Intergenarational earnings mobility in Spain

María Cervini Plá *

September 30, 2008

(very preliminary and incomplete; please do not quote)

Abstract

This paper analyse the extent and evolution of intergenerational earnings mobility in Spain considering some sample selection problems, like co-residence and employment selection. Since there are no Spanish surveys with information on both children and their fathers' earnings covering a long period, we deal with the co-residences selection problem considering two separate samples: a main sample containing information on offsprings' earnings and a set of occupational and education characteristics of their fathers and a supplemental one with data on the same set of fathers' characteristics and their earnings. The first sample we use is the Spanish sample of the EU-SILC, called ECV (Encuesta de Condiciones de Vida) and the second is the Household Budget Survey of 1980-1981 (Encuesta de Presupuestos Familiares). We combine information from the two samples by using the two-sample two-stage least square estimator described by Arellano and Meghir (1992) and Ridder and Moffit (2006). We find an elasticity of 0.40 for sons and 0.55 for daughters. Furthermore, we find a little decrease when we move to younger cohorts. Intergenerational mobility in Spain is similar to France, lower than the Nordic countries and Britain and higher than the United States. We also take into account the employment selection in the case of daughters adopting a Heckman-type correction methods. Finally, we estimate the elasticity between children's and father's earnings on average and by quantiles. We find that the influence of the father's earnings is greater when we move to the lower tail of the distribution.

Keywords: Intergenerational earnings mobility, two sample two stage least square estimator, Spain.

JEL classification: D31, J31, J62.

*Departament d'Economia Aplicada. Universitat Autònoma de Barcelona. Edificio B, 08193 Bellaterra (Barcelona), Spain. Phone number +3493 581 4575/Fax +3493 581 2292. E-mail: maria.cervini@uab.es

1 Introduction and Motivation

The degree of intergenerational mobility could be thought as an important indicator of the healthiness and success of a society. One important reason for this belief is the judgement that equal opportunity is a desirable characteristic of a good society. In this context, equal opportunity means that children from different families have equal options regarding investments in their human resource and their expected incomes (Behrman and Taubman (1990)).

Intergenerational mobility studies estimate the correlation between socioeconomic status of parents and their offspring. A high correlation would imply that people born in disadvantaged families have a smaller chance to occupy the highest socio-economic positions than people born in privileged families. A zero correlation would imply instead a high degree of mobility and more equal opportunities.

In the analysis of intergenerational mobility, researchers have used several different measures of long-run socio-economics status. Each measure has advantages and disadvantages. There are also several popular methods of examining intergenerational correlations in socio-economic status (e.g., linear regressions, quantile analysis, and transition matrices).

Economist have mainly concentrated on the relation between fathers and offsprings' permanent income, ie, intergenerational elasticity in continuous monetary variables, typically income or earnings, while sociologists use association measures between ordered categorical variables such as social and economic class positions.[1]

In this paper, following the economic approach, we focus on intergenerational mobility measured by the intergenerational elasticity of offsprings' earnings with respect to fathers' earnings.

The main objective of this paper is to study the extent and evolution of intergenerational earnings mobility in Spain considering some sample selection problems, like co-residence and employment selection. At present, there is no information about intergenerational earnings mobility in Spain for children belonging to very distant cohorts from their parents. The absence of previous findings is due to the lack of Spanish surveys with information on both children and their fathers' earnings covering a long

[1] See Solon (1999), Björklund and Jäntti (2000), Bowles and Gintis (2002), Erikson and Godthorpe (2002) for a review.

period. In general, we have information of the offspring when they live with their parents but not in their adult life because the panel are so short to follow the children during their adult life. Thus, the probability of observing offspring living with their parents decreases as the children grow older. This generate a bias in the estimation of intergenerational correlation. Following Nicoletti and Francesconi (2006) we can refer to this sample selection problem as co-residence selection.[2]

How do I deal with this problem? It is possible to estimate consistently intergenerational earnings mobility with the two-sample two-stage least square estimator. Using this estimator it is possible to combine information from two separate samples; a sample of adults (sons and daughters) with observations of their earnings and their parents' characteristics, and a sample of potential parents with observations on earnings and the same characteristics. The latter sample is used to estimate an earnings equation for parents using their characteristics as explanatory variables, while the former is used to estimate an intergenerational earnings equation by replacing the missing parents' earnings with its best linear prediction. In particular my two samples are: the Spanish sample of the EU-SILK, called Survey of Living Conditions (Encuesta de Condiciones de Vida (ECV)) and the Household Budget Survey of 1980-1981 (Encuesta de Presupuestos Familiares (EPF)). In the ECV we observe sons' (respectively daughters') earnings and a set of occupational characteristics of their parents when the children were aged between 12 and 16. This gives us a set of auxiliary variables, such as education dummies, age, occupational sector, which can be used to predict the parents' missing earnings. Since the participation of women in the labour market is particular important since the seventies, for the case of parents I will only consider fathers' earnings. Thus, I estimate the elasticity between children's earnings and fathers' earnings.

In Spain the study of intergenerational mobility has been undertaken mainly by sociologist. For example, Carabañas (1999) studied occupational mobility. From an economic point of view, there are some studies of intragenerational mobility like Cantó (2000), Rodriguez, Salas, and García (2002) and Ayala and Sastre (2002a, 2002b). These studies analyse the probability that one individual could change her level of

[2] Nicoletti and Francesconi (2006) analyze intergenerational mobility using an occupational prestige score. They find that the β coefficient is underestimated when they only consider the pairs of children and parent that are co-resident.

income during her life. The only study for Spain that analyzes intergenerational mobility, as far as I know, is Hugalde (2004). She analyzes the intergenerational income and education mobility using the Household Budget Survey (Encuesta de Presupuestos Familiares) for 1980 and 1990. However she only estimate the elasticity when the child and her father live together. She find an elasticity of income for the year 1990 of 0.44.

We use the two-sample two-stage least square estimation. These method has been already applied to study intergenerational mobility by Björklund and Jäntti (1997) in Sweden, Fortin and Lefebvre (1998) in Canada, by Grawe (2004)) in Ecuador, Nepal, Pakistan and Peru, by Lefranc and Trannoy (2004) in France and by Nicoletti and Ermisch (2007) in Britain. In all those studies, but the last two, the choice of the instrumental variables is dictated by the few variables available. However, in the last two studies a larger set of instrumental variables are available, which gives them a greater degree of freedom in choosing the instrumental variables to predict the missing fathers' earnings.

When we correct for the co-residence selection problem find an elasticity of 0.40 for sons and 0.55 for daughters. Furthermore, we find a slightly lower when we move to younger cohorts. Thus, intergenerational mobility in Spain is similar to France, lower than the Nordic countries and Britain and higher than the United State. Considering the employment selection on women we obtain a lower elasticity on younger daughters. Finally, we estimate the elasticity between children's and father's earnings on average and by quantiles. We find that the influence of the father's earnings is greater when we move to the lower tail of the distribution.

The rest of the paper is organized as follows. In the next section I briefly describe the main sources of earnings transmission. In section 3 I present a theoretical framework that allow us to understand some of the sources of earnings transmission between generations. Section 4 describes how I implement the two-sample two stage least square estimator. In section 5 I describe the data source, the selection sample and the variables used in the empirical analysis. Section 6 reports the results and finally, section 7, conclude with some final remarks.

2 Sources of earnings transmission

Why some children achieve success when they become adults while others do not? Why some children obtain better jobs and higher earnings? Which are the channels of transmission of earnings?

As Nicoletti and Ermisch (2007) point out an important number of institutions affect intergenerational mobility, like educational system, labour market, the family (in particular how it invest in children). Furthermore, public policy affects these institutions and through institutions it also affects the intergenerational mobility.

Here I will summarize the main channels of earnings transmission. Although most of these channels are the same when we analyze the transmission of income, here I will concentrate on the persistence of earnings.

One of the most important channel of the intergenerational earnings transmission derives from education. Educational choices are conditioned by individual unobserved ability (labeled talent), family cultural background, family financial resources, public resources and- more generally- social capital. As Checchi (2006) points out, most of these factors exhibit intertemporal and intergenerational persistence.

Ability is passed to children via heredity (genetic endowment). Ability can influence in the education achieved and through education affects on earnings or can influence direct in the type of job obtained or career done and in this way affect on earnings.

Education achieved also depends on cultural influences. There is a vast empirical evidence about how children of educated parents are more likely to acquire education. This may be partly due to parent imitation (if they see their parents reading a book, they get the idea that reading is a good activity), but in most cases it works through induced educational choices. An educated parent is better aware of the psychological and economic value of education, and therefore puts more pressure on his/her children to achieve more at school. In addition, if the educational system is not homogeneous, an educated parent always has some advantage in collecting information about school quality, and can reorient his/her child's choices towards better opportunities. A strengthening factor derives from marital choices: as long as there is assortative mating (namely, better-educated persons preferring to pair of with other educated persons), the cultural background within a family is made more homogeneous, and the influences received by each parent reinforce one another.

Although it is very difficult separate traits that are genetic from traits that are culturally induced, the empirical evidence obtained from the sample of twins indicates that the relative contribution from genetics to intertemporal persistence is low. Bowles and Gintis (2002) show that measured IQ test score contribute little to earnings, and use this evidence to conclude that their contribution to intergenerational persistence must be low.

Furthermore, if access to education is limited by family financial resources due to liquidity constraints, and acquired education gains access to higher-paid jobs, this opens the door to a poverty trap: poor families are prevented from investing in the education of their children by a lack of resources and the inability to access financial markets, their children remain uneducated and poor, and thus they are unable to invest in their grandchildren either.

Another source of intergenerational earnings persistence emerges from territorial segregation, and is related to family wealth. If residential choice are influenced by the evaluation of local school quality, and school quality affects house prices, then richer families will gain access to better schools by locating closer to them. Better school quality combined with a more homogeneous cultural neighbourhood will yield greater social capital, thus providing a clear advantage to children raised in that environment. Thus, the neighbourhood can influence on earnings through education (better schools) or through social capital (good neighbours allow me to obtain betters jobs).

Another channel is social capital *per se* . Obtain a good job with a good salary depend in some of the cases on the friend and social network than of the curriculum.

3 Theoretical framework

Following Checchi (2006) and Lefranc and Trannoy (2004) I present here a simple model that allow us to understand better some of the sources of intergenerational earnings transmission.

Let us suppose an individual i belonging to generation t which permanent earnings Y_{it} derives from two component ability endowment A_{it}, and human capital (means education E_{it}). If we neglect on-the-job training, education is predetermined with respect to labour market status, and therefore with respect to earnings. If we consider that ability increases labour productivity, we should observe that:

$$Y_{it} = \beta E_{it} + \varepsilon A_{it} + \mu_{1it} \tag{1}$$

Where the relationship between earnings, education and ability is assumed linear for simplicity and μ_{1it} is an i.i.d. error term, capturing the idea of luck in the labour market.

Taking into account the channels of intergenerational earnings transmission described in the previous section, we will consider four potential channels through which one generation may influence the following one. First, if ability is genetically (or mechanically) inherited, we indicate this effect with the α and $t-1$ represent the previous generation, so we have:

$$A_{it} = \delta + \alpha A_{it-1} + \mu_{2it} \tag{2}$$

This effect can be though of as all aspects of earnings determinants that "money can't buy" and at the same time are transmitted from one generation to the next. For example, transmission of IQ, social network or preferences.

Second, education can be determined by the cultural influence of the family (described by the η). Third, if the are liquidity constraints, education is also determined by the family incomes, reducing the optimal investment in education from poor families. The intuition behind this effect is that investment in child's human capital, and more generally child's upbringing, is likely to be constrained by parental resources, in the presence of imperfect capital market. For example Becker and Tormes (1979) assume that children endowment in human capital are chosen by their parents as a result of optimal allocation of the parents' permanent income. Parents' utility depends on parents' own consumption and children's permanent income.

We indicate this channel with the γ and we write

$$E_{it} = \eta E_{it-1} + \gamma Y_{it-1} \tag{3}$$

Thus, education is determined by education and income of the previous generation. But if we substitute E_{it-1} by the expression with one lag successively we can observe that education depend on earnings of the parents, grandparents and previous generations.

Forth, if we consider the possibility that family networking and neighbourhood effects give access to better job opportunities. We indicate this channel with the θ, and we can amend equation 1 by adding a further term:

$$Y_{it} = \beta E_{it} + \varepsilon A_{it} + \theta Y_{it-1} + \mu_{1it} \tag{4}$$

Taking into account all this channels, we can observe that intergenerational persistence is a dynamic system. From an empirical point of view it is not easy to distinguish between alternative explanations of intergenerational persistence on earnings. It is important to note that in a simple regression of child's income on parents's income, the coefficient will capture all effects together. Hence standard estimates of intergenerational earnings regression will provide an upward biased estimate of the causal effect of father earnings on child's earnings. Concretely we will estimate

$$Y_{it} = \beta Y_{it-1} + \mu_{it} \tag{5}$$

From a policy point of view, this distinction between the different component could be matters to predict the impact of economic policies or to know which policy could be better to improve the mobility.

4 Estimation method

4.1 The econometric model

As I explained above, following the economic approach, I focus on intergenerational mobility measured by the intergenerational elasticity of offsprings' earnings with respect to fathers' earnings. More precisely, we consider the following intergenerational mobility equation:

$$Y_i = \alpha + \beta X_i + \mu_i \tag{6}$$

where Y_i is the offspring's log earnings; X_i is the fathers' log earnings (the Y_{it-1} in the previous section); α is the intercept term representing the average change in the child's log earnings and μ is a random error identically and independently distributed (i.i.d.) with zero mean and homoskedastic. The coefficient β is the intergenerational elasticity of children's earnings with respect to their father's earnings, and it is our parameter of interest.

Let ρ be the correlation between Y and X; then β is related to ρ by the following equation:

$$\beta = \rho \frac{\sigma_Y}{\sigma_X} \tag{7}$$

where σ is the covariance. In other words, the coefficient is related to the correlation between children's and fathers' log earnings. Moreover, β is exactly equal to ρ when: $\sigma_X = \sigma_Y$.

A coefficient β equal to zero indicates a situation where all children have "equal opportunities". When $\beta = 0$ all children have an average log earnings equal to α. When β is instead different from zero, offsprings' average log earnings depend also on their fathers' earnings.

On the other hand, a value of $\beta = 1$ indicates a situation of incomplete immobility, whereby (apart of the influence of ε) the children's position in their status distribution is fully determined by their father's position.

As Lefranc and Trannoy (2004) point out the elasticity concept seems more in tune with what economists would like to measure. For example, suppose that some policy reduces all income deviations from child's generation mean by the same factor. We hope to conclude that the inheritage of parental income has decreased with such a policy. In this situation the correlation coefficient remains invariant, meanwhile the elasticity coefficient decreases.

If I had permanent income for successive generations in our sample, I would estimate equation 6 using ordinary least square directly without any problem. Unfortunately I do not have this information in one data set.

First, most data set only provide measures of current earnings and fail to provide measures of individual permanent income. Solon (1992) and Zimmerman (1992) show that the use of current earning as proxy for permanent earnings lead to downward OLS estimates of β. Different solutions have been implemented to reduce or eliminate this bias. One possibility is to work with panel data on fathers earnings and consist in using an average of father's current earnings over several years as a proxy of permanent income. Another alternative consist in use instrumental variables to estimate β. In this paper, we estimate father's earnings using auxiliary variables. Furthermore, as in the case of children, we choose adult ages trying to estimate the intergenerational earnings mobility as close as possible to the age in which earnings are similar to the permanent income.

Second, we have some selection problems that lead us to inconsistent estimations of β. In the next subsection I describe the main selection problems that I have and how we solve in this paper.

4.2 Sample selection problems

Frecuently the estimation of intergenerational earnings mobility can be biased due to different sample selection problems.

The two most important selection problems we have in short panels are the co-residence selection and the selection into employment.[3]

Following Nicoletti and Francesconi (2006), we call **co-residence selection** to the fact that we only observe earnings for pairs of parents and children when they live together in at least one wave of the panel and we do not have information for children who never co-resident with their parents during the panel. This selection problem could lead to an under-representation of the real earnings adults offspring have because if they continue living in the parental house probably is because they are still student or they do not have enough earnings to live independently. Thus, they are not a random sample. In general this selection problem cause an overestimation of intergenerational mobility (an underestimation of the elasticity between parents' earnings and offsprings' earnings). If the panel is long we do not have to deal with this selection problem because is easy to observe young children living together with their parents and follow them to adulthood to know their earnings, except if they leave the panel (attrition problems) or if they do not have job (employment selection).

In this paper we deal with this selection problem linking two samples as I explain in the next subsection. One with information of adults and characteristics (occupation,

[3]Only few papers on intergenerational mobility deal with these selection problems. For the employment selection see for example Couch and Lillard (1998), Minicozzi (2003), Ermisch, Francesconi, and Siedler (2006), Nicoletti and Francesconi (2006). For the case of co-residence selection indeed there are fewer, see Couch and Lillard (1998), Comi (2003) and Nicoletti and Francesconi (2006).

education, age) of the parents when the children had 14 years old and another sample with the same parents' characteristics but with their earnings.

The **employment selection** refers to the problem that we only have earnings for adults when they are employed. However, the decision to work or not work is not random, especially in the case of women. Thus, those who are working constitute a self-selected sample. Estimate intergenerational earnings mobility only for those who are working give us biased estimators. For daughters, we deal with this problem using a Heckman-type of correction estimation described in Vella (1998) and used in Ermisch, Francesconi, and Siedler (2006), which is based on exclusion restriction. In particular, the regressions for daughters include a cubic polynomial of the single index function that determines the selection into employment. The variables included in the selection equation are dependent children, marital status, age and father's earnings. In all regressions, these are good predictors of participation.

4.3 Intergenerational elasticity with sample selection

As we express above the co-residence selection problem can be solved if we have characteristic of the fathers, because we can use these characteristics as auxiliary variables to impute his earnings. This is what we do when we use the two-sample two-stage least squares (TS2SLS).

Since I do not have information of X but I have a set of instrumental variables Z of X, we can estimate equation 6 in two steps. Let us consider two independent samples: the first one has data on offspring log earnings, Y, and characteristics of their fathers, Z, which we call the main sample; and the second sample has data on fathers' log earnings, X, and their age, education and occupational characteristics, Z, which we call the supplemental sample . In the empirical application we combine the supplemental and the main sample to estimate the intergenerational equation 6 by using the TS2SLS estimator.

In the first step we use the supplemental sample to estimate a log earnings equation for fathers using as explanatory variables their characteristics, Z, that is:

$$X_i = Z_i\delta + v_i \tag{8}$$

In the second step we estimate the intergenerational mobility equation 6 by using the main sample and replacing the unobserved X by its predictor,

$$\hat{X} = Z\hat{\delta}, \tag{9}$$

where $\hat{\delta}$ are the coefficients estimated in the first step while Z are the variables observed in the main sample. This method can be viewed as a cold-deck linear regression imputation. Cold-deck refers to the fact that an external data source (the supplemental sample) is employed to estimate the coefficients used to impute the missing X in the main sample. This method was first proposed by Klevmarken (1982). Thus, we estimate equation 6 by using the imputed fathers' earnings.

$$Y_i = \alpha + \beta(Z_i\hat{\delta}) + u_i \tag{10}$$

Equations 8 and 10 are estimated with OLS and standard error of estimates form equation 10 are corrected for heteroscedaticity.[4] To take into account the life-cycle profiles, estimation of both equations include additional control for individual's and father's age. This estimation procedure is very similar to the IV estimation, using Z_i as instrumental variables, except for the fact that the first step estimates are taken from a different sample than the second step.

This estimator is asymptotically equivalent to the 2SIV (two-sample instrumental variable) estimator described by Angrist and Krueger (1992), Arellano and Meghir (1992) and Ridder and Moffit (2006). Both estimators are consistent under the assumptions described in Angrist and Krueger (1992). In particular both estimators are not consistent if the two samples used are not two independent random samples. Moreover, the instrumental variables common to both samples have to be identically and independently distributed in the two samples. Instrumental variable estimator is numerically identical to the two-stage least squares.[5]

The choice of the instrumental variables in some previous papers that estimate intergenerational mobility combining two different datasets was dictated by the few variables available. Björklund and Jäntti (1997) use father's education and occupation.

[4]Heteroscedasticity is taken into account using the Huber White sandwich estimator of the variance.

[5]The two types of estimators produce mathematically the same estimated coefficients when using a single sample, their equivalence holds instead only asymptotically when combining two separate samples. In our estimation procedure we use the TS2SLS to estimate the intergenerational mobility equation, but we consider standard error properly estimated to take account of the replacement of X with its prediction, see Arellano and Meghir (1992).

Grawe (2004) uses only the education levels, while Fortin and Lefebvre (1998) uses only 16 occupational groups, which, as the authors admit, can affect the quality of the imputation of earnings for fathers. One of the exceptions is Lefranc and Trannoy (2004) who use instead 8 different levels of education, 7 occupational groups and age. In Nicoletti and Ermisch (2007), the set of candidates as instrumental variables is also quite large and they try different combinations of the instrumental variables available.

As emphasized by Bound, Jaeger, and Baker (1995), when the instrumental variables are weakly correlated with the variable to be instrumented, "[...] then even a weak correlation between the instruments and the error in the original equation can lead to a large inconsistency in the IV estimates." This suggest choosing instruments such that the R^2 of the imputation regression be as higher as possible.

Nevertheless, in our case, in contrast to Bound, Jaeger, and Baker (1995), the variable to be instrumented, the fathers' log earnings X, is exogenous or at least assumed so. In other words X is independent of u and u is independent of v. Under this assumptions, the ordinary least squares (OLS) estimation of the intergenerational mobility equation produces consistent estimators. The reason why we use the TS2SLS estimator is to combine two separate samples to solve the problem of missing X. The consistency of the TS2SLS (2SIV) estimator requires that $\hat{X}$ be exogenous.

Nicoletti and Ermisch (2007) also discuss what happens when the instruments are endogenous. They arrive to the conclusion that the well-known rule for the choice of the instruments still applies. Instruments should be independent of u and with maximum multiple correlation with X, that is such that R^2 be maximum.

5 Data Sources and Sample Selection Rules

As we explained above, we combine two separate samples to estimate intergenerational mobility, a main sample and a supplemental sample.

In our case, the main sample is the Survey of Living Conditions (Encuesta de Condiciones de Vida (ECV)) for the year 2005, that is the Spanish component of the European Union Statistics on Income and Living Conditions (EU-SILC).[6]

[6] The EU-SILC is an instrument aiming at collecting timely and comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions. This instrument is anchored in the European Statistical System (ESS). The EU-SILC was launched in 2004 in 13 Member States (BE, DK, EE, EL, ES, FR, IE, IT, LU, AT, PT, FI and SE) and in NO

The ECV has annually interviewed a representative sample of about 14,000 households, keeping each household 4 years in the sample. Personal interviews are collected, at approximately one-year intervals, for adult members of all households.

From ECV we have information about son's and daughter's earnings and a set of characteristics of their fathers when the children were between 12 and 14 years old.

My supplemental sample is the Household Budget Survey of 1980-1981 (Encuesta de Presupuestos Familiares). This survey was designed with the aim of estimate consumption and the weights of the different goods used in the consumer index price. But, we also have, for the head of household, information about earnings, occupation and education level. Thus, in this sample we have data on the same set of fathers' characteristics as we have in the main sample and but additionally here we also have their earnings.

Although we have the same characteristics in both samples, we have to recode some variables to have an homogenous classification across surveys.[7]

We consider the main sample given by individuals, either head of household or spouse of the household head, born between 1955 and 1975, self-employed or in paid employment, who report positive labour earnings and they are full time workers. Thus, in the year 2005 they were between 30 and 50 years old and they are 12 or 14 years old between 1969 and 1989. This is the reason I use the Household Budget Survey of 1980-1981 as supplemental sample.

We suppose that when the child are 12 or 14 years old, the fathers are between 37 and 57 years old. Thus, when we estimate the earnings father's regression we select males between those ages.

As I explained above one problem that can bias intergenerational mobility studies is the measurement error in earnings. Theoretically, we would like to consider the intergenerational elasticity in long run permanent earnings but earnings can be observed only in a single or few specific years. The question is then, which is the age at which the current earnings should be observed to provide a proper measure of permanent earnings? Looking at the results in Haider and Solon (2006) and assuming that simi-

and IS. This first release of the cross-sectional data mainly refers to income reference year 2003 with a fieldwork carried out in 2004. The EU-SILC will reach its full scale extension with the 25 Member States plus NO, IS in 2005. It will later be completed by TR, RO, BG and CH.

[7]For a detailed description of the frequencies of the different characteristics in the main and supplemental sample see the appendix.

lar results hold for other countries, it seems reasonable to choose sons around age 40 and fathers with an age between 31 and 55. In our empirical application I follow this suggestion.

After the exclusions, I have a total of 4,352 pairs and in this sample I have fathers and children employed that reported a positive earnings.

Table 1: Descriptive statistics: Sons in the final sample

	sons 30-40	sons 40-50
Observations	1,334	1,322
annual earnings	19,728.35	22,403.7
log of annual earnings	9.72	9.84
Education		
Primary education	13.49%	19.48%
Secondary education (first step)	24.47%	25.00%
Secondary education (second step)	25.42%	24.59%
Vocational qualification	2.64%	1.73%
Higher education (university)	33.97%	29.21%
Occupation		
Higher-grade professionals	5.01%	6.6%
Higher-grade manager	11.65%	10.94%
Low grade professional	12.06%	9.97%
Routine non-manual employees high grade	7.99%	10.80%
Routine non-manual employees low grade	10.98%	9.28%
Skilled agriculture workers	2.37%	3.09%
Skilled manual workers	23.51%	22.70%
Low grade technician	12.33%	13.69%
Unskilled workers	14.09%	12.93%

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

Tables 1 and 2 present the principal descriptive statistics of our final sample of sons and daughters respectively. Tables A.2 and A.3 in the Appendix show the transition matrix between fathers and children. These tables give us an intuitive vision of the persistence of earnings or education.

Table 2: Descriptive statistics: Daughters in the final sample

	daughters 30-40	daughters 40-50
Observations	875	821
annual earnings	13,539.65	15,584.45
log of annual earnings	9.2	9.31
Education		
Primary education	10.39%	17.44%
Secondary education (first step)	19.95%	21.54%
Secondary education (second step)	21.78%	23.35%
Vocational qualification	2.35%	1.11%
Higher education (university)	45.52%	36.67%
Occupation		
Higher-grade professionals	1.59%	1.96%
Higher-grade manager	17.44%	19.54%
Low grade professional	11.68%	9.90%
Routine non-manual employees high grade	21.76%	16.89%
Routine non-manual employees low grade	21.08%	19.80%
Skilled agriculture workers	0.91%	0.85%
Skilled manual workers	4.85%	5.38%
Low grade technician	2.35%	1.71%
Unskilled workers	18.35%	23.98%

6 Results

6.1 Main Results

In this section we present the empirical results on intergenerational mobility estimation. As explained before, I use a two-sample two-stage estimation whose first step consist on the estimation of father's earnings regression using the supplemental sample. The results of this regression are presented in table 3. Then, these coefficients are used to impute fathers' earnings in the main sample. I have the same characteristics in both samples (main and supplemental) and I also have the coefficients from the supplemental sample, thus, we can estimate earnings for each father in the main sample.

Table 4 reports the second step, the coefficients of the intergenerational regression between annual earnings for children (sons and daughters) and fathers' earnings. In both regressions, father's predicted log earnings has a significant positive effect on child's earnings.

We estimate the elasticity for sons and daughters for two different cohort, those whose age are between 30 and 40, and also for the cohort born between 1955 and

Table 3: First step: estimates of father's earnings equation

Dependent variable	log father's earnings
age	0.0571 (0.0211)
age square	-0.0006 (0.0002)
Education	
Primary education	0.1873 (0.0148)
Secondary education (first step)	0.3919 (0.0276)
Secondary education (second step)	0.5254 (0.0326)
Vocational qualification	0.5581 (0.0487)
Higher education (university)	0.8455 (0.0281)
Occupation	
Higher grade manager	-0.4381 (0.0404)
Low grade professional	-0.0753 (0.0986)
Routine non-manual employees high grade	-0.0913 (0.0279)
Routine non-manual employees low grade	-0.3158 (0.0320)
Skilled agriculture workers	-0.8155 (0.0306)
Skilled manual workers	-0.1395 (0.0300)
Lower-grade technician	-0.2009 (0.0298)
Unskilled workers	-0.3177 (0.0285)
Constant	11.9961 (0.4918)
Obs	5929
R^2	0.402

Note: standard errors in parentheses. In **Education**: none (reference) and in **Occupation**: Higher-grade professionals (reference).

Table 4: Second Step: Intergenerational regression in annual earnings

	sons 30-40	sons 40-50	daughters 30-40	daughters 40-50
father's earnings	0.380 (0.042)	0.427 (0.041)	0.504 (0.066)	0.582 (0.061)
age	0.140 (0.005)	0.022 (0.005)	0.028 (0.008)	0.010 (0.008)
Constant	4.258 (0.596)	3.315 (0.605)	1.829 (0.936)	1.513 (0.895)
Obs.	1334	1322	875	821
R^2	0.061	0.08	0.072	0.10

Note:Dependant variable is log of annual labor earnings. Fathers earnings refers to the log of father annual labor earnings. Robust standard errors in parentheses.

1965, those who are between 40 and 50 in 2005. For sons (first and second columns), regression coefficients are around 0.40 and for daughters (third and fourth columns) are around 0.54.

We observe smaller correlation for the younger cohorts. There are two possible explanations for this fact, the first one, is that for younger cohort we do not observe the permanent earnings because they are at the beginning of the working career. The second hypothesis is that in Spain the intergenerational mobility has increased. The younger cohorts earnings are less correlated with father's earnings.

Comparing the estimates for sons and daughters we obtain a higher correlation for daughters. This result should not be surprising. We have to remember that our sample is restricted to full time workers and the increase of participation of woman in the Spanish labour market began at the end of seventies. It is intuitive that in older women workers are probably more common in some types of household (high educated household or very poor household), thus the correlation is higher.

Since women are getting more and more independent form a financial point of view and the women role is changing, this argument seem less relevant.

Our estimation of intergenerational earnings mobility in Spain can be compared to results obtained for other countries. However, when we want to compare our results, we should be aware of the potential impact of differences in the definition of the children's sample and the estimation method applied.

For example, in the US, the correlation ranges from 0.13 to 0.61 depending on the study considered. Solon (1999) provides an extensive survey of the US results obtained in the nineties and conclude: "all in all, 0.4 or a bit higher also seems a reasonable guess of the intergenerational elasticity in long-run earnings for men in the United States". This conclusion is obtained in studies using multi-year averages of father and child earnings, computed from panel data, as a measure of individual permanent income.

Mazumder (2001) using long panel of social security files, he point out that the larger the time used to average the earnings, the higher is the intergenerational elasticity. For example, averaging earnings over a period of 16 years leads to an elasticity of 0.613.

A good benchmark for comparing our results is provided by Björklund and Jäntti

(1997), a study that appears very close to our, both in terms of sample definition and method used. Their results show an elasticity of 0.52 for the United State and 0.28 for Sweden. Nicoletti and Ermisch (2007) applying the same methodology for Britain, obtain an elasticity that ranges from 0.20 to 0.25 for sons. In the same way, Lefranc and Trannoy (2004) find an elasticity of 0.40 for sons and 0.30 for daughters. Thus, comparing these results with our estimations, we observe that Spain presents less intergenerational mobility than France, Sweden and Britain but more than the United State.

One possible explanation why Europe shows more intergenerational mobility than the United State is the way higher education is financed. In Spain, France, Sweden the access to higher education is free, while in the United State payment of tuition may be a problem for poor household, even if generous grants are available for bright students.

But clearly this is not a definite answer, our results should be confirmed and improved using more years of the main sample to obtain a better proxy for permanent child's earnings.

Evidence available for other countries and surveyed by Solon (2002) suggests a rather high degree of intergenerational mobility in Finland (Österbacka (2001)) and Canada (Corak and Heisz (1999)), where the elasticity is around 0.2 or lower. There is some empirical evidence for Germany (see Couch and Dunn (1997)) that expresses a similar correlation to the United States.

Overall, we find an intergenerational correlation for Spain that ranks between a group of more mobile societies including Nordic countries, Canada and Britain and a group of less mobile countries which include the United States. We find an elasticity that is similar to France for sons. However, in the case of daughters I obtain larger elasticity than in France.

Table 4 reports the estimation of intergenerational earnings mobility correcting the co-residence section problem. However, this estimation forgot the point that participation in the labour market, especially for women, is not a random. Thus, in table 5 I present the result of the estimation of equation 10 correcting for the employment selection in the case of women. I use the variables married, having children and father's earnings and age to predict selection. If we compare the two last column of table 4

Table 5: Intergenerational earnings mobility for women correcting for employment selection

	daughters 30-40	daughters 40-50
father's earnings	0.369 (0.074)	0.598 (0.062)
age	0.043 (0.009)	0.009 (0.008)
Constant	3.285 (1.042)	1.287 (0.919)
Obs.	1025	992
R^2	0.072	0.10

Note:Dependant variable is log of annual labor earnings. Fathers earnings refers to the log of father annual labor earnings. Robust standard errors in parentheses.

with table 5 we observe some differences, especially in the case of younger women. The correlation between father's earnings and daughter's earnings is smaller when we correct by the employment selection with a Heckman selection model. Probably, the decision to work or not to work, is less random for the younger women cohorts.

6.2 Sources of earnings correlation

The two-sample instrumental variable estimation allows for a decomposition of the sources of earnings elasticity across generations. Following Lefranc and Trannoy (2004) and assuming that both child and father's log permanent income are observed and that each can be expressed as:

$$Y_i^g = Educ_i^g \delta_e^g + Occup_i^g \delta_o^g + v_i^g \quad for\ g \ = \ child\ or\ father \tag{11}$$

where $Educ$ is the individual's education and $Occup$ is the individual's occupation. Thus, our two-step estimate of β is simply given by:

$$\beta = \frac{cov(Y_i^c, Educ_i^f \delta_e^f + Occup_i^f \delta_o^f)}{V(Educ_i^f \delta_e^f + Occup_i^f \delta_o^f)}$$

Expanding this term, I can rewrite β as a decomposition of six terms:

$$\beta = \frac{1}{V(Educ_i^f \delta_e^f + Occup_i^f \delta_o^f)} \times \Big[\delta_e^c cov(Educ_i^c, Educ_i^f)\delta_e^f + \delta_o^c cov(Occup_i^c, Occup_i^f)\delta_o^f$$
$$+ \delta_e^c cov(Educ_i^c, Occup_i^f)\delta_o^f + \delta_o^c cov(Occup_i^c, Educ_i^f)\delta_e^f + cov(v_i^c, Educ_i^f)\delta_e^f$$
$$+ cov(v_i^c, Occup_i^f)\delta_o^f\Big]$$

It is important to notice that this composition should be seen as a descriptive device along the lines suggested in Bowles and Gintis (2002) and not as an analysis of causal effects.

The results of applying this decomposition to the estimation of earnings elasticity presented in table 4 are given in table 6

Table 6: Decomposition of earnings regression coefficient

	sons 30-40	sons 40-50	daughters 30-40	daughters 40-50
$educ_c - educ_f$	0.080	0.084	0.098	0.107
$occup_c - occup_f$	0.143	0.152	0.161	0.187
$educ_c - ocup_f$	0.065	0.071	0.081	0.094
$occup_c - educ_f$	0.055	0.082	0.105	0.110
$res_c - educ_f$	0.002	0.018	0.014	0.032
$res_c - occup_f$	0.035	0.020	0.045	0.052
total	0.380	0.427	0.504	0.582

These results can be interpreted as: assuming that the only channel of intergenerational earnings correlation would work through the correlation of father and child's education, earnings regression coefficient for our 2005 sons sample between 30 and 40, using the father sample, would be equal 0.08.

It seems from this table that for all ages and for both sons and daughters the correlation between occupations is important to understand the intergenerational correlation in earnings. Comparatively, father's education account for a smaller share of intergenerational correlation.

6.3 Quantile regressions

When we regress the children's earnings on their parent's earnings we provide a measure of intergenerational mobility at the mean. However, when we estimate quantil regressions we have a more complete picture of intergenerational transmission of earnings because we have information of the correlation between children's and parent's earnings at different points of the distribution of the children's earnings.

The mean regression explain how the conditional mean of the children earnings depend on parents earnings, however quantile regressions explain how children earnings depend on parents earnings at each specific quantile of the conditional distribution of children earnings given father earnings.

Table 7: Intergenerational mobility by quantils

	Average	10th	25th	50th	75th	90th
sons 30-40	0.380	0.428	0.339	0.391	0.356	0.394
	(0.042)	(0.109)	(0.762)	(0.032)	(0.059)	(0.067)
sons 40-50	0.427	0.656	0.435	0.468	0.502	0.485
	(0.042)	(0.107)	(0.059)	(0.044)	(0.044)	(0.051)
daughters 30-40	0.504)	0.813	0.691	0.429	0.446	0.281
	(0.066)	(0.212)	(0.124)	(0.108)	(0.065)	(0.056)
daughters 40-50	0.582	0.938	0.864	0.724	0.641	0.410
	(0.061)	(0.177)	(0.064)	(0.067)	(0.081)	(0.069)

Note: Standard error for the estimated coefficients are in parenthesis. Average refers to mean regression, whereas q-th indicates the q-th percentile regression.

For example, if the effect of having an increase in the parents log earnings is better for children with lower salaries than children with higher salaries, then the intergenerational elasticity as a mean only gives partial information of the correlation between parents and children.

In table 7 we can observe the coefficient of the father's log earnings. In the first column we show the mean regression, which inform us how important father's earnings are on average. Meanwhile, in the rest of the columns quantile regressions evaluate the influence of father's earnings at each specific quantile. I consider the 10th, 25th, 50th, 75th and 90th percentiles. We can observe that the influence of father's earnings is greater as soon as we move for the poorest quantiles of the distribution. Thus, the mobility is lower for the children born in a disadvantaged family.

7 Final remarks

In this paper I analyze the intergenerational earnings mobility in Spain considering some sample selections problems, as, for example, co-residence and employment selection. Since there is not Spanish survey with information on children and their fathers' earnings covering a long period, we deal with the co-residence selection considering two separately samples: a main sample containing information on children's earnings and a set of characteristics of the fathers, and a supplemental sample with the same characteristics for the fathers and their earnings. We combine the two samples by using the two-sample two-stage least square estimator described by Arellano and Meghir (1992)

On average we find an elasticity of 0.40 for sons and around 0.55 for daughters. Our results suggest that intergenerational mobility increases when we move to younger cohorts. Furthermore, when we compare the estimation for sons and daughters, we find more mobility for sons. One possible explanation, since our sample is restricted to sons and daughters full time workers, is perhaps full time women workers for the older cohort are probably more common in some types of household (high educated household or very poor household), thus the correlation is higher.

However, when we analyze the younger cohort, the coefficients are more similar between sons and daughters.

According to our findings, Spain shows a degree of intergenerational earnings mobility that is similar to France, lower than the Nordic countries and Britain and higher than the United States.

Furthermore, we correct the point that participation in the labour market for women is not a random estimating the earnings elasticity between daughters and fathers correcting for the employment selection with a Heckman selection model. The elasticity between father's earnings and daughter's earnings is smaller when we correct by the employment selection. Probably, the decision to work or not to work, is less random for the younger women cohorts.

Finally, when we estimate the regression by quantiles, we find that the influence of father's earnings is greater as soon as we move for the poorest quantiles of the distribution. Thus, the mobility is lower for the children born in a disadvantaged family.

References

ANGRIST, J. D., AND A. B. KRUEGER (1992): "The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples," *Journal of the American Statistical Association*, 87, 328–336.

ARELLANO, M., AND C. MEGHIR (1992): "Female labour supply and on-the-job search: an empirical model estimated using complementary data set," *The Review of Economic Studies*, 59, 537–559.

AYALA, L., AND M. SASTRE (2002a): "Europe vs. the United States: is there a trade-off between mobility and inequality?," ISER working papers 2002-26, Institute for Social and Economic Research.

——— (2002b): "What Determines Income Mobility Differences Across the European Union?," ISER working papers 2002-27, Institute for Social and Economic Research.

BECKER, G., AND N. TORMES (1979): "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy*, 87, 1153–1189.

BEHRMAN, J. R., AND P. TAUBMAN (1990): "The Intergenerational correlation between Children's Adult earnings and their Parents' Incom: Results from the Michigan Panel Survey of Income Dynamic," *Review of Income and Wealth*, 36(2), 115–127.

BJÖRKLUND, A., AND M. JÄNTTI (1997): "Intergenerational Income mobility in Sweden Compared to the United State," *American Economic Review*, 87, 1009–1018.

——— (2000): "Intergenerational mobility of socioeconomic status in comparative perspective," *Nordic Journal of Political Economy*, 26(1), 3–32.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association*, 90(430), 443–450.

BOWLES, S., AND H. GINTIS (2002): "The Inheritance of Inequality," *Journal of Economic Perspectives*, 16, 3–30.

CANTÓ, O. (2000): "Income Mobility in Spain: How much is there?," *Review of Income and Wealth*, 46(1), 86–102.

CARABAÑAS, J. (1999): *Dos estudios sobre movilidad intergeneracional.* Fundación Argendaria-Visor (ed.).

CHECCHI, D. (2006): *The economics of education: human capital, family background and inequality.* Cambridge University Press, first edn.

COMI, S. (2003): "Intergenerational mobility in Europe: evidence from ECHP," Departemental Working Papers 2003-03, Department of Economics University of Milan Italy.

CORAK, M., AND A. HEISZ (1999): "The intergenerational income mobility canadian men," *Journal of Human Resources*, 34, 504–533.

COUCH, K., AND T. DUNN (1997): "Intergenerational correlations in labor market status: A comparison of the United State and Germany," *Journal of Human Resources*, 32(1), 210–232.

Couch, K., and D. R. Lillard (1998): "Sample selection rules and the intergenerational correlation of earnings," *Labour Economics*, 5, 313–329.

Erikson, R., and J. H. Godthorpe (2002): "Intergenerational inequality: A Sociological Perspective," *Journal of Economic Perspective*, 16, 31–44.

Ermisch, J., M. Francesconi, and T. Siedler (2006): "Intergenerational economic mobility and assortative mating," *Economic Journal*, 116, 659–679.

Fortin, N., and S. Lefebvre (1998): "Intergenerational income mobility in Canada," in *Labour Market, Social Institution and the Future of Canada's Children*, ed. by M. Corak. Statistics of Canada, Ottawa.

Grawe, N. (2004): "Intergenerational mobility for whom? The experience of high- and low-earnings sons in intergenerational perspective," in *Generational Income Mobility in North America and Europe*, ed. by M. Corak. Cambridge University Press, Cambridge.

Haider, S., and G. Solon (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," NBER Working Papers 11943, National Bureau of Economic Research, Inc.

Hugalde, A. S. (2004): "Movilidad intergeneracional de ingresos y educativa en Espaa (1980-90)," Discussion paper.

Klevmarken, W. A. (1982): "Missing variables and two-stage least-squares estimation from more than one data set," *Proceeding of the American Statistical Association 1981*, Business and Economic Statistics Section, 156–161.

Lefranc, A., and A. Trannoy (2004): "Intergenerational earnings mobility in France : Is France more mobile than the US ?," IDEP Working Papers 0401, Institut d'economie publique (IDEP), Marseille, France.

Mazumder, B. (2001): "Earnings mobility in the US: A new look at intergenerational inequality," Discussion paper.

Minicozzi, A. (2003): "Estimation of sons' intergenerational earnings mobility in the presence of censoring," *Journal of Applied Econometrics*, 18, 291–314.

Nicoletti, C., and J. Ermisch (2007): "Intergenerational Earnings Mobility: changes across cohorts in Britain," *The B.E. Journal of Economic Analysis and Policy. Contributions*, 7, 1–36.

Nicoletti, C., and M. Francesconi (2006): "Intergenerational mobility and sample selection in short panels," *Journal of Applied Econometrics*, 21(8), 1265–1293.

Österbacka, E. (2001): "Family background and economic status in Finland," *Scandinavian Journal of Economics*, 103(3), 467–484.

Ridder, G., and R. Moffit (2006): "The econometrics of data combination," in *Handbook of Econometrics*, ed. by Heckman, and Learner, 6. Elsevier Science, North Holland, Amsterdam.

Rodriguez, J. P., R. Salas, and S. A. García (2002): "Movilidad social y desigualdad Economica," Discussion paper.

Solon, G. (1992): "Intergenerational income mobility in the United States," *American Economic Review*, 82(3), 393–408.

——— (1999): "Intergenerational Mobility in the Labour Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelder, and D. Card, vol. 3, chap. 29, pp. 1761–1800. Amsterdam: Elsevier.

——— (2002): "Cross-country differences in intergenerational earnings mobility," *Journal of Economic Perspective*, 16(3), 59–66.

Vella, F. (1998): "Estimating models with sample selection bias: A survey.," *Journal of Human Resources*, 3, 127–169.

Zimmerman, D. (1992): "Regression toward mediocrity in economic stature," *American Economic Review*, 82, 409–429.

Appendix

Table A.1: Distribution of father's education and occupation an coincidences between suplemental and main sample

	supplemental sample	main sample
Observation	5,032	4,352
Education		
No finish primary education	23.82	20.09
Primary education	51.28	57.65
Secondary education (first step)	8.46	6.08
Secondary education (second step)	5.90	5.84
Vocational qualification	2.07	0.49
Higher education (university)	8.47	9.85
Occupation		
Higher grade professionals	9.25	8.04
Higher grade manager	4.28	3.70
Low grade professional	3.43	5.58
Routine non-manual employees high grade	11.04	6.18
Routine non-manual employees low grade	9.85	7.25
Skilled agriculture workers	12.74	12.85
Skilled manual workers	15.88	24.99
Lower-grade technician	13.81	11.82
Unskilled workers	19.71	19.60

Note: All frequencies are weighted using the respective sampling weights.

Table A.2: Transition matrices of earnings between fathers and child

		Quantil of the father				
		1	2	3	4	5
Quantil of the son or daughter	1	30,08%	23,93%	16,98%	16,20%	13,23%
	2	24,40%	22,34%	19,17%	18,29%	16,20%
	3	19,12%	23,54%	20,26%	21,67%	15,66%
	4	15,74%	15,69%	22,64%	23,26%	22,41%
	5	10,66%	14,50%	20,95%	20,58%	32,49%

Table A.3: Transition matrices of education between fathers and child

		Education of the father					
		0	1	2	3	4	5
Education of the child	1	34,07%	13,89%	4,85%	3,04%	0,00%	0,60%
	2	34,77%	23,72%	18,12%	7,43%	8,00%	3,99%
	3	17,98%	25,22%	34,30%	31,42%	36,00%	16,37%
	4	1,90%	2,18%	1,94%	1,01%	12,00%	1,00%
	5	11,29%	34,98%	40,78%	57,09%	44,00%	78,04%