A joint Initiative of Ludwig-Maximilians-Universität and Ifo Institute for Economic Research

CESifo / PEPG Conference on

Schooling and Human Capital Formation in the Global Economy

CESifo Conference Centre, Munich
3 - 4 September 2004

Intergenerational Mobility in Europe: Evidence from ECHP

Simona Comi

CESifo
Poschingerstr. 5, 81679 Munich, Germany
Phone: +49 (89) 9224 1410 - Fax: +49 (89) 9224 1409
E-mail: office@CESifo.de
Internet: http://www.cesifo.de

Intergenerational mobility in Europe: evidence from ECHP*

13 July 2004

Simona Comi*

Università degli studi di Milano

email: comi@statistica.unimib.it

Abstract

In this paper I provide new evidence on cross-country comparison of intergenerational mobility using the first five waves of the European Community Household Panel focusing on two alternative dimensions of mobility: income and education. I estimate intergeneration earnings elasticity for sons and daughters father pairs and also some synthetic indexes from intergenerational educational transition matrices in 12 European countries. For the first time it can be possible to rank European countries according to their degree of intergenerational income mobility. It turns out that that Italy is the most immobile country in Europe, together with Portugal and Greece when considering sons and with Germany when considering daughters.
I find also that fathers behave differently in passing income and education to children accordingly to their sex and find that when a strong link between father and daughter's income is observed tipically the relation between their level of education is weak, while the reverse can be stated for sons .The paper investigates also how institutional and economic factors can affect mobility and shows that surprisingly, no relation between the income elasticity and earnings returns to human capital emerges that income elasticity seems to be negatively related to public expenditure in tertiary education positively related to the strictness of the employment protection law. Educational mobility across generations seems to be affected by the performance of the education system measured by the proportion of students fall below given benchmarks of educational achievement, it is not affected the percentage of students enrolled in private schools

1.Introduction

During the second half of the last century, social science literature investigated the process that explains why some individuals achieve success in young adulthood while others do not. Success was typically measured by schooling attainment, occupation or earnings (income) level. Sociologists were the first to study this topics and economists came later. As it is synthetically explained in Liebovitz (1974) and later in Haveman and Wolfe (1995), ability is passed to children via heredity (genetic endowment). Furthermore, parent's ability and educational attainment (via quantity and quality of time and money devoted to children) jointly with children's ability determine their educational attainment. The latter (directly and via post-school investment), together with ability and family income, affects children's earnings and income. In this paper I will concentrate on two of the possible intergenerational correlations: earnings and education correlations.

Economists have mainly concentrated on the relation between fathers and off-springs' permanent income. Starting from the milestone of Becker and Thomes in 1979, economists have tryed to measure the link between an individual socio-economic position and his father's. The interest in the transmission of economic status from one generation to another is generally motivated by the wish to determine the degree of equality of opportunity of a country. From then onward, a large part of the literature has looked for an appropriate method to measure mobility[1]. As regards income, the most used measure of mobility is the regression coefficient relating a son's log earnings to his father's. A high value indicates a very high persistence of the economic status, because an individual's position in the earnings distribution is largely a reflection of his father position in his own distribution. A low value indicates a very mobile society in which an individual's socio-economic position does not depend on his father's one. Data availability is a crucial key, in fact information about the income of the two generations is needed, and typically long panel data or cross-section with retrospective information about parents' income are used. Along the years, this literature has highlighted some crucial methodological issues. They are discussed in detail in section 2. The study of

* I am grateful to Elena Bardasi, Lorenzo Cappellari, Daniele Checchi, Stefania Damonte, Marco Francesconi, Claudio Lucifora, Cheti Nicoletti and seminar participants at the Institute for Social and Economic Research for their helpful and precious comments. This research was conducted while the author was visiting the Institute for Social and Economic Research (University of Essex) with a Marie Curie fellowship.

[1] For a review see Solon (1999)

intergenerational correlations of educational attainment is done in many different ways, such as transition matrices (or some synthetic measures) and probit (or ordered probit) estimations.

Although a large body of the economic literature has studied the correlation between father and son socio-economic status, only fewer and more recent works have analysed the differences existing in intergenerational mobility across countries or times. The evolution over time is studied by Blanden et al (2001), Mayers and Loopo (2001) and Ermisch and Francesconi (2002), the former finds that mobility decreased in Britain during the last 30 years while the other two papers find that British and American younger cohorts experienced greater mobility than older ones, and conclude that further investigation seems to be necessary.

Few studies analyse cross-country differences in intergenerational mobility. To exploit this issue, data requirements are even more stringent: in fact, similar information on both father and son income is needed for each country and, as I will explain in detail later, given that intergenerational income elasticity depends strongly on sample selection rules[2] applied to data-set, we need similarly selected sample for each country. Some studies draw the information for the sons from cross-sections, which are less sensible to sample selection rules than panel data, and use retrospective information on parental background for the fathers. Among the parental characteristics reported by the sons there are the occupation and the level of education but hardly ever the income. This information can be used, like in Checchi and Dardanoni (2002), to construct a similar index of socio-economic positions for both parents and children or, like in Bjorklund and Jantti (1997,) to infer income from a sample of older man (synthetic father) and estimate intergenerational correlation. In particular they compare Sweden to the US using a Two Sample Instrumental Variable (TSIV) methods and conclude that intergenerational mobility is higher in Sweden. Finally, also Ichino, Checchi and Rustichini (1999) compare Italy and US using parental background characteristics to build a variety of indicators, and conclude that Italy is less mobile than US.

Another and completely different way to solve the data requirement problem in cross-country comparisons can be found in Couch and Dunn (1998). Using two very similar longitudinal data-sets, the GSOEP for Germany and the PSID for US, they focus on contemporary (on the same years window) observations of parents and children, apply the same selection rules and same methodology and conclude that there

[2] In particular, it decreases with sons' age and it is sensible to the inclusion/exclusion of zero earnings years (i.e. years of unemployment). See Jenkins (1987) Couch and Lillard (1998)

is a remarkable similarity across the two countries. A similar approach is used by Grawe (2001). Using a quantile regression approach, he produces a pair-wise comparison between US and many other countries for which he can find similar longitudinal data-set, such as Germany, Canada, Uk Malaysia, Ecuador, Nepal, Pakistan and Peru.

In this paper I provide a new evidence on cross-country comparison using for the first time the same dataset, the European Community Household Panel, for 12 countries. Although this data-set produces estimation that suffer of many potential biases, such as life cycle bias due to the young age of children, if the distortions are similar across countries, then the results can be useful and produce a better understanding of the forces that shape different societies. The main advantage of these data is that the same "community" questionnaire is adopted by the national data collection units in each participating country, which increases comparability among the countries. And in a cross-country comparison view we can try to relate intergenerational income and education mobility to the national educational systems. For instance, we can see for example if differences in the predominance of public schools and free of charge higher education affect the degree of income and educational mobility.

This paper proceeds as follows. In section 2 I briefly discuss the econometric issues related to the measure of earnings intergenerational elasticity, in section 3 I describe the data and the sample I use and section 4 contains the results. Section 5 is devoted to the measure of educational mobility and finally in section 6 I summarize the results and try to find some correlations between the degree of earnings and educational intergenerational mobility and characteristics and institutional settings of each country.

2. Earnings elasticity: estimation methods

The economic literature on intergenerational mobility has faced along the years many estimation problems. Earlier studies simply estimated

$$Y_{1i} = \beta Y_{0i} + \varepsilon_i \qquad (1)$$

where Y_{1i} is a measure of the permanent economic status of the son and Y_{0i} is the corresponding measure for the father. Early works on this topic used a single year income as a proxy for permanent income but later it has been shown that the estimated β was downward biased. In particular, if parents and children are observed at different points in their life-cycle then the age effect of parents and children should be removed (two possibility: putting age and age square of both children and fathers in (1) or regressing earnings on age and age2 and using residuals). And then, consider the following expressions:

$$Y_{0is} = y_{0i} + v_{0is} \qquad (2)$$

$$y_{1it} = y_{1i} + v_{1it} \qquad (3)$$

In equation (2) father earnings in year s is composed by a permanent component that reflects the true long-term earnings capacity (y_{0i}), a component that captures both transitory shock that might affect that particular year earnings and error due simply to inaccurate report of earnings (v_{0is}). Equation (3) does the same for son's earnings. If we are interested in the relation (1) but we estimate it using single-year measure the coefficient will be biased downward by the attenuation factor:

$$p\lim \hat{\beta} = \beta\left(\frac{\sigma_{y_0}^2}{\sigma_{y_0}^2 + \sigma_{v0}^2}\right) \qquad (4)$$

To avoid this bias Solon (1992) proposes to use an average of father earnings (typically 5 years) because this will reduce (but not eliminate completely) the biases generated by both transitory shock and measurement error (σ_{v0}).[3]

Jenkins (1987) and Grawe (2001) show that the estimation may be sensitive to life cycle biases even after controlling for measurement error and age. In particular, Grawe (2001) points out that since income variance grows over the life cycle, estimates of income persistence based on data from mature fathers will naturally be lower than those based on young fathers and finds that a great part of the differences in estimated intergenerational correlations for the US is explained by the differences in the age of father and sons at the point of measurement. If the estimates are to be compared, the selection criteria must control for age of the father in all samples, and ECHP naturally allows to control for this problem. Reville (1995) shows that the intergenerational correlation decreases with son's age, and according to Solon (2002) this happens because the kind of measurement error in son's early earnings as a proxy for the permanent earnings is not of the same kind seen before. In fact, those young sons which will have the highest level of socio-economic status will have more rapid earnings growth than the ones that will be poorer in the future. This kind of measurement error is "mean reverting" and it is negatively correlated with long-run earnings. Bound et al. (1994) show that mean reverting measurement error in a regression dependent variable

[3] For a discussion see Mazumder (2001)

compresses its variation and consequently leads to a tendency to underestimate the magnitude of the regression slope coefficient. Solon (2002) concludes that averaging in this case will even worsen the estimation bias because in averaging I could use also the very early earnings of my son's samples. Given these considerations, with ECHP I expect to find lower levels of intergenerational correlation than other study, because sons are observed in their early life and father in their later years of labour market experience.

Finally, my results can be affected by sample selection bias because to match children and father in ECHP they should cohabit at least in one wave. If in southern country children stay in their parents house up to their thirty (see Iacouvu , 2002) the leaving age is far much lower in northern and nordic countries. To take account of this potential bias I correct the estimation using the method proposed by Woodrige (1999)

As estimation strategy, I consider the model introduced and incorporate age profile in equations (2) and (3):

$$Y_{0is} = y_{0i} + \alpha_0 + \gamma_0 A_{0is} + \phi_0 A^2_{0is} + v_{0is} \quad (5)$$

$$y_{1it} = y_{1i} + \alpha_1 + \gamma_1 A_{1it} + \phi_1 A^2_{1it} + v_{1it} \quad (6)$$

where A_{0is} is the age of the father from family i in year s and A_{1it} is the age of son from family i in year t. Solving (5) and (6) for y_{0i} and y_{1i} and substituting them into equation (1), I get:

$$Y_{1it} = (\alpha_1 - \beta\,\alpha_0) + \beta\, Y_{0is} + \gamma_1 A_{1it} + \phi_1 A^2_{1it} - \gamma_0 \beta\, A_{0is} -$$
$$- \phi_0 \beta\, A^2_{0is} - \beta\, \upsilon_{0is} + v_{1it} + \varepsilon_i \quad (7)$$

Equation (7) relates son's observed earnings in year t to father's observed earnings in year s and to age controls for both father and son. As we have seen, the estimation of (7) generates a biased $\hat{\beta}$ and that's why, as Solon (1992) suggests, in addition to (7) I also estimate:

$$Y_{1it} = (\alpha_1 - \beta\,\alpha_0) + \beta\, \overline{Y}_{0i} + \gamma_1 A_{1it} + \phi_1 A^2_{1it} - \gamma_0 \beta\, \overline{A}_{0i} -$$
$$- \phi_0 \beta\, \overline{A}^2_{0i} - \beta\, \upsilon_{0is} + v_{1it} + \varepsilon \quad (8)$$

where, for any variable m_{0is}, $\overline{m}_{0i} = \sum_{j=s}^{s+T} m_{0ij} / T$

In equation (8) the averages over T years of father age and earnings are used instead of their single-year measures. Solon (1989) shows also that the bias decreases as T increases.

Finally, in order to correct the sample selection bias, for each wave I estimate the probability to be matched with the father in that wave *t*, s_{it}:

$$P(s_{it}=1|X_{it})=f(x_{it}d_t) \tag{9}$$

where X_{it} is a matrix of personal characteristic including age, gender, a tertiary education dummy, a dummy to indicate if the children is a student, a dummy indicating wether they work and a set of regional dummies[4]. From this first stage, I compute the inverse Mill's ratio λ_{it} that will be added to the final specifications to be estimated with random effect:

$$Y_{1it}= (\alpha_1- \beta\, \alpha_0)+ \beta\, Y_{0is}+ \gamma_1 A_{1it} +\phi_1 A^2_{1it} - \gamma_0\beta A_{0is} - \\ -\phi_0\beta A^2_{0is} - \beta\, \upsilon_{0is}+\rho\lambda_{it}+ v_{1it} + \varepsilon_i \tag{10}$$

pAnd with father averages:

$$Y_{1it}= (\alpha_1- \beta\, \alpha_0)+ \beta\, \overline{Y}_{0i}+ \gamma_1 A_{1it} +\phi_1 A^2_{1it} - \gamma_0\beta\, \overline{A}_{0i} - \\ - \phi_0\beta\, \overline{A}^2_{0i} - \beta\, \upsilon_{0is}+ v_{1it} +\rho\lambda_{it} + \varepsilon_i \tag{11}$$

If the variance in log earnings is the same for both generation then the intergenerational elasticity obtained, $\hat{\beta}$¸ is also the intergenerational correlation, which is the measure of intergenerational mobility mainly used in sociology literature. The two measures are roughly comparable even if the variance in income differs substantially across generations as shown by Solon (1992). Bowles and Gintis (2001) suggest that the regression coefficient is a preferred measure since it does not conflate changes in cross-sectional inequality with the association in earnings across generation.

3.Earnings elasticity: Data and sample selection

The European Community Household Panel (ECHP) is a large household survey that covers most members countries in the European Union. Rather than trying to harmonise output from national surveys, the European statistical agency (Eurostat) adopts an input oriented approach and uses the same community questionnaire as the

[4] In ECHP for privacy reason some countries (Denmark, The Netherlands, Irelands) do not have Nuts code.

base for the national versions of the survey. The data are collected by the National Collection Units and finally checked by Eurostat (European Community (1991)). A desirable feature of ECHP is that the definitions of and questions on earnings, the reference period and the survey methods are common across countries. This format increases comparability, but does not eliminate all problems, as the interpretation of common questions can vary across countries because of country – specific institutions and history (OECD, 1991). Educational attainment in ECHP is a categorical variable [5] and there's no information on years of schooling but it is still possible to study the way parents' education affects children's one. I use the first 5 waves, from 1994 to 1998[6]. The survey is composed of a household and a personal file, and the same individuals and families are interviewed over time. In the first wave (in 1994) a sample of some 60,500 nationally representative households - i.e. approximately 130,000 adults aged 16 years and over - were interviewed in the 12 Member States. Austria (in 1995) and Finland (1996) have joined the project since then, Sweden remaining the only exception. For the fourth wave of the ECHP, i.e. in 1997, the original ECHP surveys were stopped in three countries, namely Germany, Luxembourg and in the United Kingdom. In these countries, existing national panels were used and comparable data were derived from the GSOEP and BHPS - back from 1994 onwards, and I use these samples. I excluded Luxembourg and Finland from my study because of their small size.

In this paper I consider son –father and daughter- father pairs and allow families to contribute as many father-child pairs to each sample as meet the selection rules. Sons and daughters are matched to their father using the relational file provided in every waves. So I include in the sample every individual that in at least a wave was linked to somebody as a child, aged between 16 and 35 and his/her father (i.e. every male that in at least a wave was coded as parent and that has an age between 35 and 70). I then exclude observations during any year in which the child was enrolled in school or the parent to whom he or she is matched was enrolled in school or retired. Finally I exclude both self-employed and unemployed children and fathers. In calculating averages of earnings across years, I include as many years of valid data as were available for each individual. After exclusions, I have a total of 15011 pairs and in this sample I have fathers and children employed that reported a positive earnings.

[5] ISCED0-2, ISCED 3 –secondary , ISCED 4-7 or tertiary level education
[6] The autor used ECHP while she was visiting the Institute of Social and Economic Research in 2001-2002

The first concern that arises in selecting the samples is that countries included in this paper have different social habits as regards cohabitation with parents and home leaving ages (Soro-Bonmati, 1999 and Iacovou, 2001). This evidence exposes my results to a possible sample selection bias because in Northern and Nordic countries children leave home very early to go to college and never come back while in Southern countries, Italy for example, they stay with their parents till they are 30 and I may have an higher probability to match them. Table 1 shows that the percentage of children with an age between 18 and 35 that can be matched to their father is far more higher in Mediterranean countries: compare the 67% of Italian sons with the 19% of Danish. In column III I report the age at which 50% of the sample are living away from parents home reported in column III from Iacovou (2001): from these figure clearly emerges that in countries where children tend to stay longer in their parents home I am able to match a greater share. That is why I need to correct for sample selection bias, correction that is almost never done in intergenerational income mobility literature.

Table 1:Sample selection, average age and final numbers of pairs.

Country	**Starting sample: matched pairs, wave1**				**Average Age**		**Age at which 50% of the sample are living away from home***		**Final Sample: Number of pairs**		**Expected years in education for 15 years old****
	Sons N	**Sons %**	**Daughters N**	**Daughters %**	**Sons**	**Daug.**	**Men**	**Women**	**Son-father**	**Daughter-father**	
Germany	687	*35*	584	25	21,5	21,5	24.8	21.6	1373	890	4,5
Denmark	166	*19*	159	16	20,4	20,4	21.4	20.3	293	157	3,7
Nether.	433	*31*	316	19	21,1	21,1	23.3	21.2	522	301	2,7
Belgium	189	*25*	301	26	23,8	23,8	25.8	23.8	266	129	6,2
France	456	*27*	719	29	23,6	23,6	24.1	22.2	540	257	6,8
Uk	370	*24*	263	15	21,4	21,4	23.5	21.2	646	523	2,7
Ireland	820	*53*	981	52	22,8	22,8	26.3	25.2	842	631	4,8
Italy	2142	*67*	1726	54	23,5	23,5	29.7	27.1	1158	630	5,8
Greece	967	*55*	759	37	23,5	23,5	28.2	22.9	439	284	6,0
Spain	1907	*60*	1590	5	23,8	23,8	28.4	26.6	1166	686	5,2
Portugal	1067	*58*	858	49	23,1	23,1	28.0	25.2	1265	659	4,8
Austria	700	53	487	38	20,7	20,7	27.2	23.4	792	562	3,9
Total	-	-	-	-	22,4	22,4	-	-	9302	5709	-

*Taken from Iacovou (2001). **Taken from OECD Education at a glance[7].

Given this sample selection problem, also the average age of children and fathers should be different across countries. But the existing differences in age may

[7] OECD calculates the age –specific proportion of young people still in education and then total it to 15-29 years old to yield the expected years in education.

reflect differences in educational system and enrolment rates of countries rather than a self-selection in moving from home before I can match them to their parents. In the last column of table 2 I report the expected years in education for a 15 years old in each country. As my sample only includes sons and daughters already in the labor market, countries in which the youth population tends to stay longer in education and enter the labor market later in their life exhibit an higher average age of children. In Belgium, for example, where the expected years in education are 6,2 children have an average age of 24, while in The Netherkand, where the expected school life of 2,7, they have an average age around 20.

Furthermore, to avoid measurement errors Solon (1992) suggests to take averages over different years of earnings data in order to obtain better estimates of permanent earnings capacity, and to do so, he includes observation only if fathers and sons are continuously employed over the entire period. In this way sons and fathers that report only one year of zero are removed from the sample creating an additional selection bias with an unclear direction (for a discussion see Couch and Lillard, 1998). In fact, it is more common that low income earners become unemployed, and so their exclusion will increase the average income in the sample. And since unemployment is a national phenomenon also a cross-country selection bias will be added if I exclude unemployed. Following Counch and Lillard's procedure, in the next section I will provide results for fathers earning averaged including years of unemployment.

Another possible selection bias is due to the exclusion of self-employed from the sample. Dunn (1996) studies the intergenerational persistence of self-employment and finds that the intergenerational link is strong. But self –employment reported earnings are far more exposed to measurement error than employees' and the earnings variable I use is the monthly gross salary which is not available for self-employed[8] in ECHP. Standard analyses of intergenerational mobility exclude self-employed from the sample, and I will do the same to produce comparable results[9].

The earnings variable I use in all the specification is the current gross monthly earnings which is almost directly collected (not imputed) and is not distorted by the national taxation systems.

[8] Self employed income is the yearly (not monthly) income of the previous year.
[9] Table A1 in appendix shows the proportion of self-employed by country.

4. Earnings elasticity: estimates

Table 2 and 3 present Random Effect estimations of equation (10) and (11) for son-father pairs and daughter-father pairs[10]. In each table I report the estimated regression coefficient $\hat{\beta}$ and its Huber –White standard error (to account for the fact that there are instances in both data sets where more than one child is matched to the same parent) obtained with clustering on individuals (because observations are independent across individual but not necessarily independent within the history of the same individual). The average father income used in column(II) is calculated using all the available non negative years. When a year of unemployment is reported an income of 1 unit is imputed (see Couch and Lillard, 1998)

Table 2: Son-father pairs: $\hat{\beta}$ from random effect estimates of (10) and (11).

Country	Father yearly earnings			Father earnings averaged including years of unemployment		
	$\hat{\beta}$	Sample(1)	Mobility Rank	$\hat{\beta}$	Sample(1)	Mobility Rank
Germany	.15 (.050)	1373	4	.12 (.038)	1509	6
Denmark	-.05 (.094)	293	-	-.033 (.088)	316	-
The Netherlands	-.06 (.069)	522	-	-0.009 (.060)	552	-
Belgium	.23 (.068)	266	8	.11 (.054)	278	5
France	.17 (.042)	540	5	.06 (.046)	567	2
Uk	.09 (.045)	646	3	.09 (.037)	716	4
Ireland	.07 (.036)	842	2	.045 (.028)	992	1
Italy	.27 (.031)	1158	10	.19 (.032)	1259	9
Greece	.24 (.039)	439	9	.12 (.041)	534	7
Spain	.17 (.031)	1166	6	.082 (.026)	1370	3
Portugal	.19 (.024)	1265	7	.14 (.023)	1466	8
Austria	.07 (.044)	792	1	.004 (.044)	836	-

Notes: (1) samples are different because of the exclusion of some outliers from the regressions. Outliers are detected using the Hadi procedure of STATA 7

[10] Selection equations and full estimates are available upon request from the author.