.1097
ρ_2^{LM}	0.8671	0.1140	0.4564	0.1391	0.4143	0.1572	0.4308	0.1572
ρ_3^{LM}	1.2216	0.1316	0.5748	0.1626	0.5380	0.1860	0.6355	0.1925
ρ_4^{LM}	2.3412	0.1488	0.9904	0.1991	0.4225	0.2735	0.9508	0.3229
δ_3^{LM}	-1.7609	0.0732	-1.7080	0.0718	-1.9336	0.0930	-1.9100	0.0898
δ_4^{LM}	-4.1827	0.1234	-4.4783	0.1571	-4.8104	0.1766	-4.8957	0.2047

Table 9. Sons Math

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{ER}	3.7837	0.1161	5.1617	0.1745	5.8778	0.2112	7.2252	0.4740
α_2^{ER}	0.1260	0.0603	2.7629	0.1430	3.7055	0.1558	4.5904	0.2125
α_3^{ER}	-	-	-0.4708	0.0816	0.9414	0.1122	3.2890	0.2108
α_4^{ER}	-	-	-	-	-2.0537	0.2983	0.4091	0.1388
α_5^{ER}	-	-	-	-	-	-	-2.9811	0.7107
δ_3^{ER}	-2.0859	0.0838	-2.4660	0.1195	-2.4651	0.1059	-2.7146	0.1474
δ_4^{ER}	-4.0245	0.1126	-4.6476	0.1510	-4.7652	0.1475	-5.0015	0.1800
α_1^{LR}	1.1343	0.1753	2.3968	0.2360	3.6066	0.3043	4.3060	0.3744
α_2^{LR}	-1.2773	0.1019	0.5244	0.1846	1.7598	0.2408	2.6201	0.2847
α_3^{LR}	-	-	-1.4654	0.1107	-0.3637	0.1632	-1.1907	0.2647
α_4^{LR}	-	-	-	-	-2.4432	0.2873	-0.6035	0.1608
α_5^{LR}	-	-	-	-	-	-	3.0243	0.5374
ρ_2^{LR}	1.2380	0.1308	0.8613	0.1539	0.5141	0.1741	0.3800	0.1789
ρ_3^{LR}	2.0867	0.1602	1.6692	0.1847	1.2482	0.2041	1.1053	0.2220
ρ_4^{LR}	3.6859	0.1784	2.8772	0.2095	2.2226	0.2346	-1.9933	0.2508
δ_3^{LR}	-2.5767	0.1028	-2.5349	0.0959	-2.6384	0.1109	-2.6245	0.1105
δ_4^{LR}	-5.0011	0.1334	-5.1866	0.1428	-5.3922	0.1603	5.3774	0.1564

TABLE 10. Daughters Reading

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{ER}	3.2408	0.1024	4.4681	0.1426	4.9987	0.1727	5.1789	0.1990
α_2^{ER}	-0.1731	0.0565	1.6788	0.1021	2.5469	0.1383	2.2806	0.1479
α_3^{ER}	-	-	-1.3916	0.1144	0.9202	0.1239	3.8558	0.3396
α_4^{ER}	-	-	-	-	-2.2182	0.2358	0.8401	0.1243
α_5^{ER}	-	-	-	-	-	-	-2.2538	0.2460
δ_3^{ER}	-1.8416	0.0732	-2.1758	0.0892	-2.1810	0.0884	-2.2401	0.0920
δ_4^{ER}	-3.4048	0.0968	-4.0315	0.1224	-4.0869	0.1239	-4.2927	0.1569
α_1^{LR}	1.1503	0.1429	2.7816	0.2238	3.0218	0.2460	2.9084	0.2481
α_2^{LR}	-0.8300	0.0785	1.1892	0.1756	1.9570	0.2116	2.0529	0.2145
α_3^{LR}	-	-	-1.4469	0.1109	0.5205	0.1785	1.1566	0.3152
α_4^{LR}	-	-	-	-	-1.8929	0.1772	0.4585	0.1747
α_5^{LR}	-	-	-	-	-	-	-1.9478	0.1877
ρ_2^{LR}	1.2366	0.1123	0.2815	0.1631	0.1723	0.1719	0.2157	0.1695
ρ_3^{LR}	1.9857	0.1414	1.0068	0.1798	0.9196	0.1852	1.0390	0.1894
ρ_4^{LR}	3.7814	0.1601	2.5650	0.2031	2.4028	0.2135	2.5973	0.2262
δ_3^{LR}	-2.0306	0.0768	-2.2206	0.0884	-2.1932	0.0862	-2.1974	0.0862
δ_4^{LR}	-4.5549	0.1167	-4.8273	0.1297	-4.7927	0.1282	-4.8239	0.1302

TABLE 11. Sons Reading

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{NP}	2.1927	0.0701	2.3358	0.0856	2.4576	0.1057	2.6060	0.1424
α_2^{NP}	1.1808	0.0639	1.9152	0.0821	2.1045	0.0874	2.2485	0.1082
α_3^{NP}	-	-	0.9918	0.0731	1.6103	0.0904	1.8993	0.1064
α_4^{NP}	-	-	-	-	0.4267	0.1189	1.5234	0.1020
α_5^{NP}	-	-	-	-	-	-	0.2182	0.1512
δ_3^{NP}	-1.4035	0.0502	-1.4217	0.0510	-1.4592	0.0533	-1.4642	0.0538
δ_4^{NP}	-2.6472	0.0624	-2.6749	0.0634	-2.7238	0.0657	-2.7308	0.0661
α_1^{NS}	2.1218	0.0696	2.2920	0.0859	2.3387	0.1042	2.3371	0.1376
α_2^{NS}	0.5990	0.0595	1.7404	0.0818	2.0691	0.0891	2.2418	0.1076
α_3^{NS}	-	-	0.3134	0.0709	1.0483	0.0868	1.8239	0.1123
α_4^{NS}	-	-	-	-	-0.2783	0.1284	0.8161	0.0990
α_5^{NS}	-	-	-	-	-	-	-0.4473	0.1628
δ_3^{NS}	-1.2287	0.0463	-1.2614	0.0482	-1.2876	0.0497	-1.2891	0.0498
δ_4^{NS}	-2.4909	0.0616	-2.5325	0.0634	-2.5638	0.0645	-2.5660	0.0646

TABLE 12. Daughters Non Cognitive

	2LC		3LC		4LC		5LC	
	coeff	se	coeff	se	coeff	se	coeff	se
α_1^{NP}	1.4226	0.0612	1.5678	0.0725	1.6752	0.0852	1.6849	0.0854
α_2^{NP}	0.0399	0.0544	0.7484	0.0678	1.1288	0.0897	1.1952	0.1069
α_3^{NP}	-	-	-0.3062	0.0776	0.3714	0.0900	0.8594	0.1953
α_4^{NP}	-	-	-	-	-0.4150	0.0956	0.3689	0.0896
α_5^{NP}	-	-	-	-	-	-	-0.4312	0.0979
δ_3^{NP}	-1.1372	0.0405	-1.1498	0.0410	-1.1584	0.0414	-1.1597	0.0415
δ_4^{NP}	-2.2937	0.0574	-2.3085	0.0577	-2.3264	0.0583	-2.3294	0.0585
α_1^{NS}	1.5214	0.0614	1.6289	0.0723	1.6688	0.0843	1.6610	0.0839
α_2^{NS}	0.4411	0.0545	1.2138	0.0714	1.3662	0.0884	1.2654	0.1034
α_3^{NS}	-	-	-0.0352	0.0758	1.0307	0.0937	1.7457	0.2022
α_4^{NS}	-	-	-	-	-0.2707	0.0973	1.0024	0.0938
α_5^{NS}	-	-	-	-	-	-	-0.2889	0.0997
δ_3^{NS}	-1.2061	0.0410	-1.2507	0.0431	-1.2593	0.0436	-1.2626	0.0438
δ_4^{NS}	-2.3938	0.0565	-2.4539	0.0584	-2.4623	0.0587	-2.4698	0.0591

TABLE 13. Sons Non Cognitive