

## STOCHASTIC MONOTONICITY IN INTERGENERATIONAL MOBILITY TABLES

VALENTINO DARDANONI,<sup>a\*</sup> MARIO FIORINI<sup>b</sup> AND ANTONIO FORCINA<sup>c</sup>

<sup>a</sup> *Università degli Studi di Palermo, Italy*

<sup>b</sup> *University of Technology Sydney, Australia*

<sup>c</sup> *Università degli Studi di Perugia, Italy*

### SUMMARY

The aim of this paper is to test for stochastic monotonicity in intergenerational socio-economic mobility tables. In other words, we question whether having a parent from a high socio-economic status is never worse than having one with a lower status. Using existing inferential procedures for testing unconditional stochastic monotonicity, we first test a set of 149 intergenerational mobility tables in 35 different countries and find that monotonicity cannot be rejected in hardly any table. In addition, we propose new testing procedures for testing conditional stochastic monotonicity and investigate whether monotonicity still holds after conditioning on a number of covariates such as education, cognitive and non-cognitive skills. Based on the NCDS cohort data from the UK, our results provide evidence that monotonicity holds, even conditionally. Moreover, we do not find large differences in our results when comparing social class and wage class mobility. Copyright © 2010 John Wiley & Sons, Ltd.

*Received 10 October 2008; Revised 14 August 2009*

### 1. INTRODUCTION

The extent to which individuals inherit their socio-economic status in a society has important implications for the debate concerning equal opportunities and social justice. Inspection of typical mobility tables and theoretical reasoning indicate that in most societies there is a general tendency for children from higher-status parents to somehow fare ‘better’ in social achievement than children from lower-status parents. To substantiate this general idea, first one needs to agree on the exact meaning of this statement; second, one has to devise an appropriate testing procedure for verifying the hypothesis; and finally, one has to apply the testing procedure to real-world data.

The first section of this paper will propose a precise definition of the idea that a child is better off by having a parent with a higher social status, based on the theory of monotone Markov chains. We then employ a rich dataset on intergenerational social mobility, which has been made available in Ganzeboom *et al.* (1989), to test the monotonicity assumption using the stochastic dominance testing procedure of Dardanoni and Forcina (1998). This large dataset contains 149 mobility tables combining information from 35 different countries and different years. It is a very comprehensive dataset for comparative mobility analysis, and it has the distinctive advantage of employing a consistent and well-defined definition of social status. Perhaps not surprisingly, we find that for most societies the monotonicity assumption cannot be rejected.

In the second part of the paper, we consider testing for stochastic monotonicity conditional on an appropriate set of covariates  $z$ . Starting with Becker and Tomes (1979), researchers have

---

\* Correspondence to: Valentino Dardanoni, Via F. Laurana 115, 90143 Palermo, Italy. E-mail: vdardano@unipa.it

proposed economic models of intergenerational mobility. It is by now widely accepted that parental transmission of skills, beliefs, motivation and social connections are all important in explaining the strong dependence between father and son social status. It is therefore natural to test whether the positive dependence between father's and son's status is still present after conditioning for some of these characteristics. On a similar line of thought, some researchers have turned their attention to the concept of equality of opportunity (EoP).<sup>1</sup> Dardanoni *et al.* (2006) for example, following the seminal contribution of Putterman *et al.* (1998) and Roemer (2000), describe EoP by distinguishing between circumstances and effort. Circumstances are aspects of the environment affecting the socio-economic status and for which society does not wish to hold individuals responsible. Effort is the set of actions affecting the status for which individuals are responsible. EoP implies that differences in status are ethically acceptable if they are due to differential effort but not if they are due to differential circumstances. This requires independence of parent and offspring socio-economic status conditional on those covariates that we consider effort.

The main theoretical challenge is then to devise statistical inference procedures to test for stochastic monotonicity conditional on observed covariates.<sup>2</sup> To this purpose, we extend the Dardanoni and Forcina (1998) test for stochastic monotonicity by allowing for conditioning on covariates. Our approach exploits recent advances in marginal modeling (see, for example, Bergsma and Rudas, 2002; Bartolucci *et al.*, 2007), along the lines of Bartolucci *et al.* (2001), who consider (unconditional) testing for a notion of positive dependence (*positive quadrant dependence*) which is weaker than the monotonicity assumption analyzed in this paper.

The importance of stochastic monotonicity in economics is highlighted in a recent paper by Lee *et al.* (2009), who propose a new test of stochastic monotonicity of a given continuous random variable  $Y$  with respect to another continuous random variable  $X$ . They then consider stochastic monotonicity in intergenerational income mobility as a relevant field of application of their procedure by using the Panel Study of Income Dynamics (PSID). Our approach can be considered complementary to theirs, since in order to apply our approach to continuous  $X$  and  $Y$  some grouping is required; on the other hand, if  $X$  and  $Y$  are discrete and ordered, our approach can be applied without the need to replace categories with arbitrary scores. In addition, our approach allows, under some parametric restrictions, conditioning on continuous and discrete covariates, while the approach taken by Lee *et al.* (2009) allows conditioning on covariates only when these are discrete and take only few values.<sup>3</sup>

We apply our methodology using the National Child Development Survey (NCDS), a UK dataset, which follows a cohort born in 1958 over its lifetime. Information on social class and wages is available both for the cohort members and their parents. The data also provide information on the educational attainment, cognitive and non-cognitive skills of the cohort members. Given the amount of characteristics observable to the researcher, the data are particularly fit to test for conditional dependence. Our results indicate that even though our control variables explain part of the intergenerational mobility mechanism, stochastic monotonicity in this sample holds both unconditionally and conditionally.

---

<sup>1</sup> Defining the appropriate concepts of equality of opportunity and testing them empirically is an area which is undergoing much current research (see, for example, Bourguignon *et al.*, 2003; Peragine, 2004; Lefranc *et al.*, 2006; Fleurbaey, 2008).

<sup>2</sup> Stochastic monotonicity is a property of a *single* mobility table; Formby *et al.* (2004) provide an excellent discussion of the statistical properties of various partial orderings and summary measures used to compare the degree of mobility of *different* tables.

<sup>3</sup> See Remark 2.2 in their paper.

## 2. MONOTONE TRANSITION MATRICES

Let  $X$  and  $Y$  denote, respectively, father's and son's lifetime socio-economic status, and assume they take  $k$  distinct values, which correspond to the  $k$  socio-economic classes ordered from worst to best. Consider then a standard discrete Markov chain of intergenerational social mobility: if the unit of observation is a family of father and son, the chain can be described by the equation  $\mathbf{p}'_y = \mathbf{p}'_x \mathbf{P}$ , where  $\mathbf{P}$  denotes the  $(k \times k)$  transition matrix, with typical element  $P_{ij} = \Pr(\text{son in } j \mid \text{father in } i) \geq 0$ , and  $\mathbf{p}_x, \mathbf{p}_y$  denote respectively the marginal distributions of father's and son's social status. The typical row  $i$  of an intergenerational transition matrix indicates the probability distribution faced by a son whose father belongs to social class  $i$ . As argued above, it is natural to expect that when social states are ordered sons whose fathers are in a higher social class are somewhat at an advantage with respect to the sons whose fathers are in a lower class. In a stochastic setting, this translates into the assumption that the 'lottery' faced by the son of a father in class  $i + 1$  is better than the 'lottery' faced by the son of a father in class  $i$ . A natural definition of a 'better lottery' in this context is given by the stochastic dominance ordering  $\succeq$ : given two  $(k \times 1)$  probability vectors  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , we say that  $\mathbf{q}_2 \succeq \mathbf{q}_1$  if  $q_{11} + \dots + q_{1j} \geq q_{21} + \dots + q_{2j}$  for all  $j < k$ .

The stochastic dominance ordering gives a precise definition to the intuitive notion of differential advantage. Let  $s$  denote a real-valued  $(k \times 1)$  vector of 'social status scores', where the typical element  $s_j$  is a quantitative measure of the value of social class  $j$ , and let  $\mathbf{p}_i$  (a row vector) denote the  $i$ th row of  $\mathbf{P}$ . An equivalent characterization of the stochastic dominance ordering is obtained in terms of expected social status:  $\mathbf{p}_{i+1}s \geq \mathbf{p}_i s$  for any increasing vector  $s$  if and only if  $\mathbf{p}_{i+1} \succeq \mathbf{p}_i$ .<sup>4</sup>

The intuitive notion of background advantage is captured in the discrete Markov chain by the so called 'monotonicity' assumption. The transition matrix of a discrete Markov chain with ordered states is called monotone (for applications, see Keilson and Kester, 1977; Conlisk, 1990; Dardanoni, 1993, 1995) if each row stochastically dominates the row above:  $\mathbf{p}_k \succeq \mathbf{p}_{k-1} \succeq \dots \succeq \mathbf{p}_1$ . Note that under the assumption of a constant transition matrix this relationship also holds for the grandfather, great-grandfather and so on since, if  $\mathbf{P}$  is monotone, so is  $\mathbf{P}^t$  for  $t = 1, 2, \dots$

A relevant, though extreme monotone transition matrix is the so-called 'equal opportunities' transition matrix (see, for example, Prais, 1955), where at time  $t$  each son faces an identical probability distribution regardless of his father's background. Given the transition equation, the equal opportunities transition matrix is equal to  $\mathbf{1}\mathbf{p}'_y$ , so that the stochastic dominance constraint is satisfied as an equality. This particular matrix will play an important role in the hypothesis testing of the monotonicity assumption.

## 3. TESTING UNCONDITIONAL MONOTONICITY

There is now an extensive statistical literature (see, for example, Robertson *et al.*, 1988; Silvapulle and Sen, 2005) on estimation and hypothesis testing in problems involving stochastic orderings. In particular, Robertson and Wright (1982) derive testing procedures based on maximum likelihood estimates of two stochastically ordered distributions, and Dykstra *et al.* (1991) obtain the maximum likelihood estimates of several multinomial distributions under uniform stochastic ordering, which is stronger than stochastic dominance. Dardanoni and Forcina (1998) extend these results and propose a nonparametric test for stochastic dominance which can be used to test monotonicity of the Markov chain of intergenerational mobility. In particular, Theorem 2 in their paper gives

<sup>4</sup>This is a well-known result in the stochastic dominance literature; see, for example, Lehmann (1955).

conservative bounds to the distribution of the likelihood ratio test statistic for testing monotonicity against an unrestricted alternative.

We perform Dardanoni and Forcina's procedure on a sample of cross-classification tables presented in Ganzeboom *et al.* (1989). This dataset, which contains 149 intergenerational class mobility tables from 35 countries, is one of the most comprehensive and well-structured datasets on intergenerational social mobility to date. Ganzeboom, Luijks and Treiman present the cross-classification of father's occupation by son's current occupation for representative national samples of men aged 21–64, with the characteristic that the tables conform to a well-specified six-category scheme. The six social classes, in ascending order of socio-economic status, are the following: (1) self-employed farmers and (unskilled) agricultural workers; (2) unskilled and semi-skilled manual workers; (3) lower-grade technicians, manual supervisors and skilled manual workers; (4) small proprietors with and without employees; (5) routine non-manual workers; (6) large proprietors, higher and lower professionals and managers. The use of a common and well-structured classification of social classes results in a substantial degree of comparability among the different tables.

By using the iterative quadratic programming algorithm described by Dardanoni and Forcina (1998), we computed the likelihood ratio for testing monotonicity against the unrestricted alternative for each table; convergence to the fifth decimal place of the likelihood function is usually obtained within the first four iterations. Let  $L_h$  denote the value of the likelihood ratio in the  $h$ th table of the set; following Dardanoni and Forcina (1998, Theorem 2), the  $p$ -value under the conservative unconditional chi-bar-squared distribution may be computed as

$$p_h = \sum_{i=0}^{(k-1)^2} \binom{(k-1)^2}{i} 2^{-(k-1)^2} Pr[\chi_i^2 > L_h]$$

The values of  $L_h$  and  $p_h$  for the 149 tables are plotted in Figure 1.

Computation of  $\alpha$  size critical values would be slightly more complicated as it requires setting  $p_h = \alpha$  and solving the equation above for  $L_h$  by numerical inversion. It emerges that out of 149 intergenerational class mobility tables monotonicity is rejected at the 1% significance level only for the transition matrices of Hungary 1962, Philippines 1968, Poland 1972 and Spain 1975. In addition, the monotonicity hypothesis is rejected at the 5% level for Hungary 1973 and 1983 and India 1963c. Thus it appears that monotonicity of the intergenerational transmission mechanism can generally be considered as an assumption supported by the real world.

#### 4. CONDITIONAL MONOTONICITY

The degree of intergenerational mobility for a given society as a whole has been the object of several studies. Either within the Galton model of regression to the mean, or using a transition matrix approach, almost all studies have found that parent's and offspring's adult status are not independent, but exhibit some form of positive association (for recent surveys on empirical findings in intergenerational mobility see, for example, Corak (2004) or the special issue on intergenerational mobility of the *B.E. Journal of Economic Analysis and Policy*, 2007, vol. 7(2)). The previous section has confirmed this 'fact of life', where positive dependence is precisely formulated in terms of stochastic dominance, and formal statistical inference procedures have been employed.

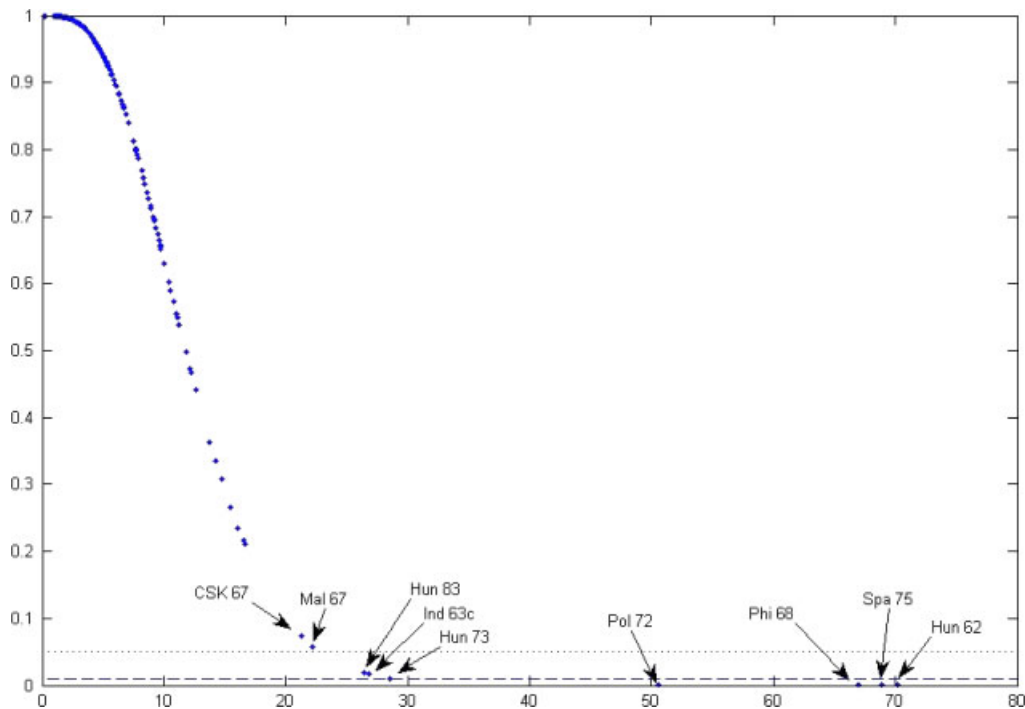


Figure 1.  $p$ -Values against likelihood ratios for testing monotonicity; horizontal lines mark 5% and 1% significance levels. This figure is available in color online at [wileyonlinelibrary.com/journal/jae](http://wileyonlinelibrary.com/journal/jae)

Starting with Becker and Tomes (1979), researchers have proposed economic models of intergenerational mobility to uncover the mechanism behind the transmission of social status. Becker and Tomes (1986), Solon (1999), Mulligan (1999), Han and Mulligan (2001) and Restuccia and Urrutia (2004) are all attempts in that direction. At a basic level, a simple model that assumes intergenerational transmission of ability and a human capital return to parental investment that is increasing in the child's ability can already generate a high degree of immobility. Adding imperfect capital markets to the model results in even less mobility. These models also show that the degree of mobility can be highly nonlinear across the father's and child's socio-economic distribution, whether socio-economic status is measured by wage, consumption or education. On a more intuitive ground, Bowles and Gintis (2002), Erikson and Goldthorpe (2002) and Blanden *et al.* (2007) suggest that more than a simple transmission of ability might be in place. Other factors such as race, geographical location, wealth, risk aversion, discounting of the future, non-cognitive skills, but also height and beauty, can be transmitted and generate the correlation in status. Not surprisingly, most of these papers have also tried to investigate the black box empirically. However, none of these studies can explain more than 60% of the overall correlation. Bjorklund *et al.* (2006) use unique Swedish data with information on adopted children's biological and adoptive parents to estimate intergenerational mobility associations in earnings and education. They find that both pre- and post-birth factors contribute to intergenerational earnings and education transmission. The distinction between nature and nurture is particularly important if we are asked to judge meritocracy and equality of opportunity.

Taking a slightly different line of thought, Dardanoni *et al.* (2006) discuss different notions of equality of opportunity based on the distinction between circumstances and effort: ‘Agreement is widespread that equality of opportunity holds in a society if the chances that individuals have to succeed depend only on their own efforts and not upon extraneous circumstances that may inhibit or expand those chances. What is contentious, however, is what constitutes effort and circumstances.’ In their paper they describe four channels through which parents affect status in an intergenerational context: social connections, the formation of social beliefs and skills, the transmission of native ability and the instillation of preferences and aspirations. Various notions of EoP depend on whether these channels are regarded as circumstances or effort. In other words, if we consider all those channels as circumstances out of an individual’s control, then EoP implies perfect intergenerational mobility. This is perhaps the strongest definition of EoP where parent’s and offspring’s status must be independent. Less stringent notions of EoP allow for some of those channels to be influenced by the offspring. In turn, this requires independence conditional on those covariates  $z$  that we consider individual effort.  $z$  could include measures of preferences and aspirations, native ability, social beliefs and skills, and social connections.

#### 4.1. A Statistical Model

As before, let  $X$  and  $Y$  denote, respectively, father’s and son’s social class, let  $z$  be a vector of covariates that may affect the joint distribution determined by  $\pi(z)$ , the vector containing the probabilities  $P(X = i, Y = j|z)$  arranged in lexicographic order, with the  $Y$  categories running faster. In words, we are assuming that the data-generating process is multinomial, a specification which implies the following restrictions:

1. Conditionally on covariates, the response variables  $(X, Y)$  of separate father-son pairs are independent.
2. The dependence of  $\pi(z)$  on  $z$  is correctly specified.

Because assumption 1 is standard and plausible in most contexts, the crucial issue is how to model the dependence of transition probabilities on covariates.

If  $z$  was discrete and a sufficient number of observations were available for each distinct configuration of  $z$ , Dardanoni and Forcina’s (1998) unconditional test procedure could be performed for each subpopulation or for the set as a whole. However, this is unlikely to happen whenever, as in the application we consider in this paper, the number of covariates is reasonably large and/or certain covariates assume a large number of distinct values. An approximate solution would be to replace the true covariate values with a few categories corresponding to suitable ranges of values. However, in spite of the coarseness of the covariate reduction process, there could still be several tables containing observed zero frequencies which may preclude testing the monotonicity assumption. An application of this procedure is discussed in Section 6, where we propose an empirical solution to the typical difficulties that may arise in applications.

When covariates assume a large number of distinct values, a meaningful approach is to choose a suitable *link function* which maps the saturated model for the bivariate distribution of  $(X, Y)|z$  into a linear regression model. The link function that we propose is based on the mapping of the conditional distribution of  $(X, Y)|z$  into a set of  $(k - 1)$  row and  $(k - 1)$  column marginal parameters and  $(k - 1)^2$  association parameters. For the row and column marginal parameters we

propose using *global logits* (see, for example, Agresti, 2002) which suit the ordered qualitative nature of these variables:

$$\rho_i(\mathbf{z}) = \log \left[ \frac{P(X > i|\mathbf{z})}{P(X \leq i|\mathbf{z})} \right], \quad i = 1, \dots, k - 1,$$

$$\xi_j(\mathbf{z}) = \log \left[ \frac{P(Y > j|\mathbf{z})}{P(Y \leq j|\mathbf{z})} \right], \quad j = 1, \dots, k - 1$$

Global logits can be seen as the natural generalization of the standard binary logits to an ordered variable: global logits are binary logits computed on successive splits of the response categories into a ‘low’ and a ‘high’ set.

For the association parameters we propose using *local–global log-odds ratios* (see, for example, Agresti, 2002), which may be defined as a contrast between conditional global logits:

$$\tau_{ij}(\mathbf{z}) = \log \frac{P(Y > j|X = i + 1, \mathbf{z})}{P(Y \leq j|X = i + 1, \mathbf{z})} - \log \frac{P(Y > j|X = i, \mathbf{z})}{P(Y \leq j|X = i, \mathbf{z})}, \quad i, j = 1, \dots, k - 1$$

The following result (discussed in Douglas *et al.*, 1990) provides a motivation for using the local–global log-odds ratios.

**Lemma 1** *The conditional transition matrix defined by  $\boldsymbol{\pi}(\mathbf{z})$  is monotone if and only if  $\tau_{ij} \geq 0$  for all  $i, j = 1, \dots, k - 1$ .*

**Proof:** By definition, monotonicity of  $\boldsymbol{\pi}(\mathbf{z})$  implies that  $P(Y > j|X = i + 1, \mathbf{z}) \geq P(Y > j|X = i, \mathbf{z})$ ; the result follows by mapping each side of the inequality into global logits, a function which is strictly monotone.

Henceforth, whenever appropriate, we will simply denote as *logits* the global logits  $\rho$  and  $\xi$ , and *log-odds ratios* the local–global log-odds ratios  $\tau$ . Collect now all the logit and log-odds ratio parameters into the vectors  $\boldsymbol{\rho}(\mathbf{z})$ ,  $\boldsymbol{\xi}(\mathbf{z})$  and  $\boldsymbol{\tau}(\mathbf{z})$  (by letting the  $j$  index run faster than  $i$ ) and let

$$\boldsymbol{\lambda}(\mathbf{z}) = [\boldsymbol{\rho}(\mathbf{z})', \boldsymbol{\xi}(\mathbf{z})', \boldsymbol{\tau}(\mathbf{z})']'$$

this has dimension  $2(k - 1) + (k - 1)^2 = k^2 - 1$ , which equals the number of free parameters in  $\boldsymbol{\pi}(\mathbf{z})$ .

The mapping between  $\boldsymbol{\pi}(\mathbf{z})$  and  $\boldsymbol{\lambda}(\mathbf{z})$  may be written in compact matrix form as

$$\boldsymbol{\lambda}(\mathbf{z}) = \mathbf{C} \log[\mathbf{M}\boldsymbol{\pi}(\mathbf{z})]$$

where the log operator is computed elementwise; Bartolucci *et al.* (2007) describe an algorithm for constructing the matrix of row contrasts  $\mathbf{C}$  and the marginalization matrix  $\mathbf{M}$ . In addition, they show that the mapping is invertible and differentiable for all strictly positive  $\boldsymbol{\pi}(\mathbf{z})$  (see their Theorem 1). Thus the set of marginal and association parameters  $\boldsymbol{\lambda}(\mathbf{z})$  is a one-to-one mapping of  $\boldsymbol{\pi}(\mathbf{z})$  with no modeling restriction. However, as argued above, when  $\mathbf{z}$  takes a large number

of distinct configurations, to gather information from such sparse data the dependence of  $\lambda(\mathbf{z})$  on covariates may be constrained to satisfy a linear model of the form

$$\begin{aligned}\rho_i(\mathbf{z}) &= \alpha_i^X + z'_X \boldsymbol{\beta}_i^X, \quad i = 1, \dots, k-1 \\ \xi_j(\mathbf{z}) &= \alpha_j^Y + z'_Y \boldsymbol{\beta}_j^Y, \quad j = 1, \dots, k-1 \\ \tau_{ij}(\mathbf{z}) &= \alpha_{ij}^{XY} + z'_{XY} \boldsymbol{\beta}_{ij}^{XY}, \quad i, j = 1, \dots, k-1\end{aligned}$$

where  $\mathbf{z}_X$ ,  $\mathbf{z}_Y$  and  $\mathbf{z}_{XY}$  denote respectively the subset of observed covariates  $\mathbf{z}$  which are supposed to affect the marginal distribution of  $X$  and  $Y$  and their dependence structure. This may be written in a compact matrix form as

$$\lambda(\mathbf{z}) = \mathbf{C} \log[\mathbf{M} \boldsymbol{\pi}(\mathbf{z})] = \mathbf{Z} \boldsymbol{\psi} \quad (1)$$

where  $\boldsymbol{\psi}$  is obtained by stacking one below the other the parameters  $\boldsymbol{\alpha}^X$ ,  $\boldsymbol{\beta}^X$ ,  $\boldsymbol{\alpha}^Y$ ,  $\boldsymbol{\beta}^Y$ ,  $\boldsymbol{\alpha}^{XY}$ ,  $\boldsymbol{\beta}^{XY}$  and  $\mathbf{Z}$  is block diagonal.

To gain additional insights into the substantive properties of the local–global log-odds ratios, it may be instructive to consider the following version of the standard Galton intergenerational regression model adapted to the discrete ordered nature of  $X$  and  $Y$ :

$$Y^* | (X = i, \mathbf{z}) = \alpha + \beta_i + z\gamma' + \varepsilon$$

where  $Y^* | (X = i, \mathbf{z})$  is the latent continuous counterpart of the observed social class of a son whose father is in social class  $X = i$  and has covariates  $\mathbf{z}$ . If, in addition, we assume that the error term  $\varepsilon$  has a standard logistic distribution, then the above latent version of the Galton model implies that the local–global log odds ratio  $\tau_{ij}(\mathbf{z}) = \beta_{i+1} - \beta_i$  and there is stochastic monotonicity if and only if  $\beta_{i+1} \geq \beta_i$ . When compared to our model, the ordered logit version of the Galton model is much more restrictive: the log-odds ratio which is the measure of immobility depends only on the father's social class (and not on the son's) and is not affected by covariates. The motivation for the substantially more complex model proposed in this paper is that, because the main purpose of the analysis is to assess monotonicity, we would like to control for observed covariates as much as possible.

## 4.2. Hypotheses of Interest

A convenient feature of the parametrization defined above is that the hypothesis of stochastic monotonicity conditionally on relevant covariates can be expressed in the form of linear inequality constraints on an appropriate sub-vector of the  $\boldsymbol{\psi}$ . In the sequel, hypotheses of interest and corresponding subsets of the parameter space will be denoted by the same symbol. In particular, the hypothesis of stochastic monotonicity can be written as

$$\mathcal{H}_1 : \tau_{ij} = \alpha_{ij}^{XY} + z'_{XY} \boldsymbol{\beta}_{ij}^{XY} \geq 0 \quad \forall \mathbf{z}_{XY}; \quad i, j = 1, \dots, k-1 \quad (2)$$

The constraints implied by  $\mathcal{H}_1$  may be written in matrix form as  $\mathbf{D} \boldsymbol{\psi} \geq \mathbf{0}$ , where  $\mathbf{D}$  is obtained by stacking one below the other the matrices  $(\mathbf{0}_{am} \mathbf{I}_a \mathbf{1}_a z'_{XYu})$  for  $u = 1, \dots, n$ , where  $a = (k-1)^2$ ,  $m$  is the overall size of  $(\boldsymbol{\alpha}^X, \boldsymbol{\beta}^X, \boldsymbol{\alpha}^Y, \boldsymbol{\beta}^Y)$  (the marginal part of the model, which is unconstrained) and  $\mathbf{z}_{XYu}$  is the vector  $\mathbf{z}_{XY}$  observed in unit  $u$ .



In typical applications the matrix  $D$  may have many more rows than columns (for example, in the first application discussed below, there are 7768 inequalities with only 76 variables) and some of these might be redundant. This happens if a subset of the rows of  $D$  are a linear combination of other rows with non-negative coefficients. Thus detecting redundant inequalities is not a trivial task (see, for example, Schrijver, 1986); a random search algorithm applied to our context did not detect any redundancy. In any case, redundancy can only affect computational efficiency and our estimation algorithm can handle a large number of inequality constraints and be fast.

The hypothesis of equality of opportunities can be written as

$$\mathcal{H}_0 : \tau_{ij} = \alpha_{ij}^{XY} + z'_{XY} \beta_{ij}^{XY} = 0 \quad \forall z_{XY}; i, j = 1, \dots, k - 1 \quad (3)$$

which, equivalently, can be written as

$$\mathcal{H}'_0 : \alpha^{XY} = \mathbf{0} \text{ and } \beta^{XY} = \mathbf{0}$$

Finally, we will denote the unrestricted model by  $\mathcal{H}_2$ .

### 4.3. Estimation

Suppose now we have independent observations  $(X_u, Y_u, z_u)$  for a sample of  $n$  units. Let  $\mathbf{t}(z_u)$  be the frequency table corresponding to the observation on unit  $u$  written into a vector with the same lexicographic rule used for  $\boldsymbol{\pi}(\mathbf{z})$ ; clearly this will be a vector of 0's except for a 1 in the position corresponding to the observed pair  $X_i(u), Y_j(u)$ . To simplify notation, in the sequel we write  $\mathbf{t}(u)$  instead of  $\mathbf{t}(z_u)$ ; a similar convention will be adopted for any vector which depends on  $z_u$ . Under independent sampling, conditionally on  $z_u$ ,  $\mathbf{t}(u)$  has a multinomial distribution with vector of probabilities  $\boldsymbol{\pi}(u)$ . An algorithm for maximizing the multinomial log-likelihood

$$L = \sum_u \mathbf{t}(u)' \log[\boldsymbol{\pi}(u)] \quad (4)$$

is described by Colombi and Forcina (2001) and Dardanoni and Forcina (2008), and is based on an extension of an algorithm due to Aitchison and Silvey (1958). Essentially, at each step the algorithm does the following, until convergence:

- compute a quadratic approximation of the log-likelihood in terms of the canonical (log-linear) parameters;
- compute a linear approximation of the canonical parameters in terms of  $\boldsymbol{\psi}$ ;
- solve a weighted least squares problem.

When inequality constraints are present, the weighted least squares problem to be solved at each step requires a quadratic optimization which is itself iterative: there are many algorithms for quadratic optimization under inequality constraints, which are usually very fast and reliable.

### 4.4. Hypothesis Testing

In the following let  $\boldsymbol{\psi}_2$  denote the unrestricted maximum likelihood estimate (MLE) of  $\boldsymbol{\psi}$ ,  $\boldsymbol{\psi}_1$  be the MLE of  $\boldsymbol{\psi}$  under the stochastic monotonicity hypothesis  $\mathcal{H}_1$  and  $\boldsymbol{\psi}_0$  be the MLE of

$\boldsymbol{\psi}$  under the equality of opportunity hypothesis  $\mathcal{H}_0$ . Let  $\mathbf{F}(\boldsymbol{\psi})$  denote the expected information matrix with respect to  $\boldsymbol{\psi}$ . From standard asymptotic results it follows that, if  $\mathcal{H}_0$  is true and, as  $n$  increases,  $\mathbf{F}(\boldsymbol{\psi})/n$  is of full rank,  $\boldsymbol{\psi}_2$  has an asymptotic normal distribution  $N(\boldsymbol{\psi}, \mathbf{F}(\boldsymbol{\psi})^{-1})$ . Therefore, hypotheses on single elements of  $\boldsymbol{\psi}$  may be tested by comparing the estimate with the corresponding standard error. Joint testing may be based on the asymptotic distribution of the LR statistic. Recall the well-known result that the LR for testing the unrestricted model against  $\mathcal{H}_0$

$$T_{02} = 2(L(\boldsymbol{\psi}_2) - L(\boldsymbol{\psi}_0)) \quad (5)$$

has asymptotic  $\chi_r^2$  distribution, where  $r$  is the sum of the dimensions of  $\boldsymbol{\alpha}^{XY}$  and  $\boldsymbol{\beta}^{XY}$ .

When inequalities are involved, the testing problem may be split into testing the unrestricted model  $\mathcal{H}_2$  against  $\mathcal{H}_1$  and testing  $\mathcal{H}_1$  against  $\mathcal{H}_0$ . The corresponding LR statistics may be written as

$$T_{01} = 2(L(\boldsymbol{\psi}_1) - L(\boldsymbol{\psi}_0)) \quad (6)$$

$$T_{12} = 2(L(\boldsymbol{\psi}_2) - L(\boldsymbol{\psi}_1)) \quad (7)$$

It is also useful to recall the following (see, for example, Shapiro, 1988; Wolak, 1991).

**Definition 1** Let  $\mathbf{b} \sim N(\mathbf{0}, \mathbf{V})$  be a  $k$ -dimensional normal random vector, and let  $\mathcal{C}$  be a convex cone in  $R^k$ . The squared norm of the projection of  $\mathbf{b}$  onto  $\mathcal{C}$  is a chi-bar-squared random variable  $\bar{\chi}^2(\mathcal{C}, \mathbf{V})$

$$\bar{\chi}^2(\mathcal{C}, \mathbf{V}) = \mathbf{b}'\mathbf{V}^{-1}\mathbf{b} - \min_{\mathbf{a} \in \mathcal{C}} (\mathbf{b} - \mathbf{a})'\mathbf{V}^{-1}(\mathbf{b} - \mathbf{a}) \quad (8)$$

and has distribution function:

$$Pr(\bar{\chi}^2(\mathcal{C}, \mathbf{V}) \leq x) = \sum_{i=0}^k w_i(\mathcal{C}, \mathbf{V}) F_{\chi}(x, i) \quad (9)$$

where  $F_{\chi}(x, i)$  denotes the distribution function of a chi-square with  $i$  d.f. and  $w_i(\mathcal{C}, \mathbf{V})$  is the probability that the projection of  $\mathbf{b}$  onto  $\mathcal{C}$  belongs to a face of dimension  $i$ .

When the above is applied to our context by noting that  $\mathcal{H}_1$  corresponds to a convex cone in the parameter space, the following result can be easily derived from Dardanoni and Forcina (1999):

**Proposition 1** Under the assumption that the true value  $\boldsymbol{\psi}^o$  belongs to the interior of  $\mathcal{H}_0$ , the asymptotic distributions of  $T_{01}$  and  $T_{12}$  are

$$\begin{aligned} T_{01} &\longrightarrow \bar{\chi}^2(\mathcal{H}_1, \mathbf{F}^{-1}(\boldsymbol{\psi}^o)) \\ T_{12} &\longrightarrow \bar{\chi}^2(\mathcal{H}_1^o, \mathbf{F}^{-1}(\boldsymbol{\psi}^o)) \end{aligned} \quad (10)$$

where  $\mathcal{H}_1^o$  denotes the dual of  $\mathcal{H}_1$  in the metric determined by the information matrix at  $\boldsymbol{\psi}^o$ .<sup>5</sup>

<sup>5</sup> That is,  $\mathcal{H}_1^o = \{v : v'\mathbf{F}^{-1}(\boldsymbol{\psi}^o)\mathbf{u} \leq 0, \forall \mathbf{u} \in \mathcal{H}_1\}$ .

Asymptotic  $p$ -values for these statistics depend on the probability weights  $w_i(\mathcal{H}_1, \mathbf{F}^{-1}(\boldsymbol{\psi}^o))$ . Unfortunately, except in very small problems, no closed form expression is available for the computation of these weights. However, reliable estimates may be obtained by Monte Carlo simulations as described by Dardanoni and Forcina (1998).

It is worth recalling briefly the idea upon which the estimation of the probability weights may be based. First transform the original problem with  $\mathbf{D}\boldsymbol{\psi} \geq \mathbf{0}$  and information matrix  $\mathbf{F}$  into the canonical form  $\overline{\mathbf{D}}\boldsymbol{\xi} \geq \mathbf{0}$  and identity information matrix of size  $d$ , the dimension of the space. Then, at each step:

- sample at random points from a  $N(\mathbf{0}, \mathbf{I}_d)$ ;
- project onto  $\overline{\mathbf{D}}\boldsymbol{\xi} \geq \mathbf{0}$  and let  $\overline{\mathbf{D}}_0$  be the subset of rows of  $\overline{\mathbf{D}}$  such that  $\overline{\mathbf{D}}_0\hat{\boldsymbol{\xi}} = \mathbf{0}$ ;
- let  $g$  be the size of the orthogonal complement of  $\overline{\mathbf{D}}_0$ ;
- set  $w_g, g = 0, \dots, d$ , equal to the relative frequency out of  $m$  draws of each  $g$ .

Since  $\mathcal{H}_1$  is a composite hypothesis, one should search for the value of  $\boldsymbol{\psi} \in \mathcal{H}_1$  which gives the least favorable asymptotic null distribution for  $T_{12}$  and, as Wolak (1991) has shown, this value does not necessarily belong to  $\mathcal{H}_0$ . Dardanoni and Forcina (1998) discuss some practical solutions to this problem. Finally, note that the joint distribution of  $T_{01}$  and  $T_{12}$  can also be derived (see Dardanoni and Forcina, 1999, for details), where use of this joint distribution for hypotheses testing is also compared with alternative testing procedures.

## 5. CONDITIONAL MONOTONICITY IN THE NCDS DATASET

In order to test for conditional monotonicity we use the National Child Development Study (NCDS), an ongoing survey that originally targeted over 17,000 babies born in Britain in the week 3–9 March 1958. Surviving members of this birth cohort have been surveyed on seven further occasions in order to monitor their changing health, education, social and economic circumstances: in 1965 (age 7), 1969 (age 11), 1974 (age 16), 1981 (age 23), 1991 (age 33), 1999 (age 41) and 2004 (age 46). At the age of 7, 11 and 16, mathematics, reading and general skills tests were taken by the cohort members, while at the age of 7 and 11 information on non-cognitive skills was also collected.

From the age of 16 individuals could leave education and enter the labor market. For those who stayed, the surveys from 1981 onwards, together with a 1978 school survey, provide information on the qualifications attained. Data on wages and social class were gathered at age 23, 33, 41 and 46. To study intergenerational mobility we also need data on parental socio-economic status. The first four surveys (1958, 1965, 1969, 1974) contain data about parental background including age, education (1974), wage (1974), social class of father (1965, 1969, 1974) and mother (1974). These datasets therefore bring together information on socio-economic status for two consecutive generations.<sup>6</sup>

<sup>6</sup> The NCDS data managers have also collected information on the cohort members' children in 1991. However, at that time these children were still very young and had not yet entered the labor market. No further information on these and new children of the cohort members was gathered in the 1999 and 2004 surveys.

### 5.1. Measurements of Socio-economic Status

To apply our stochastic monotonicity tests, we first have to find suitable variables representing socio-economic status  $X$  and  $Y$ . Since true socio-economic status is not observed, intergenerational mobility scholars typically employ either wage (income) or social class in their analysis.

Economists often look at wages or income as the most important observed measure of socio-economic status. However, both can be very sensitive to measurement error or temporary shocks such as short periods of unemployment, health shocks or even short business cycles. In the standard linear model, using current wages rather than true lifetime socio-economic status can result in attenuation bias. Researchers try to solve this problem either by using average wage (income), whenever the data provide repeated observations, or by using an instrumental variable approach (see, for example, Zimmerman, 1992, for a discussion on the effect measurement error on measured mobility in the linear regression model). Note that the attenuation bias holds also in our discrete mobility tables setting (Carroll *et al.*, 2006, contains a thorough discussion of measurement error in nonlinear models); see, for example, Neuhaus, 1999, for an analysis in the logit model.

On the other hand, sociologists prefer to use social class as a measure of lifetime socio-economic status (see, for instance, Erikson and Goldthorpe, 2002). They argue that not only is social class less sensitive to temporary shocks, but also it includes immaterial aspects such as prestige and power. The main limitations of social class originate from its subjectivity, since it is the researcher, using a combination of labor market occupation, education and other factors, that imputes the social class of the individual, sometimes also in an ordered manner, from the more prestigious occupation downwards. The way occupations are coded into social classes can sometimes affect the results. Moreover, the prestige associated with a social class can be time varying, i.e. being in, let us say, the skilled manual category in the 1960s is very different from being in this category today, and this is a relevant problem in the case of intergenerational mobility, where we look at individuals born 20–40 years apart. Finally, within a class there could clearly be a large degree of heterogeneity; a painter and a carpenter may both be defined as skilled manual workers, but of course the socio-economic status of, say, Picasso is very different from that of an unknown painter. Yet some of these problems affect wages (income) too. A miner might earn even more than an academic professor due to the risk associated with his job, yet not many professors would choose to become miners.

Choosing how to measure socio-economic status inevitably depends on the data available. In our data there is not enough information to construct a reliable measure of a father's permanent wage since this is observable only at one point in time. To overcome this problem Dearden *et al.* (1997) regress current wage on non-time-varying factors such as education and social class, and then use the predicted variable as a measure of permanent wage. However, while there is no guarantee that this procedure eliminates attenuation bias, it also leads to a mix of wage and social class mobility (because social class is used to predict wages) and it is not really suited to test for conditional mobility (because it directly uses education to predict wages). Moreover, as we show below, there are several individuals for whom wages are not available while we can observe social class. For these reasons, we consider social class a more reliable measure of lifetime socio-economic status than wages, but in the application below we will also check our main results for the case of wage mobility.

In this application we use the 1991 data on the cohort member socio-economic status coupled with the 1974 data on father's status. These are also the NCDS surveys used by Dearden *et al.*

(1997) and Blanden *et al.* (2007) in their studies on wage mobility. We first select all male cohort members for which we observe cognitive and non-cognitive skills at age 7, 11 and 16 (cognitive skills only), educational attainment and father's age. We then select those individuals for whom we observe both social class in 1991 and father's class in 1974.<sup>7</sup>

Table I shows summary statistics for the social class measures. Social class is a status variable grouping occupations into six broad categories, ordered on a skill basis.<sup>8</sup> In the data, sons are more skilled than their parents were and the distribution of sons' socio-economic status is clearly stochastically larger than that of their fathers.

Table II shows summary statistics for parents' and sons' weekly net wages, sons' highest educational qualification and fathers' age. The NCDS collected information on parental net wages only in 1974, with separate questions about fathers' and mothers' wages and other sources of income. Note that, in the original coding, wage was grouped into 12 wage bands and we assign to each observation the median value of the observed band. Finally, in the data there are several individuals for whom fathers' social class is available but wages are not, while the opposite is quite rare. There are a number of explanations. Some individuals (or their fathers) are self-employed

Table I. Social class: males

	Son		Father	
Professional	7.21	(7.21)	5.97	(5.97)
Intermediate	35.84	(43.05)	21.63	(27.60)
Skilled Non-manual	12.92	(55.97)	10.87	(38.47)
Skilled Manual	28.89	(84.86)	45.01	(83.47)
Semiskilled	12.77	(97.63)	13.18	(96.65)
Unskilled	2.37	(100.00)	3.35	(100.00)
Observations	1942		1942	

*Note:* Values are percentages. Numbers in parentheses are cumulated percentages.

Table II. Summary statistics

	Obs.	Mean	SD
Son's net wage	1341	301.07	102.79
Father's net wage	1486	232.56	83.93
No qualification	1942	0.45	0.49
O-levels	1942	0.33	0.46
A-levels	1942	0.09	0.29
Higher education	1942	0.13	0.33
Father's age	1942	46.55	6.12

*Note:* Father and son net wages are in January 2001 prices.

<sup>7</sup> We select males to make our results comparable to previous studies.

<sup>8</sup> Social class variables are derived according to the Registrar Classification (RG). This classification imputes social class using only information about occupation. This is a quite common and simple grouping methodology, even though some sociologists have proposed alternative ones. The Goldthorpe class schema, for instance (see Erikson and Goldthorpe, 2002) aims to capture qualitative differences in employment relations. Unfortunately, the classes distinguished by this schema are not consistently ordered according to some inherent hierarchical principle. Therefore the Goldthorpe class schema does not suit our statistical model.

and their wages are not reliable. For other individuals the wage is not available either because they were unemployed or because they chose not to report it.

## 5.2. Preliminary Analysis

Before trying to control for covariates, it may be interesting to examine briefly the unconditional mobility table for the whole population. To be consistent with the analysis to follow, we group social classes into three categories roughly corresponding to a high/medium/low-skilled partition:

- 1 Semi-skilled + Unskilled
- 2 Skilled non-manual + Skilled manual
- 3 Professional + Intermediate

Having more categories may allow a more detailed analysis of the pattern of mobility. However, since the unrestricted model is rather complex because we want to control for as many covariates as possible, with more categories the maximum likelihood algorithm has convergence difficulties due, probably, to the sparseness of the data.

Table III shows the mobility table using sons' social class in 1991 and fathers' class in 1974. In the table we also include the chi-square statistic for independence, with degrees of freedom in parentheses. The chi-square statistic is very large, leading to rejection of the hypothesis of independence.

Table IV shows the log-odds ratios in the unconditional table and the corresponding standard errors computed with a first-order approximation (Delta method).

This indicates that, as could be expected, monotonicity cannot be rejected at any significance level because all the estimated log-odds ratios are positive and significantly different from 0; thus the independence hypothesis can be rejected with overwhelming evidence. The question is, from the statistical point of view, how much of the positive association between  $X$  and  $Y$  is induced by heterogeneity which could be accounted for by controlling for the observed

Table III. Social class mobility table

Father/son class	1	2	3	
1	4.53	8.08	3.91	16.53
2	8.81	26.06	21.01	55.87
3	1.80	7.67	18.13	27.60
	15.14	41.81	43.05	100.00

Note: 1942 observations. Numbers in table are percentages. Chi-square (4) = 192.73.

Table IV. Unconditional local–global log-odds ratios

	Rows 1/2		Rows 2/3	
	$\tau_{11}$	$\tau_{12}$	$\tau_{21}$	$\tau_{22}$
Estimates	0.7025	0.6641	0.9851	1.1551
SE	0.1503	0.1455	0.1937	0.1105

covariates. Furthermore, as discussed above in Section 4, for the purpose of evaluating equality of opportunities in a society one might want to test independence of fathers' and sons' social class *after* conditioning on an appropriate set of covariates.

### 5.3. Control Variables

As we explained in Section 4.1, our aim is to test for dependence conditional on some characteristics of the parents and offspring. Since most economic models of intergenerational mobility assume that the transmission of the ability endowment across generations is one of the main reasons behind immobility (see, for example, Becker and Tomes, 1979; Grawe and Mulligan, 2002), a starting point is to investigate monotonicity conditional on cognitive skills. However, as Becker and Tomes (1979), Bowles and Gintis (2002), Erikson and Goldthorpe (2002) and Dardanoni *et al.* (2006) suggest, cognitive ability is just one dimension of the endowment stock. Recently Heckman *et al.* (2006) and Cunha and Heckman (2007) show that non-cognitive skills can also explain a diverse array of outcomes such as schooling choices, wages, employment and work experience. It is quite likely that non-cognitive skills are also transmitted across generations, if not genetically, because of parental behavior and education. Finally, the human capital models predicts that high-status parents invest more in their children. In turn, this implies that these children have more human capital. Therefore we choose to test for monotonicity conditional on the educational attainment, cognitive and non-cognitive skills of the offspring (son). Since the fathers were of different ages at the moment of the survey, we also control for fathers' age.

Given the education system faced by the 1958 cohort, its educational attainment is measured by four dummy variables corresponding to 'No qualification', 'O-levels', 'A-levels' and 'Higher education'. In the UK, schooling is compulsory up to the age of 16, when individuals can, at the end of the scholastic year, stay in education or enter the labor market. Those who stay on at age 16 enroll for O-levels or CSE qualifications, which are taken immediately at the end of the scholastic year. These students are still aged 16 when they obtain the qualification. In the autumn term of the same year, those who successfully obtained five or more O-levels/CSEs can enroll for A-levels. These last 2 years, until individuals are aged 18. Passing two A-levels constitutes the minimum level required for entry into higher education. Once the student has completed A-levels, he can gain admission to a university, polytechnic or college of higher education, where a first degree is obtained. The time needed to gain a degree varies by subject but in the majority of cases it is 3 years.

To control for cognitive skills we use mathematics and reading test scores. These tests were taken by the cohort members at the age of 7, 11 and 16. We use all these multiple age  $\times$  skills observations, but in order to reduce the dimension of the control variables space at each age we replace the original maths and reading scores with the principal component. In all cases the principal component explains no less than 90% of the total variance. For non-cognitive characteristics things are a bit more complex. Both at age 7 and age 11 there are 12 scores measuring features such as depression, anxiety and hostility, as reported by teachers in schools. (No score is available at age 16.) In order to keep our problem computationally tractable, we do a factor analysis of the non-cognitive scores using the iterated principal factor method. Out of 12 scores, only two eigenvalues are larger than 1, with the third being sensibly smaller. Therefore, at each age point, we retain only two factors. In Table V we show the rotated loading factors. The first factor captures the skill to relate to other individuals, either adults or other children. The

second factor captures emotional problems. There are no large differences between age 7 and 11. The final factors are obtained using the regression method. Since all the original scores indicate deficiencies rather than skills, we pre-multiply each factor by  $-1$ . Therefore, the larger the factor, the more (less) likely was the cohort member to have some of these skills (deficiencies).

Finally, since fathers' age was recorded only in the original 1958 survey, we restrict our sample to those cohort members living with a biological father during their childhood. Summing up, we use 11 covariates: namely, three dummies for educational attainment; three continuous variables capturing cognitive skills at 7, 11 and 16 years of age; four continuous variables capturing non-cognitive skills (seen as lack of 'social' and 'emotional' problems) at 7 and 11 years of age; and one discrete variable measuring fathers' age.

#### 5.4. Conditional Social Class Mobility

We now are ready to examine independence and monotonicity conditional on the available set of covariates by fitting a suitable regression model. In order to avoid additional modeling restrictions, apart from those implied by linearity on the scale of the link function, we allow the effect of covariates to be specific to each parameter of the table; more precisely:

- the two logits for the marginal distribution of the father are allowed to depend on father's age, so they require two intercepts and two slopes;
- the two logits for the marginal distribution of the son are allowed to depend on all covariates, so they require two intercepts and  $2 \times 11$  slopes;
- the four log-odds ratios are allowed to depend on all covariates, so they require four intercepts and  $4 \times 11$  slopes.

Parameter estimates are given in Table IX in the Appendix. Since all the covariates are centered, the constant terms have a direct interpretation: they measure the log-odds ratios for the individual with average covariates. For  $\xi$ , education and cognitive skills coefficients have the expected positive sign, meaning that the sons with better education and cognitive skills are more likely to have a

Table V. Loading factors: non-cognitive scores

	Age 7		Age 11	
	F1	F2	F1	F2
Unforthcomingness	-0.0690	0.7279	-0.0807	0.7161
Withdrawal	0.0649	0.6693	0.0908	0.6856
Depression	0.2727	0.6757	0.3496	0.6305
Anxiety for acceptance of adults	0.4061	-0.1096	0.4257	-0.0487
Hostility toward adults	0.6042	0.2118	0.6455	0.1988
Writing off adults and standards	0.4362	0.5162	0.4696	0.4690
Anxiety for acceptance by kids	0.6662	-0.0429	0.6862	-0.0386
Hostility toward children	0.6736	0.0942	0.6593	0.1311
Restlessness	0.5467	0.1447	0.5258	0.1190
Inconsequential behavior	0.7761	0.2020	0.7771	0.1931
Miscellaneous symptoms	0.2354	0.5894	0.3329	0.5408
Miscellaneous nervous symptoms	0.2636	0.1910	0.3057	0.1689

Note: Number of observations = 14,931 (age 7), 14,158 (age 11). Retained factors = 2. Number of parameters = 23.



higher status. None of the non-cognitive skills factors is statistically significant on its own. When interpreting the coefficients for the log-odds ratios ( $\tau$ ), one has to remember that a larger  $\tau$  indicates less mobility. For instance, the higher education coefficient is positive and statistically significant for  $\tau_{11}$  and  $\tau_{12}$ . Intuitively, this result suggests that those individuals with higher education in the medium class were very unlikely to have a father in the low class, and at the same time those individuals with higher education in the low class were unlikely to have a father in the medium class.

We have applied the testing procedures described in Section 4.4 above for testing monotonicity and equality of opportunity on the table as a whole and also by restricting the constraints to adjacent pairs of rows. The results are displayed in Table VI and indicate that, even when we condition to the large set of available covariates, there is overwhelming evidence that the equality of opportunity hypothesis cannot hold. On the other hand, the evidence against the monotonicity hypothesis is so weak and entirely compatible with the random variability implied by the conditional multinomial model.

Table VI. Test results: social mobility

Rows	$\mathcal{H}_0/\mathcal{H}_1$			$\mathcal{H}_1/\mathcal{H}_2$		
	$T_{12}$	d.f.	$p$ -value	$T_{01}$	d.f.	$p$ -value
1–2	30.42	24	0.0003	10.73	24	0.8497
2–3	41.07	24	0.0000	11.38	24	0.8149
All	80.79	48	0.0000	22.48	48	0.9436

### 5.5. Wage Mobility

For completeness, in this section we replicate the analysis using the wage mobility table illustrated in Table VII. As we discussed above, the sample size is much smaller. As in the case of social class, we group the individuals into three categories based on their wage percentile. Unfortunately, given that the original father's wage variable was coded into 12 bands, it is not possible to exactly partition it into three terciles. The chi-square statistic is smaller than for social class (compare Table III) though we still reject independence.

In the Table X in the Appendix we show the estimated coefficients and standard errors for the unrestricted model. A glance at the table reveals that the effect of covariates on the marginal distributions and the association parameters seems broadly comparable to the case of social mobility: the main difference is that for wage mobility the intercepts of the log-odds ratios  $\tau_{ij}$  ratios are generally smaller (with the exception of  $\tau_{21}$ ), indicating a lower level of (conditional) association between  $X$  and  $Y$ . Table VIII shows the results of the hypothesis tests. Interestingly,

Table VII. Wage class mobility table

Father/son class	1	2	3	
1	14.04	10.78	8.79	33.61
2	12.41	10.51	10.42	33.33
3	7.61	11.32	14.13	33.06
	34.06	32.61	33.33	100.00

Note: 1104 observations. Chi-square (4) = 37.03.

Table VIII. Test results: wage mobility

Rows	$\mathcal{H}_0/\mathcal{H}_1$			$\mathcal{H}_1/\mathcal{H}_2$		
	$T_{01}$	d.f.	$p$ -value	$T_{12}$	d.f.	$p$ -value
1–2	8.96	24	0.3389	14.70	24	0.5554
2–3	29.98	24	0.0005	25.91	24	0.0779
All	37.85	48	0.0008	39.80	48	0.2991

note that, when comparing the first two rows of the mobility table for stochastic dominance, we now cannot reject conditional independence. Moreover, as argued above, since three out of four intercepts of the log-odds ratios (the log-odds for the average individual) are smaller than those obtained in the social class case, we may be led to conclude that there seems to be less father-to-son dependence in wages than in social classes. This difference might be due to measurement error or temporary shocks affecting wages more than social class. However, we cannot rule out that sample selection might also be driving this difference. Nevertheless, overall we still reject independence, while we cannot reject stochastic monotonicity.

## 6. NONPARAMETRIC CONDITIONAL MONOTONICITY TEST

As discussed in Section 4.1 above, as a side check we have also tried to test conditional monotonicity with a nonparametric approach; that is, without imposing any restriction on the dependence of  $\lambda(\mathbf{z})$  on  $\mathbf{z}$ . First, we tried to approximate all covariates which are either continuous or take many different values with a limited number of discrete categories.

Even with a rather crude approximation based on two categories for father's age, three for cognitive and non-cognitive abilities, which implies a total of  $2 \times 3 \times 3 \times 4 = 72$  different covariate configurations, there are 63 non-empty social class mobility tables but only 36 of these have a pattern of observed 0's which allows maximum likelihood estimation.<sup>9</sup>

A reasonable strategy is the following: for each sparse table search for a non-sparse one which has approximately similar conditional distributions by row and merge the first into the second. This approach produces 29 conditional mobility tables. We have nevertheless tested  $\mathcal{H}_0$  and  $\mathcal{H}_1$  for this set of tables as a whole. The  $T_{12}$  test equals 33.23 with  $p$ -value equal to 0.8324, while  $T_{01} = 187.13$  and a  $p$ -value of the order  $10^{-5}$ , indicating that conditional stochastic monotonicity cannot be rejected, but there is overwhelming evidence against equality of opportunity.

## 7. CONCLUSIONS

The aim of this paper was to test for stochastic monotonicity in intergenerational socio-economic mobility tables. To do so we apply and extend the methodology discussed in Dardanoni and Forcina (1998) and Bartolucci *et al.* (2001). We first test for unconditional stochastic monotonicity using a set of 149 intergenerational mobility tables in 35 different countries, where it emerges that monotonicity cannot be rejected in hardly any table. We then explain how a number of controls such

<sup>9</sup> Recall that the maximum likelihood estimate of the local global odds ratio is undefined whenever the estimated probabilities that appear on the numerator and the denominator are both 0.

as education, cognitive and non-cognitive skills can be used to investigate whether monotonicity still holds after conditioning on these factors. In the economics literature, no previous work on intergenerational mobility tables has dealt with continuous controls. Since current research on mobility is focusing on the determinants of dependence in socio-economic status between parents and offspring, conditioning on discrete and continuous covariates is increasingly important.

To apply our test of conditional monotonicity we use the NCDS, a UK cohort data with information on the socio-economic status of the cohort members and their parents, and individuals' educational qualifications, cognitive and non-cognitive skills. Our tests show evidence of stochastic monotonicity, both unconditionally and conditionally. While it is not surprising that the unconditional joint distribution exhibits monotonicity, it is interesting to find that such a strong form of dependence exists even conditional on educational achievement and cognitive and non-cognitive skills. This result reinforces the findings of Solon (1999), Bowles and Gintis (2002), Restuccia and Urrutia (2004), Dardanoni *et al.* (2006) and Blanden *et al.* (2007) indicating that part of the mechanism linking parents' and offspring's socio-economic status is still a black box. Finally, we observe only minor differences between social and wage class tables; if anything, we find that there seems to be more dependence when using social rather than wage class.

#### ACKNOWLEDGEMENTS

We thank two referees for detailed and helpful comments which have significantly helped to improve the paper. Valentino Dardanoni thanks MIUR and Università di Palermo for financial support. Antonio Forcina thanks the Department of Economics, Finance and Statistics of the University of Perugia for hospitality.

#### REFERENCES

- Agresti A. 2002. *Categorical Data Analysis*. Wiley: Hoboken, NJ.
- Aitchison J, Silvey SD. 1958. Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics* **29**: 813–828.
- Bartolucci F, Forcina A, Dardanoni V. 2001. Positive quadrant dependence and marginal modeling in two-way tables with ordered margins. *Journal of the American Statistical Association* **96**: 1497–1505.
- Bartolucci F, Colombi R, Forcina A. 2007. An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statistica Sinica* **17**: 691–711.
- Becker GS, Tomes N. 1979. An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy* **87**: 1153–1189.
- Becker GS, Tomes N. 1986. Human capital and the rise and fall of families. *Journal of Labor Economics* **4**: S1–S39.
- Bergsma WP, Rudas T. 2002. Marginal models for categorical data. *Annals of Statistics* **30**: 140–159.
- Bjorklund A, Lindahl M, Plug E. 2006. The origins of intergenerational associations: lessons from Swedish adoption data. *Quarterly Journal of Economics* **121**: 999–1028.
- Blanden J, Gregg P, Macmillan L. 2007. Accounting for intergenerational income persistence: noncognitive skills, ability and education. *Economic Journal* **117**: C43–C60.
- Bourguignon F, Ferreira FHG, Menendez M. 2003. *Inequality of outcomes and inequality of opportunities in Brazil*. DELTA Working Paper 2003-24, Department and Laboratory of Applied and Theoretical Economics, Paris.
- Bowles S, Gintis H. 2002. The inheritance of inequality. *Journal of Economic Perspectives* **16**: 3–30.
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement Error in Nonlinear Models*. Chapman & Hall: London.

- Colombi R, Forcina A. 2001. Marginal regression models for the analysis of positive association. *Biometrika* **88**: 1007–1019.
- Conlisk J. 1990. Monotone mobility matrices. *Journal of Mathematical Sociology* **15**: 173–191.
- Corak M. 2004. *Generational Income Mobility in North America and Europe*. Cambridge University Press: Cambridge, UK.
- Cunha F, Heckman JJ. 2007. The technology of skill formation. *American Economic Review* **97**: 31–47.
- Dardanoni V. 1993. Measuring social mobility. *Journal of Economic Theory* **61**: 372–394.
- Dardanoni V. 1995. Income distribution dynamics: monotone Markov chains make light work. *Social Choice and Welfare* **12**: 181–192.
- Dardanoni V, Forcina A. 1998. A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association* **93**: 1112–1123.
- Dardanoni V, Forcina A. 1999. Inference for Lorenz curve orderings. *Econometrics Journal* **2**: 49–75.
- Dardanoni V, Forcina A. 2008. Multivariate ordered regression. Mimeo University of Perugia: <http://www.stat.unipg.it//forcina/English.html>.
- Dardanoni V, Fields GS, Roemer JE, Puerta MLS. 2006. How demanding should equality of opportunity be, and how much have we achieved? In *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, Morgan SL, Grusky DB, Fields GS. Stanford University Press: Chicago, IL; 59–82.
- Dearden L, Machin S, Reed H. 1997. Intergenerational mobility in Britain. *Economic Journal* **107**: 47–66.
- Douglas R, Fienberg SE, Lee MT, Sampson AR, Whitaker LR. 1990. Positive dependence concepts for ordinal contingency tables. In *Topics in Statistical Dependence*, Block HW, Sampson AR, Sanits TH. Hayward, CA; 189–202.
- Dykstra R, Kocher S, Robertson T. 1991. Statistical inference for uniform stochastic ordering in several populations. *Annals of Statistics* **19**: 870–888.
- Erikson R, Goldthorpe JH. 2002. Intergenerational inequality: a sociological perspective. *Journal of Economic Perspectives* **16**: 31–44.
- Fleurbaey M. 2008. *Fairness, Responsibility, and Welfare*. Oxford University Press: Oxford.
- Formby JP, Smith WJ, Zheng B. 2004. Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics* **120**: 181–205.
- Ganzeboom HBG, Luijkx R, Treiman DJ. 1989. Intergenerational class mobility in comparative perspective. *Research in Social Stratification and Mobility* **8**: 3–79.
- Grawe ND, Mulligan CB. 2002. Economic interpretations of intergenerational correlations. *Journal of Economic Perspectives* **16**: 45–58.
- Han S, Mulligan CB. 2001. Human capital, heterogeneity and estimated degrees of intergenerational mobility. *Economic Journal* **111**: 207–243.
- Heckman JJ, Stixrud J, Urzua S. 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* **24**: 411–482.
- Keilson J, Kester A. 1977. Monotone matrices and monotone Markov chains. *Stochastic Processes and their Applications* **5**: 231–241.
- Lee S, Linton O, Whang Y-J. 2009. Testing for stochastic monotonicity. *Econometrica* **77**: 585–602.
- Lefranc A, Pistoletti N, Trannoy A. 2006. *Inequality of opportunities vs. inequality of outcomes: are Western societies all alike?* Working Paper 54, ECINEQ, Society for the Study of Economic Inequality, Palma de Mallorca.
- Lehmann E. 1955. Ordered families of distributions. *Annals of Mathematical Statistics* **26**: 399–419.
- Mulligan CB. 1999. Galton versus the human capital approach to inheritance. *Journal of Political Economy* **107**: S184–S224.
- Neuhaus JM. 1999. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika* **86**: 843–855.
- Peragine V. 2004. Ranking income distributions according to equality of opportunity. *Journal of Economic Inequality* **2**: 11–30.
- Prais SJ. 1955. Measuring social mobility. *Journal of the Royal Statistical Society, Series A* **118**: 56–66.
- Putterman L, Roemer JE, Silvestre J. 1998. Does egalitarianism have a future? *Journal of Economic Literature* **36**: 861–902.
- Restuccia D, Urrutia C. 2004. Intergenerational persistence of earnings: the role of early and college education. *American Economic Review* **94**: 1354–1378.

- Robertson T, Wright FT. 1982. Bounds on mixtures of distributions arising in order restricted inference. *Annals of Statistics* **10**: 302–306.
- Robertson T, Wright F, Dykstra R. 1988. *Order Restricted Statistical Inference*. Wiley: New York.
- Roemer JE. 2000. Equality of opportunity. In *Meritocracy and Economic Inequality*, Arrow K, Bowles S, Durlauf S (eds). Princeton University Press: Princeton, NJ; 17–32.
- Schrijver A. 1986. *Theory of Linear and Integer Programming*. Wiley: New York.
- Shapiro A. 1988. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* **56**: 49–62.
- Silvapulle MJ, Sen PK. 2005. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley-Interscience: Hoboken, NJ.
- Solon G. 1999. Intergenerational mobility in the labor market. In *Handbook of Labor Economics*, Vol. 3, Ashenfelter O, Card D (eds). Elsevier: Amsterdam; 1761–1800.
- Wolak FA. 1991. The local nature of hypotheses tests involving inequality constraints. *Econometrica* **59**: 981–995.
- Zimmerman DJ. 1992. Regression toward mediocrity in economic stature. *American Economic Review* **82**: 409–429.

## APPENDIX

Table IX. Unrestricted model: social class (1974)

Parameter	Marginal							
	$\rho_1$		$\rho_2$		$\xi_1$		$\xi_2$	
Constant	1.6352	(26.4510)	-0.9620	(-18.9698)	2.1078	(24.1718)	-0.2674	(-5.2418)
Father's age	-0.0338	(-3.5990)	-0.0003	(-0.0396)	-0.0290	(-2.8677)	-0.0004	(-0.0438)
O-level	—	(—)	—	(—)	0.5343	(2.9851)	0.2329	(1.8124)
A-level	—	(—)	—	(—)	1.6015	(3.0112)	1.0127	(4.7179)
High. educ.	—	(—)	—	(—)	0.7777	(2.2970)	1.3276	(6.5289)
C. skills 7	—	(—)	—	(—)	-0.0798	(-1.0457)	-0.0537	(-0.8778)
C. skills 11	—	(—)	—	(—)	0.1519	(1.4253)	-0.0109	(-0.1395)
C. skills 16	—	(—)	—	(—)	0.4170	(3.9301)	0.5326	(6.1489)
Nc. skills 7 (1st)	—	(—)	—	(—)	-0.0127	(-0.1694)	-0.0089	(-0.1438)
Nc. skills 7 (2nd)	—	(—)	—	(—)	0.0371	(0.4785)	0.0749	(1.0838)
Nc. skills 11 (1st)	—	(—)	—	(—)	-0.0309	(-0.4194)	0.0038	(0.0592)
Nc. skills 11 (2nd)	—	(—)	—	(—)	0.1022	(1.4225)	0.0955	(1.4514)

Parameter	Odds							
	$\tau_{11}$		$\tau_{12}$		$\tau_{21}$		$\tau_{22}$	
Constant	0.7785	(3.4852)	0.5249	(3.0082)	0.4118	(1.5316)	0.6935	(5.7366)
Father's age	0.0255	(0.9328)	0.0517	(2.0670)	-0.0878	(-3.1548)	-0.0375	(-1.9414)
O-level	-0.4090	(-0.8632)	-0.0218	(-0.0527)	0.4563	(0.9364)	0.1094	(0.3608)
A-level	0.9018	(0.6769)	0.0666	(0.1026)	-0.1924	(-0.1311)	-0.1019	(-0.1968)
High. educ.	1.7188	(1.9630)	1.7835	(2.9363)	0.8091	(0.6344)	-0.1658	(-0.3114)
C. skills 7	0.1127	(0.5435)	0.0896	(0.4560)	-0.2430	(-1.1552)	0.0658	(0.4555)
C. skills 11	-0.0001	(-0.0004)	0.2169	(0.9342)	-0.2386	(-0.8128)	-0.1569	(-0.8348)
C. skills 16	-0.0305	(-0.1055)	-0.8564	(-2.9761)	-0.0636	(-0.2181)	-0.1141	(-0.5569)
Nc. skills 7 (1st)	-0.1527	(-0.7321)	0.0168	(0.0837)	0.4423	(2.1842)	0.1249	(0.8676)
Nc. skills 7 (2nd)	0.2082	(0.8959)	0.0596	(0.2756)	0.4707	(2.2672)	0.2638	(1.6378)
Nc. skills 11 (1st)	0.1114	(0.5616)	-0.2076	(-0.9010)	-0.5944	(-2.4360)	0.0147	(0.0987)
Nc. skills 11 (2nd)	0.0560	(0.2812)	0.1678	(0.8019)	-0.0958	(-0.5054)	-0.0340	(-0.2208)

Note: Columns 1 and 2 correspond to the father's marginals. *t*-ratios in parenthesis.

Table X. Unrestricted model: wage class (1974)

Parameter	Marginal							
	$\rho_1$		$\rho_2$		$\xi_1$		$\xi_2$	
Constant	0.7025	(10.8281)	-0.7153	(-11.0466)	0.7790	(11.3201)	-0.7829	(-11.2551)
Father's age	-0.0604	(-5.3298)	-0.0491	(-4.1184)	0.0103	(0.8803)	-0.0031	(-0.2556)
O-level	—	(—)	—	(—)	0.1923	(1.1480)	0.0386	(0.2096)
A-level	—	(—)	—	(—)	0.6051	(2.0390)	0.5021	(1.9052)
High. educ.	—	(—)	—	(—)	0.7753	(2.7621)	0.7313	(2.9104)
C. skills 7	—	(—)	—	(—)	0.0499	(0.6402)	0.1655	(1.9331)
C. skills 11	—	(—)	—	(—)	0.0086	(0.0813)	0.0760	(0.7121)
C. skills 16	—	(—)	—	(—)	0.3645	(3.2359)	0.2502	(2.0866)
Nc. skills 7 (1st)	—	(—)	—	(—)	0.0552	(0.6698)	0.1992	(2.0448)
Nc. skills 7 (2nd)	—	(—)	—	(—)	-0.0268	(-0.3110)	-0.1316	(-1.4177)

Table X. (Continued)

Parameter	Marginal							
	$\rho_1$		$\rho_2$		$\xi_1$		$\xi_2$	
Nc. skills 11 (1st)	—	(-)	—	(-)	-0.0434	(-0.5449)	-0.1690	(-1.9518)
Nc. skills 11 (2nd)	—	(-)	—	(-)	0.2199	(2.7225)	0.3527	(3.5865)
Parameter	Odds							
	$\tau_{11}$		$\tau_{12}$		$\tau_{21}$		$\tau_{22}$	
Constant	0.1589	(0.8712)	0.1128	(0.5894)	0.5186	(2.6211)	0.3566	(1.9529)
Father's age	-0.0529	(-1.7568)	-0.0502	(-1.5437)	0.0973	(2.8043)	0.0001	(0.0024)
O-level	0.4536	(1.0696)	0.6547	(1.3205)	0.3801	(0.8438)	-0.4857	(-1.0260)
A-level	2.3471	(2.4014)	1.7889	(2.5130)	-2.8069	(-2.8402)	-2.1742	(-3.0755)
High. educ.	0.6454	(0.9285)	0.6428	(0.9749)	-0.2310	(-0.3092)	-0.3686	(-0.5742)
C. skills 7	0.4129	(2.0222)	0.2796	(1.1973)	-0.1209	(-0.5801)	-0.3467	(-1.5591)
C. skills 11	-0.5185	(-1.8872)	-0.2244	(-0.7904)	0.6906	(2.3944)	0.1713	(0.6224)
C. skills 16	0.1640	(0.5558)	0.1466	(0.4545)	-0.5773	(-1.8904)	-0.2207	(-0.7104)
Nc. skills 7 (1st)	-0.0837	(-0.3861)	-0.1728	(-0.6398)	0.0475	(0.2135)	0.2570	(1.0288)
Nc. skills 7 (2nd)	-0.2386	(-1.0826)	-0.0087	(-0.0363)	0.0420	(0.1831)	0.4537	(1.8804)
Nc. skills 11 (1st)	-0.1253	(-0.5918)	-0.4016	(-1.5518)	-0.4297	(-1.7630)	-0.0807	(-0.3501)
Nc. skills 11 (2nd)	0.0871	(0.4103)	-0.1781	(-0.6943)	-0.1454	(-0.6697)	0.3068	(1.1944)

Note: Columns 1 and 2 correspond to the father's marginals. *t*-ratios in parenthesis.