# EDUCATION TRANSMISSION: A TWO-LATENT-CLASS MODEL

ANTONIO FORCINA AND SALVATORE MODICA

Abstract. Intergenerational education transmission is evaluated by introducing and estimating a finite-mixture discrete model whose system of equations includes simultaneously family and child unobservables. The data are from the English National Child Development Survey.

Our main broad finding is that the flow of education is not automatic from parents to children; on the other hand it goes through, to a non-negligible extent, if family pushes. In other words, there seems to be value in giving value to education in the family.

*JEL Classification Numbers* I21, C35

*Keywords* Education Transmission, Finite Mixtures Models, Causal Effects.

## 1. Introduction

Dealing with intergenerational education transmission inevitably involves coming to grips with parents' and children's unobservables. The existing literature, which we describe in the next section with the aid of some formal structure, takes two main avenues to deal with the problem. One, starting with Behrman and Rosenzweig (2002), is to control for parents' endowments only, substituting out the child's endowment which ultimately depends on parents' characteristics besides chance. A well know method of doing this is to use twin parents, but IV methods are also relevant as we report shortly. The other route, see the influential paper by Cameron and Heckman (1998), is to ignore parents' unobservables and focus directly on the determinants of the child's education including her own endowment, in which case the need to go back to parents' unobservable disappears since the latter only act through the child's characteristics included in the specification. Each approach addresses a specific question: Behrman and Rosenzweig deal with the different roles played by mothers and fathers; Cameron and Heckman look at the strength of family effects through the different stages of children's learning. A third line of research, represented by the somewhat dated paper by Behrman and Taubman (1989), investigates the

*Date*: October 10, 2011.

relative weight of the different channels through which education flows from a generation to the next, concentrating on 'genetic' versus 'induced' transmission.

The present paper introduces a new approach by estimating a finite-mixture discrete model whose system of equations includes simultaneously family and child unobservables. By so doing the different causal dependencies can be appraised in an integrated framework, and the issues so far mentioned can be studied on the same footing in the context of a single population. The results we present are based on the English National Child Development Survey, which concerns individuals born in the UK in March 1958 and surveyed, together with their parents, through all their educational career.

The broad message we get from the data we have analyzed is the following. First of all, family unobservables do play a significant role, but not overwhelming: education transmission through parents' endowment is of roughly the same size as the educational increments due to variations of parents' education or rearing efforts. Thus room for policy is significant. With policy in view, in our interpretation the main lesson we draw from this study is that, within the family, there is value in giving value to education. What family pushes goes through, to a non-negligible extent. In the social environment we have studied, which dates back to the seventies, this clearly, and luckily, emerges in an education transmission mostly confined to the male community: the parent whose education has real impact is the father, and his influence affects only his sons. We say luckily, because it would surely be harder to extract such a lesson from present-day data about a society where gender role has changed so substantially. The other good news is that parents' education is not all that matters: interest has a comparable impact, which reinforces the value-of-education-value hypothesis. One thing we have to report on the negative side: for the transition to higher education the weight of family background appears to be substantial.

The rest of the exposition is as follows. The next section puts the paper in context and relates it to the relevant literature. Section 3 describes the dataset we study. Section 4 describes the model we estimate, and section 5 presents estimation results. In section 6 we compute causal effect by applying the methodology introduced by Pearl, see Pearl (2000). Section 7 contains some final remarks. An Appendix contains details of the estimation procedure and a table with the complete set of estimates.

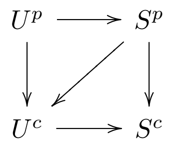
## 2. Context, Literature and Setup

To discuss the basic issues concerning intergenerational education transmission it may be useful to start from a simple model based on two structural equations where $S^p, U^p$ and $S^c, U^c$ denote schooling and unobservable endowments respectively for the parents and the child and $\epsilon^p$, $\epsilon^c$ are exogenous errors

$$S^c = f(S^p, U^c, \epsilon^c), \tag{1}$$

$$U^c = g(S^p, U^p, \epsilon^p). \tag{2}$$

This model says that a child's education depends on her own endowment and her parents' education, and in turn, the child's endowment depends on her parents' schooling and endowment. This model is equivalent to the following causal Directed Acyclic Graph (DAG) where each node corresponds to a structural equation where the output variable is determined by its parents (see for example Pearl, 2000):




The DAG makes clear where, within this simplified context, the estimation difficulty lies: the observed correlation between $S^p$ and $S^c$ is partly due to the effect of endowment on schooling within each generation (horizontal arrows) combined with the transmission effect from $U^p$ to $U^c$ (left vertical arrow). Thus, the stronger the endowment transmission effect, the weaker the scope of education policy.

We briefly review some models and results in the literature to which the present paper aims to contribute. To estimate the effect of policy education, one way to proceed is to substitute from equation (2) into (1) to get the reduced form equation

$$S^c = f(S^p, U^p, \epsilon)\,. \tag{3}$$

Estimation of the causal effect of $S^p$ on $S^c$ by this equation, once properly specified, requires controlling for the unobservable parents' endowment. Though this is not an easy task, at least three interesting attempts in this direction have been proposed. Behrman and Rosenzweig (2002) take differences between subjects with twin mothers and try to account for assortative mating in order to control for differences between education of fathers; Plug (2004) uses data on adoptees under the assumption that there should be no endowment transmission, although, as noted by Holmlund et al. (2011), association may be induced by selective placement of adoptees; finally, Black et al. (2005) analyze a dataset where differences in parent's education was exogenously induced by reforms in municipal schooling laws so that the can be used as an instrument. These papers concentrate mainly on gender effects: which parent's education matters more for the child's? The difficulties faced by these methods are discussed in detail by Holmlund et al. (2011) who apply the three methods to a single data set and show that they produce conflicting results.

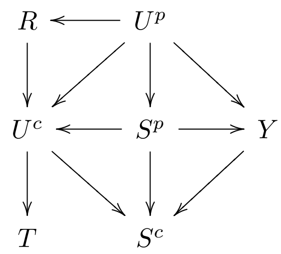
Cameron and Heckman (1998) address the issue of how the influence of family background on the probability of transition from one grade of education to the next evolves. They also discuss a discrete choice model where the effect of family background on education of the child is adjusted for individual heterogeneity. Though this model resembles (1), the heterogeneity is assumed independent from the observed covariates, so it could be interpreted as the component of $U^c$ which is not determined by family background.

By directly controlling for both $U^c$ and $U^p$ we can also discuss the basic nature-nurture question mentioned above. The title of the authoritative paper by Behrman and Taubman (1989), "Is Schooling Mostly in the Genes", reveals quite explicitly its content. To be precise they do not address the education transmission problem, since their family background includes only father's occupation and family size; and their methods are also quite different from ours, for they use a two-stage procedure whereby effect of environment is estimated first using OLS, and in the second stage genoytpe relations are estimated by maximum likelihood using observations on relatives.

The present paper formulates and estimates a system of structural equations describing the causal relations between the variables affecting education. The model explicitly includes both latent endowments $U^c$ and $U^p$, and it is thus general enough to discuss all the issues we have mentioned so far. We make certain parametric assumptions in order for the resulting mixture model to be identifiable; and then use Pearl's theory of causation to compute the causal effects of interest in terms of transition probabilities from one educational level. to the next. The type of finite mixture models used in this paper differ from those used for instance by Cameron and

Heckman (1998) mainly because we estimate the full set of structural equations determining all endogenous variables including the two latent endowments. This allows us to extract information on the unobservables from several related observable variables.

A slightly simplified version of the system we deal with is represented in the DAG below where $R$, $Y$ denote, respectively, interest and income of parents and $T = (NC, C)$ are test results of non cognitive and cognitive abilities of the child:




The actual model which we fit involves 14 different structural equations including the two unobservables $U^p$ and $U^c$ and is presented and discussed in section 4, where we also specify the parametric assumptions used in the estimation.

A word on interpretation of unobservables is in order. The variable $U^p$, family endowment, is essentially identified by the variables it affects, so it is meant to capture the family environment in which children grow up. It is in principle a cross classification of various characteristics of the family, but in practice it turns out to be naturally ordered in a scale of 'quality'.

The child's unobservable $U^c$ is identified mainly through cognitive and non-cognitive test scores, so it is not to be interpreted as strictly reflecting an individual intrinsic endowment; it is rather a mixture of this and other unobservables like motivation and acquired knowledge useful for schooling advancement.

## 3. Data

We use data from the National Child Development Survey (NCDS). The dataset is produced by a UK cohort study targeting 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 achievement, various tests results and family background was collected. A complete description of the data is available at `esds.ac.uk/longitudinal/access/ncds`.

Some variables are inherently discrete (notably schooling level) while others would be more naturally described as continuous, like income and tests scores. Because the finite mixture model approach used in this paper can be applied only when all endogenous variables are categorical, continuous variables were turned into discrete. Though, clearly, a continuous variable contains more information relative to a discrete approximation, there are two reason why a model involving categorical variables may involve less parametric restrictions than one based on the original continuous measurements. First, a continuous variables used as explanatory in a regression model implies linearity unless additional polynomial terms are introduced; instead, when used a dummy variable corresponding to discrete categories, it can capture patterns of non linearity in a non parametric way. Also, models involving a continuous variable as response are usually based on the rather restrictive assumption of normality while, when used as categorical, the discrete distribution is assumed to be multinomial, that is completely unrestricted, at least in the first stage.

The original sample contains 18560 observations, but more than 80% have at least a missing entry. Incompleteness is scattered across the great many variables included in the survey; however, the marginal distributions of the summary statistics for the most relevant variables in the complete-case sub-sample do not differ significantly from the same distributions in the whole sample. Thus, excluding the observation with missing data should not introduce substantial selection bias. On the other hand the complete observations still amount to almost 3000 subjects. We have thus decided to analyze the subsample of complete data, which consists precisely of 2801 subjects, 1471 males (sons) and 1330 females (daughters).

Our main dependent variable is the amount of education achieved by each individual, which takes four levels: no qualification, O-level, A-level and higher education. In our sample, males have a 50-30-10-10 percentage distribution over the levels, females more like 40-30-5-15.

Children are tested at the age of 7 and 11 for mathematics, reading and non-cognitive skills, and again at 16 for math and reading, and we use the test scores for identification of the unobservable endowment. More specifically, after taking principal components (which in all cases explain no less than 90% of the total variance) for math and reading we combine scores at 7 and 11 into an ordered variables 'early math' and 'early reading'. Math and reading scores at 16 are coded in two further ternary variables 'late math' and 'late 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. For these variables we use quantiles as cut-points.

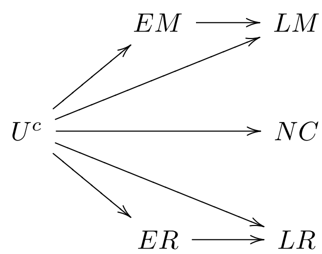
Parents' schooling is defined as the age at which they left school (12 to 21 years); we extract three levels corresponding to significant educational steps: leaving up to 14 years of age; after 14 but not later than 16; after 16. For mothers this results in a 45-45-10 percentage distribution in our sample; fathers are distributed roughly 60-30-10. As usual there are many missing data on family income; to alleviate the problem, since few mothers in the dataset have an income, we neglect mother's income (thus avoiding to drop data with missing mother's income) and concentrate on fathers'. We group their income in three categories, with cut-points 150 and 300, with resulting distribution of roughly 15-70-15.

The NCDS contains also information on parents' interest in their children's education, as reported by teachers; this turns out to be an important variable, which may be related to rearing efforts or concern, and in general to the value that family gives to the child's education. Parents' interest is originally classified into as many as 5 categories; we extract three binary parents' interest variables, one for each surveyed age (7, 11 and 16); the low-interest percentage is roughly 60%.

## 4. Definition and estimation of the Model

In this section we describe in detail the system of structural equations that constitute our model and the likelihood function used for estimation. A simplified version of the structural equations was displayed in the DAG on pag. 5; note however that a causal DAG is mainly a way of coding the assumptions of conditional independence (exclusion-restrictions). Below we complete the model specifications by defining, for each equation, the link function and the actual regression model. Models involving latent variables are usually not identifiable unless several assumptions of conditional independence hold, see for instance Allman et al (2009). Identification of our model is achieved, essentially, by the conditional independencies coded in the DAG and by the parametric restrictions in the regression models to be described below; we used the result of Rothemberg (1971) for local identifiability by checking that the expected information matrix is of full rank and well conditioned.

Parents' education is a two-element vector $S^p = (S^m, S^f)$, superscripts $m, f$ for mother and father; parents' interest is reported at different ages of the child: $(I^7, I^{11}, I^{16})$; cognitive tests are summarized by four variables, two for math and two for reading, the early ones being $(EM, ER)$, and the late ones being $(LM, LR)$. Finally, we denote by $Y, NC$ Income and Non-cognitive tests, respectively. To visualize the markovian dependence allowed in the system we are going to spell out, recall that in the DAG on page 5 there is only one arrow from $U^c$ to a variable $T$ encompassing all tests. The actual dependence of tests on $U^c$ is described in the following DAG:




The above DAG together with the one on page 5 describe the dependence and conditional independence relations present in the model.

The full system of structural equations composing the model is the following:

$$
\begin{aligned}
&U^p = f_{U^p}(\epsilon^{U^p})\\
&S^m = f_{S^m}(U^p, e^{S^m}), \quad S^f = f_{S^f}(U^p, S^m, e^{S^f})\\
&I^i = f_{I^i}(U^p, e^{I^i}), \quad i = 7, 11, 16\\
&Y = f_Y(U^p, S^m, S^f, e^Y)\\
&U^c = f_{U^c}(U^p, I^7, I^{11}, I^{16}, S^m, S^f, \epsilon^{U^c})\\
&ET = f_{ET}(U^c, \epsilon^{ET}), \quad LT = f_{LT}(U^c, ET, \epsilon^{LT}), \quad T = M, R\\
&NC = f_{NC}(U^c, \epsilon^{NC})\\
&S^c = f_{S^c}(S^m, S^f, Y, U^c, \epsilon^{S^c})
\end{aligned}
$$

where $f_v$ denote a suitable link function specific to the $v$ variable.

The system allows for assortative mating; however, because this relation is not of interest in itself, we simply allow the education of the father to depend on the education of the mother. A more structural role is played by the dependence of late tests results on early ones, since this should provide some insights into the dynamics of child endowment. On the other hand, as the reader may have noticed, rearing concern $R$ does not depend on parents' education $S^p$; the reason is that the data show no marked dependence of $R$ on $S^c$, suggesting that $R$ depends instead on family environment and values (hence on $U^p$). Finally, due to the limited information on individual parents, we try to control for an overall measure of parents' endowment $U^p$ which is assumed to capture the overall, qualitative endowment of the parents as residual heterogeneity.

In the estimated model all variables are discrete; we recall the number of categories for each variable starting from the observable responses: for parents, education $S^m, S^f$ and income $Y$ have 3 levels, interest $I^i,\ i = 7, 11, 16$ has 2; for children, cognitive tests $EM, ER, LM, LR$ have 3 levels, non-cognitive response $NC$ has 2, schooling achievement $S^c$ has 4. As to the unobservables, we have used 3 categories both for $U^p$ and $U^c$.[1] For convenience, the initial level of each variable is always coded as 0.

In the first stage the joint distribution of the 14 variables involved in the above system is assumed to be multinomial with a vector of probabilities $\boldsymbol{\pi}$; though this vector is huge, it can be factorized according to the DAG and its components can be given a parametric, linear form using logit link functions suitable for each variable (see e.g. Agresti (2002) ch.7). Once a non linear regression model is specified for each variable in the system, one can maximize the full likelihood by maximizing the likelihood of each response conditionally on its direct causes which appear in the corresponding regression model.

As usual, for $p \in (0, 1)$ the logit function is defined as

$$\text{Logit}(p) = \ln(p/(1-p)) = G^{-1}(p)$$

where $G$ is the standard logistic cdf. Because all observable variables in the system are naturally ordered, we use cumulative (or 'global') logits defined as $\text{Logit}\, P(Z > j)$, for any ordered variable $Z$. The levels of unobservable variables, instead, are assumed to correspond to unordered qualitative types, so we use logits computed with reference to the initial category: $\text{Logit}[P(U = j \mid U = 0)]$.

In order to model the dependence of logits on a vector of discrete covariates, we assume that the effect of covariates is additive on the appropriate logit scale. For each logit there will be an

[1]We have tried several different options and the choice was based on the Bayes Information Criterion.

intercept parameter corresponding to the value of the logit when all covariates are set at their initial level and, for each covariate with $k$ levels, there will be $k-1$ parameters measuring how the logit changes when the covariate is set to $j$ instead of $j-1$.

Given these rules, it is redundant to display all equations and we may concentrate on a few clarifying instances. For example, for $S^m$, which depends only on $U^p$, we have 2 baseline logits for $S^m$ ($S^m > 0$ and $S^m > 1$) and two incremental effects of $U^p$ changing from 0 to 1 and from 1 to 2. To specify the equation it is convenient to use the indicator function $\mathbf{1}_{\{x\geq y\}}$ equal to 1 if $x \geq y$ and 0 otherwise. Then the equation has the following form (recall that $j = 0, 1$):

$$\text{Logit}(S^m > j \mid U^p = u) = \alpha_0(S^m) + j\alpha_1(S^m) + \mathbf{1}_{\{u\geq 1\}}\beta_1(U^p, S^m) + \mathbf{1}_{\{u\geq 2\}}\beta_2(U^p, S^m).$$

The next example is the equation for $S^c$ (where, for compactness, we write, for instance $\beta_1(S^m)$ in place of $\beta_1(S^m; S^c)$) has the form

$$\begin{aligned}\text{Logit}(S^c > j \mid S^m = s, S^f = t, Y = y, U^c = u) &= \alpha_0 + \mathbf{1}_{\{j\geq 1\}}\alpha_1 + \mathbf{1}_{\{j\geq 2\}}\alpha_2 \\ &+ \mathbf{1}_{\{s\geq 1\}}\beta_1(S^m) + \mathbf{1}_{\{s\geq 2\}}\beta_2(S^m) + \mathbf{1}_{\{t\geq 1\}}\beta_1(S^f) + \mathbf{1}_{\{t\geq 2\}}\beta_2(S^f) \\ &+ \mathbf{1}_{\{y\geq 1\}}\beta_1(Y) + \mathbf{1}_{\{y\geq 2\}}\beta_2(Y) + \mathbf{1}_{\{u\geq 1\}}\beta_1(U^c) + \mathbf{1}_{\{u\geq 2\}}\beta_2(U^c)\,.\end{aligned}$$

In words, it contains 11 parameters, 3 for the $S^c$ logit intercepts and 2 for the incremental effects of each of the 4 explanatory variable passing from 0 to 1 and from 1 to 2.

As a last example, consider the equation for $U^c$, with a notation similar to the one used above, it has 2 logits for the response ($u = 0, 1$), and the incremental effects of the explanatory variables:

$$\begin{aligned}&\text{Logit}\,P(U^c > u \mid U^p = v, I^7 = i, I^{11} = j, I^{16} = k, S^m = s, S^f = t) = \\ &\alpha_0 + \mathbf{1}_{\{u\geq 1\}}\alpha_1 + \mathbf{1}_{\{v\geq 1\}}\beta_1(U^p) + \mathbf{1}_{\{v\geq 2\}}\beta_2(U^p) + i\beta(I^7) + j\beta(I^{11}) + k\beta(I^{16}) \\ &+ \mathbf{1}_{\{s\geq 1\}}\beta_1(S^m) + \mathbf{1}_{\{s\geq 2\}}\beta_2(S^m) + \mathbf{1}_{\{t\geq 1\}}\beta_1(S^f) + \mathbf{1}_{\{t\geq 2\}}\beta_2(S^f)\,,\end{aligned}$$

where recall that $v, i, j, k = 0, 1$.

## 5. Main Estimation Results

To test for possibly significant differences due to role effects we estimate the system of equations on page 8 separately for sons and daughters. The system contains 74 parameters. Estimates and standard errors for the equation of main interest, that of $S^c$, are reported in Table 1 and commented below; the complete set of estimated coefficients is in Table 3 in the Appendix.

More extended discussion of estimation results is postponed to next section, where an analysis of causal effects is carried out in terms of the relevant transition probabilities. Since the child's endowment plays a major role it will be interesting to compare the relative strength of the transmission channels through which it is raised, increasing the probabilities of transition to higher levels of education for the children. We then proceed to evaluate the four channels identified by the model through which education flows from family to children: endowment, interest, education and income.

Coming back to the $S^c$ equation in Table 1, the first noteworthy outcome is that income does not seem to be of much relevance. On the other hand, as expected the role of the child's endowment is strong for both daughters and sons, though the relative strength of $\beta_1$ and $\beta_2$ is reversed in the two cases. The other interesting feature of the $S^c$ equation is that the impact of the three-level parents' education contains a pronounced cross-gender effect: only fathers' education seems to count, but only for their sons'. It is stronger when the parent's education becomes high. Mothers' education has no significant impact on children's, confirming the results of Behrman and Rosenzweig (2002). Given the sociological context of English families in the seventies where these data are collected, with the prevalence of males in the working environment, the effect of fathers on sons appears to reflect a strong role effect, whereby the educated father strongly pushes his sons' schooling achievement independently of the latter's endowment, the more so if the father is highly educated.

We remark that in the context of a multi-equation system as the one we are estimating a single equation does not contain all information on the variables involved. In particular, the reader may have noticed that parents' interest is not present in the $S^c$ equation above. It will be clear in the next section that the causal effect of the interest variables on $S^c$ is definitely present and

TABLE 1. Coefficients of the $S^c$ Equation

| | Daughters | | Sons | |
|---|---|---|---|---|
| | coeff | se | coeff | se |
| $\alpha_0$ | 2.3737 | 0.2507 | 2.5194 | 0.2681 |
| $\alpha_1$ | -2.3476 | 0.1132 | -2.7011 | 0.1567 |
| $\alpha_2$ | -0.6647 | 0.0633 | -0.9800 | 0.0829 |
| $\beta_1(S^m)$ | 0.1585 | 0.1491 | -0.0600 | 0.1423 |
| $\beta_2(S^m)$ | 0.3622 | 0.2009 | -0.1118 | 0.2003 |
| $\beta_1(S^f)$ | 0.0818 | 0.1539 | 0.3571 | 0.1465 |
| $\beta_2(S^f)$ | 0.1118 | 0.2050 | 0.6452 | 0.2132 |
| $\beta_1(Y)$ | 0.1162 | 0.1926 | 0.0254 | 0.1867 |
| $\beta_2(Y)$ | 0.2433 | 0.1699 | 0.2533 | 0.1762 |
| $\beta_1(U^c)$ | 2.8394 | 0.2107 | 1.7948 | 0.2175 |
| $\beta_2(U^c)$ | 1.8843 | 0.1831 | 3.0267 | 0.2144 |

important, and can be assessed with the appropriate analytical tool in the DAG representing the system.

## 6. Causal Effects on Transition Probabilities

When measuring causal effects within the system, which is represented by the DAGs on pages 5 and 8, it is convenient to formulate questions and compute appropriate answers within a formal language of causality which we briefly summarize below (see for example Pearl, 2000, Chapter 3). A Markovian causal model (of which the one in section 4 is an example) is specified by a system of structural equations

$$Z_i = f_i(pa_i, \epsilon_i) \quad i = 1, \ldots, n\,,$$

where $pa_i$ are the direct causes ('parents') of $Z_i$ and $(\epsilon_1, \ldots, \epsilon_n)$ are background exogenous variables which are assumed to be independent. The specification above leads to the recursive decomposition of the joint distribution $P$ given by

$$P(z_1, \ldots, z_n) = \prod_i P(z_i \mid pa_i)\,.$$

Computation of the causal effect of a subset of variables $X = (Z_i)_{i\in I}$ on $Y = (Z_i)_{i\in J}$, with $J$ disjoint from $I$, requires, essentially, three steps: first one determines the intervention distribution when $X$ is set to a value $x$; this amounts to delete the factors of index $i \in I$ in the joint distribution, the result being the distribution under the so called *do* operator, taking the value

$$P(z_1, \dots, z_n \mid do(x)) = \prod\nolimits_{i \notin I} P(z_i \mid pa_i)$$

for $(z_1, \dots, z_n)$ consistent with $x$ and zero otherwise. Next we need to marginalize the distribution with respect to the variables with $i \notin I \cup J$,

$$P(y \mid do(x)) = \sum\nolimits_{i \notin I \cup J} P(z_1, \dots, z_n \mid do(x))\,.$$

Finally we need to choose how to compare distributions of $Y$ for different values of $x$. The two most obvious choices are differences or ratios of the relevant probabilities; we have chosen the latter.

6.1. **Direct and indirect effects.** In a complex system causal effects may arise through several different pathways, as in the DAG displayed in Section 2 which is the present object of interest. There $S^p$ affects $S^c$ directly, or by affecting $U^c$ or $Y$ which in turn affects $S^c$. The effect of $U^p$ is exerted through many channels, but we want to compute its effect on $S^c$ while observed family background is held fixed, to capture the effect of family unobserved environment. In the literature different definitions of direct effects have been considered; the one used here, encompassing all cases we wish to discuss, is due to Pearl, who calls it 'natural direct effect', and is defined as follows (see Pearl, 2000, Definition 4.5.1). Suppose we are interested in the causal effect of a set of variables $X$ on $Y$ exerted through all paths except those going through a set of mediating variables $M = (Z_i)_{i \in K}$, with $K$ disjoint from $I, J$. Then one first computes the distribution when we set $X = x$ and $M = m$

$$P(y \mid do(x), do(m)) = \sum\nolimits_{i \notin I \cup J \cup K} P(z_1, \dots, z_n \mid do(x), do(m))\,;$$

the effect of $M$ is then averaged out, with weights provided by the distribution of $M$ when $X$ is set to its baseline value by intervention, that is $P(m \mid do(x))$.

6.2. **Estimated Causal effects.** We now apply the method outlined above to measure several causal effects on children's schooling based on the assumed system of structural equations and the estimated probability distributions. The system of structural equations, or the resulting DAG, is required to determine the collection of variables that can mediate indirect causal pathways; in addition, the conditional distribution associated with each structural equation is derived to be used in the actual computation of the intervention distribution of interest.

We compare pairs of intervention distributions for $S^c$ when each explanatory variable of interest is shifted from one category to the next. The results, which we next discuss, are reported in Table 2. The comparisons are expressed as ratios of survival probabilities, so for example the upper-left value of 1.5633 says that the probability that a girl reaches education level at least 1 when $U^p = 1$ is 1.5633 times larger than when $U^p = 0$. The three general conclusions of our analysis are the following:

Table 2. Causal effects on $S^c$

| | Daughters | | | Sons | | |
|---|---|---|---|---|---|---|
| | $S^c > 0$ | $S^c > 1$ | $S^c > 2$ | $S^c > 0$ | $S^c > 1$ | $S^c > 2$ |
| *of family endowment* $U^p$, *given measured family background* $(R, S^p, Y)$ | | | | | | |
| from 0 to 1 | 1.5633 | 2.5020 | 2.8852 | 1.1778 | 1.4881 | 1.5413 |
| from 1 to 2 | 1.0791 | 1.2482 | 1.2897 | 1.0121 | 1.0283 | 1.0305 |
| from lowest to highest | 1.6869 | 3.1230 | 3.7211 | 1.1920 | 1.5302 | 1.5882 |
| *of Interest* $R = (I^7, I^{11}, I^{16})$ | | | | | | |
| from lowest to highest | 1.3640 | 1.7243 | 1.8120 | 1.7379 | 2.6502 | 2.7337 |
| *Separately for the three components* | | | | | | |
| $I^7$ from 0 to 1 | 1.0398 | 1.0682 | 1.0733 | 1.0780 | 1.1254 | 1.1263 |
| $I^{11}$ from 0 to 1 | 1.1761 | 1.3242 | 1.3554 | 1.2824 | 1.5079 | 1.5182 |
| $I^{16}$ from 0 to 1 | 1.1023 | 1.1794 | 1.1943 | 1.2373 | 1.4142 | 1.4215 |
| *of parents' schooling* $S^p = (S^m, S^f)$ | | | | | | |
| from lowest to highest | 1.2105 | 1.5477 | 1.7016 | 1.7120 | 2.7250 | 3.3245 |
| *Separately, of mother's schooling* $S^m$ | | | | | | |
| from 0 to 1 | 0.9383 | 0.9110 | 0.9144 | 0.9893 | 0.9799 | 0.9699 |
| from 1 to 2 | 1.2005 | 1.4555 | 1.5480 | 1.1840 | 1.2881 | 1.2741 |
| from lowest to highest | 1.1264 | 1.3260 | 1.4155 | 1.1713 | 1.2622 | 1.2357 |
| *and of father's schooling* $S^f$ | | | | | | |
| from 0 to 1 | 1.0593 | 1.1162 | 1.1330 | 1.1746 | 1.3373 | 1.4384 |
| from 1 to 2 | 1.0259 | 1.0605 | 1.0748 | 1.3195 | 1.6863 | 1.9446 |
| from lowest to highest | 1.0867 | 1.1837 | 1.2177 | 1.5499 | 2.2551 | 2.7971 |
| *of income* $Y$ | | | | | | |
| from 0 to 1 | 1.0275 | 1.0628 | 1.0773 | 1.0071 | 1.0129 | 1.0178 |
| from 1 to 2 | 1.0546 | 1.1279 | 1.1596 | 1.0720 | 1.1293 | 1.1844 |
| from lowest to highest | 1.0836 | 1.1987 | 1.2493 | 1.0796 | 1.1439 | 1.2055 |

*On Parents' Education.* Our findings confirm that fathers' education has globally more impact than mothers' (Behrman and Rosenzweig, 2002). This global result may however be misleading, because it covers strong cross-gender effects which reveal subtleties of the education transmission mechanisms which emerge only if one considers daughters and sons separately. We return to this point below.

*On the Strength of Effects through Stages.* We do not find the inverted-U shape pattern reported by Cameron and Heckman (1998); we find instead increasing marginal effects. This holds for parents' education, where in particular the effect of fathers' education on the transition from high school to college of sons is remarkably high, probability jumping from .07 to .21 as the three-level father's education goes from low to high. And it holds for parents' rearing efforts, which present a similar pattern.

*On Nature-vs-Nurture.* Unlike Behrman and Taubman (1989) the weight of family environment is not found to be overwhelming. Indeed the effects of passing from lowest to highest level are roughly similar for family endowment, interest and schooling (income seems to have a less relevant role). Of course, this may depend on the specification and interpretation of the $U^p$ unobservable, for we honestly cannot talk about genetics on the basis of the data we use.

In more detail the following facts seem worth of notice:

*Effect of* $U^p$. The three-level latent family background has more impact when it goes from low to average. For females, on whom it acts more, the change brings the probability of getting *some* educational attainment from .54 to .84, and that of reaching higher education from .11 to .34 (for males the corresponding values are .75 to .88 and .24 to .36). From a policy point of view this seems to suggest that intervention in degraded family environments may have appreciable effects on children's schooling achievements. It is also interesting that the effect varies significantly with gender: girls seem to 'absorb more' from the environment.

*Effect of* $R$. Interest has on boys effects comparable in size to those of family background on girls (as we see from the lowest-to-highest comparison). Moreover, the effect of interest too depends on gender. For boys parents' interest at 16 is as important as it is at 11; it raises the probability of the relevant transitions, to some level and to higher education, from .45 to .55 and from .09 to .13. For girls the effect at 11 is already lower than it is for boys, and it also decreases at 16. This seems to say that interest in boys is more result-oriented, which fits well with the findings on the effects of $S^p$ which we discuss next. For girls a possible interpretation is that interest reported to the teachers does not fully reflect support at home.

*Effect of $S^p$.* Parents' education has more effect on sons than on daughters. For boys the lowest-to-highest level change raises relevant probabilities, always to some level and to higher education, from .44 to .75 and from .07 to .25 (for girls it is .56 to .68 and .15 to .25). Of interest here is the marked cross-dependence of parents' and children's gender. Fathers' education has more impact on sons' and mothers' on daughters'. However, the male-male effect is much larger than the other: the three-level father education passing from medium to high doubles the probability that the son gets higher education from .11 to .21; for daughters the same effect of mothers' change raises that probability from .14 to .22. Moreover, the mother's education still helps the son to some extent, while the father's seems not to in his relation with the daughter. Thus there is a pronounced role effect in the male community in these data. Remember however that they reflect a social environment dating back a few decades ago; we would not be surprised if something has changed now.

As the reader will have noticed there is a row of ratios smaller than one in the table, corresponding to mothers' education passing from 0 (leaving school up to 14 years of age) to 1 (leaving up to 16). We do not venture in artificial explanations of these numbers, the truth being that in a system with 74 covariates one cannot expect that no coefficient 'goes wrong'. In this case the anomaly is probably due to unmodeled cross-correlation of this variable with others (rearing efforts being a possibility); but we had to make choices to obtain an estimable model, and could not include all possible interdependencies. The reader will judge for herself the plausibility of the global picture emerging from the exercise.

*Effect of $Y$.* As already remarked, family income is found to have a modest effect on children's education, both for sons and for daughters. This should be not too surprising, for after all it is not income *per se* which can raise children's achievements, and the system already contains the relevant transmission channels of parents' education and interest and the child's own schooling potential reflected in her endowment $U^c$. The three-level income is relevant when it becomes high, but even then it raises relevant probabilities only of a couple of percentage points.

## 7. Conclusion

In the society we have studied, which dates back to the seventies, education is almost exclusively transmitted from fathers to sons. Girls only absorb what they can from the family environment. In a male-dominated environment, we interpret this finding as saying that what family pushes goes through, to a non-negligible extent. Moreover, parents' interest matters as much as the parents' formal education, and income is not at all crucial except for higher education (for which however there are valid widespread alternative sources of funding). Thus the lesson appears to be that what matters is the family's attitude towards educational values. In a word, there is value in giving value to education.

## Appendix

**The EM algorithm.** The basic theory for estimation of a set of structural equations involving categorical observed and latent variables is based on the EM algorithm whose basic properties are well known (see for example Heckman and Singer, 1980, p. 302); a detailed discussion of the algorithm within a system of non linear structural equations is provided by Bergsma et al (2010), Section 6.5. Essentially, in the M (maximization) step, each node of the DAG is visited and the corresponding structural equation is fitted by maximum likelihood on the basis of the frequency distribution of each node on its parents as estimated in the E step. The E (expectation) step consists in computing the probability of belonging to the different categories of the latent variable, conditionally on the parameter estimates and on the categories of the observed variables.

The EM algorithm, in itself, does not provide estimated of standard errors. The estimates which we provide are computed by inverting the expected information matrix for the observed frequency distribution.

**Software.** A set of MatLab functions available from the authors have been prepared to perform the following tasks

(1) define the structure of a DAG and a system of non parametric structural equations possibly involving latent variables,

(2) define the set of link functions and specify the corresponding linearize model for each variable of interest;

(3) compute the maximum likelihood estimates by the EM algorithm and the expected information matrix;

(4) given the estimated distribution and the DAG, compute the direct and total effects for any selected sets of causes, mediators and responses.

**All Estimated Coefficients.** Here are all estimated coefficients from the system on page 8, equation by equation.

Table 3. All Coefficients

| | Daughters coeff | Daughters se | Sons coeff | Sons se |
|---|---|---|---|---|
| **Up** | | | | |
| a0 | 0.4844 | 0.1391 | 0.2896 | 0.1340 |
| a1 | -1.7145 | 0.1376 | -2.1829 | 0.1039 |
| **I 7** | | | | |
| a0 | 0.9429 | 0.1997 | 0.6217 | 0.1857 |
| b1(Up) | 2.2594 | 0.3358 | 2.9598 | 0.4460 |
| b2(Up) | 1.2392 | 0.2451 | 0.6025 | 0.2262 |
| **I 11** | | | | |
| a0 | 2.5109 | 0.4331 | 1.4182 | 0.2300 |
| b1(Up) | 2.4611 | 0.3487 | 3.3789 | 0.5661 |
| b2(Up) | 2.4931 | 0.4598 | 1.1250 | 0.2707 |
| **I 16** | | | | |
| a0 | 1.6306 | 0.2522 | 1.5454 | 0.2422 |
| b1(Up) | 2.7674 | 0.3951 | 2.2290 | 0.2544 |
| b2(Up) | 1.4133 | 0.3007 | 1.3905 | 0.2727 |
| **Sm** | | | | |
| a0 | 2.8580 | 0.2709 | 2.8140 | 0.2212 |
| a1 | -2.9359 | 0.1455 | -2.8686 | 0.1145 |
| b1(Up) | 0.7036 | 0.1750 | 0.6295 | 0.1437 |
| b2(Up) | 2.5027 | 0.2776 | 2.4124 | 0.2288 |

| | | | | |
|---|---|---|---|---|
| | | Sf | | |
| a0 | 1.4016 | 0.3852 | 4.2236 | 0.5462 |
| a1 | -2.7161 | 0.1759 | -3.2944 | 0.2488 |
| b1(Up) | 0.2087 | 0.2082 | 0.4638 | 0.1749 |
| b2(Up) | 2.9051 | 0.3717 | 5.7344 | 0.5335 |
| b1(Sm) | 1.6785 | 0.1471 | 1.8157 | 0.1423 |
| b2(Sm) | 0.5724 | 0.2537 | -0.8211 | 0.2678 |
| | | Y | | |
| a0 | 3.3764 | 0.4053 | 6.0278 | 0.5633 |
| a1 | -3.9939 | 0.1312 | -4.3703 | 0.1412 |
| b1(Up) | 0.6166 | 0.1884 | 0.9908 | 0.1880 |
| b2(Up) | 1.7081 | 0.3874 | 3.8858 | 0.5297 |
| b1(Sm) | 0.1970 | 0.1497 | 0.0816 | 0.1475 |
| b2(Sm) | -0.2624 | 0.2424 | -0.2829 | 0.2523 |
| b1(Sf) | 0.5323 | 0.1624 | 0.4722 | 0.1627 |
| b2(Sf) | 0.4575 | 0.2859 | -1.3545 | 0.4990 |
| | | Uc | | |
| a0 | 0.0949 | 0.8419 | 1.4641 | 0.6855 |
| a1 | -3.0600 | 0.3039 | -2.5192 | 0.1678 |
| b1(Up) | 2.8051 | 0.5843 | 1.4523 | 0.4476 |
| b2(Up) | 1.2421 | 0.6426 | 0.1656 | 0.5751 |
| b(I7) | -0.2345 | 0.2152 | -0.3143 | 0.1976 |
| b(I11) | -0.9265 | 0.2339 | -1.0002 | 0.2089 |
| b(I16) | -0.5668 | 0.2332 | -0.8651 | 0.1738 |
| b1(Sm) | 0.5925 | 0.2022 | -0.0279 | 0.1419 |
| b2(Sm) | -0.6575 | 0.2957 | -0.8102 | 0.2473 |
| b1(Sf) | -0.2310 | 0.1961 | -0.2161 | 0.1492 |
| b2(Sf) | -0.0092 | 0.3903 | -0.6482 | 0.4565 |
| | | EM | | |
| a0 | 3.8429 | 0.2086 | 3.7846 | 0.1789 |
| a1 | -2.6018 | 0.1278 | -2.5657 | 0.1206 |
| b1(Uc) | 2.8296 | 0.1876 | 2.8756 | 0.2046 |
| b2(Uc) | 2.3515 | 0.2027 | 2.4003 | 0.1728 |
| | | LM | | |
| a0 | 3.7758 | 0.3548 | 3.5850 | 0.2966 |
| a1 | -3.2564 | 0.1759 | -2.6998 | 0.1229 |
| b1(Uc) | 2.0644 | 0.2211 | 1.5640 | 0.2148 |
| b2(Uc) | 3.1387 | 0.2905 | 3.0301 | 0.2331 |
| b1(EM) | 0.3476 | 0.1984 | 0.5053 | 0.1843 |
| b2(EM) | 0.7452 | 0.2117 | 0.6517 | 0.1857 |

| | | | | | |
|---|---|---|---|---|---|
| | | | ER | | |
| a0 | 4.2970 | 0.2250 | | 3.7980 | 0.1831 |
| a1 | -2.9802 | 0.1577 | | -2.6814 | 0.1285 |
| b1(Uc) | 3.4529 | 0.2182 | | 3.3093 | 0.2487 |
| b2(Uc) | 2.1346 | 0.1948 | | 2.5773 | 0.1763 |
| | | | LR | | |
| a0 | 2.5832 | 0.3218 | | 3.0390 | 0.3005 |
| a1 | -3.0829 | 0.1476 | | -2.9077 | 0.1326 |
| b1(Uc) | 2.4067 | 0.2634 | | 2.5648 | 0.2745 |
| b2(Uc) | 2.0108 | 0.2156 | | 2.0336 | 0.1995 |
| b1(ER) | 0.6436 | 0.2358 | | 0.2479 | 0.2226 |
| b2(ER) | 1.3591 | 0.1749 | | 1.3494 | 0.1759 |
| | | | NC | | |
| a0 | 2.5415 | 0.2182 | | 1.6402 | 0.1361 |
| b1(Uc) | 1.6382 | 0.1771 | | 1.5266 | 0.1635 |
| b2(Uc) | 0.8467 | 0.2736 | | 1.1739 | 0.1745 |
| | | | Sc | | |
| a0 | 2.3737 | 0.2507 | | 2.5194 | 0.2681 |
| a1 | -2.3476 | 0.1132 | | -2.7011 | 0.1567 |
| a2 | -0.6647 | 0.0633 | | -0.9800 | 0.0829 |
| b1(Sm) | 0.1585 | 0.1491 | | -0.0600 | 0.1423 |
| b2(Sm) | 0.3622 | 0.2009 | | -0.1118 | 0.2003 |
| b1(Sf) | 0.0818 | 0.1539 | | 0.3571 | 0.1465 |
| b2(Sf) | 0.1118 | 0.2050 | | 0.6452 | 0.2132 |
| b1(Y) | 0.1162 | 0.1926 | | 0.0254 | 0.1867 |
| b2(Y) | 0.2433 | 0.1699 | | 0.2533 | 0.1762 |
| b1(Uc) | 2.8394 | 0.2107 | | 1.7948 | 0.2175 |
| b2(Uc) | 1.8843 | 0.1831 | | 3.0267 | 0.2144 |

## References


Agresti, Alan (2002): *Categorical Data Analysis*, Second Edition, John Wiley & Sons

Allman, Elisabeth S., Matias, Catherine and Rhodes, John A. (2009): "Identifiabnility of parameters in latent structure models with many observed variables", *The Annals of Statistics* **37**, pp. 3099-3132.

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

Arcidiacono, Peter and John Bailey Jones (2003): "Finite mixture distributions, sequential likelihood and the EM algorithm", *Econometrica* **71**, pp. 933–946

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

Behrman, Jere R. and Paul Taubman (1989): "Is schooling 'Mostly in the Genes'? Nature-Nurture decomposition using data on relatives", *Journal of Political Economy* **97**, pp. 1425–1446

Bartolucci, Francesco, Roberto Colombi and Antonio Forcina (2006): "An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints", *Statistica Sinica*, to appear

Bartolucci, Francesco and Forcina, Antonio (2005): "Likelihood inference on the underlying structure of IRT models", *Psychometrika* **30**, pp. 140–159

Bergsma, Wicher, Croon, Marcel and Hagenaars, Jaques, A. (2010): *Marginal models for dependent, clustered and longitudinal data*, Springer, London, New York.

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

Cameron, Stephen V., and James J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy* **106**, pp. 262–333

Cunha, Flavio, James J. Heckman and Salvador Navarro (2007): "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds", *International Economic Review* **48**, pp. 1273-1309

Heckman, James and Singer, B. (1980): "A Method for minimizing the impact of distributional assumptions in econometric models for duration data", *Econometrican* **52**, pp. 271–320

Holmlund Helena, Mikael Lindahl and Erik Plug (2011): "The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods", *Journal of Economic Literature* **49**, pp. 615–651

Mare, Robert D. (1980): "Social Background and School Continuaiton Decisions", *Journal of the American Statistical Association* **75**, pp. 295-305

Pearl, Judea (2000): *Causality*, Cambridge University Press

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

Rosemberg, Thomas J. (1971): "Identification in parametric models", *Econometrica* **39**, pp. 577-591.

Facoltà di Economia, Università di Perugia
*E-mail address*: forcina@stat.unipg.it

Facoltà di Economia, Università di Palermo
*E-mail address*: modica@unipa.it