Income mobility in Spain between 1999 and 2008: evidence from tax data

María Gil Izquierdo*

Universidad Autónoma de Madrid

Fidel Picos Sánchez**

REDE-Universidade de Vigo

[preliminary version, please do not cite]

Abstract

The aim of this paper is to analyse income mobility in Spain using the Spanish PIT Return Panel (1999-2008). Although the use of tax records for income mobility analysis is quite common in other countries, almost all of the mobility studies in Spain have used survey data. Tax data have some drawbacks, but they have the advantage of providing more accurate measures of income and high representativity for high income earners. The preliminary results show lower mobility than in previous studies, although further research should be carried out to identify the possible effects of methodological choices and differences in data characteristics that can affect the results.

* Corresponding author. María Gil Izquierdo. Departamento de Economía Aplicada, Fac. CC. EE., Avda. Tomás y Valiente, 5, Cantoblanco. 28049 Madrid (Spain). Phone: +34914973524. E-mail: maria.gil@uam.es

** Fidel Picos Sánchez. REDE-Departamento de Economía Aplicada, Facultade de Ciencias Empresariais e Turismo, Universidade de Vigo. As Lagoas s/n, 32004 Ourense (Spain). Phone: +34988368771. E-mail: fidel@uvigo.es.

1. Introduction

Income mobility analysis is nowadays a well-established research line, with both theoretical and empirical contributions that have underpinned the study of longitudinal trends in inequality. For the last decades, the availability of new and more complete longitudinal databases in many countries (especially Anglo-Saxon countries) has definitely boosted the production of results related to changes in income distribution and therefore, the conclusions we may obtain from these results.

The situation in Spain is somehow different. Panel databases have not experienced the same level of development as cross-section databases. Obviously, the former are much more costly in all terms than the latter. The scarcity of this source of information has therefore limited the possibilities to empirically contrast the patterns of mobility in Spain. Panel data have been mainly built from living conditions surveys, addressed to individuals or households, with information along the years about their incomes and some other socioeconomic characteristics.

The aim of this paper is to contribute to the income mobility literature for the Spanish case[1]. The added value of our study is the use of data from a panel of individual income tax returns: the Spanish PIT Return Panel (1999-2008). The use of tax records for mobility analysis purpose is rare in Spain (we are aware of just one study using a previous panel), though their use is quite common in other countries, especially in the United States.

As one can expect, there are advantages and disadvantages of using panel data from surveys and from tax records. Tax records are not conducted with statistical purposes, and for this reason, information about personal characteristics is not as rich as the information obtained from a questionnaire survey. Instead, tax records are superior in two major aspects: they have accurate information about personal income[2] and they

---

[1] We will not be analyzing the intergenerational income mobility, that is, the transmission of economic, social, educational, etc. status from parents to descendents. Pascual (2009), Cervini-Plà (2011) or Gil and De Pablos (2010) provide results in these terms for Spain.
[2] Of course, taking apart the lack of information derived from tax fraud.

provide detailed information about the upper part of the income distribution[3]. This allows us to focus on middle and upper income mobility, which we consider a relevant contribution of this study.

The organization of the paper is as follows. After this introduction, we revise some previous works in Spain. Then we explain the data, the methodology and indices that we will be using. The fourth section offers the results of our study. We close summarizing the main conclusions.

2. Previous works in Spain

Evaluation of income mobility represents an increasing research line in last decades, closely related to the analysis of income inequality, though not as developed. Albeit theoretical studies about the measurement of mobility appears in the fifties (Prais, 1955), and even quite few empirical applications are conducted in later years (Schiller, 1977; Shorrocks, 1981; Lillard and Willis, 1978, to mention some), it is not until the nineties when this literature becomes more popular. Difficulties to define some mobility concepts, a lesser axiomatic development of mobility indices or the shortage of longitudinal data sets, are some of the drawbacks that have conditioned the evolution of empirical applications (Fields and Ok, 1999) until that date. Different studies have contributed over time to partially solve some of these problems, and so an increasing number of papers have been dedicated to measure income mobility, from different perspectives and using different datasets.

The empirical analysis of mobility studies relies, to a large extent, on the quality of the longitudinal database. Obviously, recording the behaviour of the same individuals for a longer than one period requires much more effort and it implies higher costs (not to say for some consecutive years). There are mainly two ways to obtain longitudinal data: from panel surveys and from administrative records. The former ones follow the same individual or household for the whole period (pure panel) or for a fraction of it

---

[3] Not so complete for the downward part of the distribution, due to the usual non-compulsory filing for low income earners.

(rotating panel) and they normally focus on socioeconomic conditions of the unit of analysis. The latter are not collected with statistical purposes, but they may be used with this aim, for instance tax records. For mobility analysis purposes, survey databases have been commonly used, though in the last years administrative records become a more widespread source of information, above all in the United States (see for example US Department of the Treasury (1992a, 1992b), Carrol et al, 2006; Auten and Gee, 2007 and 2009; and Splinter et al., 2009).

Studies for Spain have mainly used data from surveys to analyse income mobility. However, panel data availability is limited to three datasets: the Spanish Household Panel Survey[4] (*Encuesta de Presupuestos Familiares,* ECPF), the European Community Household Panel[5] (*Panel de Hogares de la Unión Europea*, PHOGUE) and the European Union Statistics on Income and Living Conditions[6] (*Encuesta de Condiciones de Vida*, EU-SILC).

Using these surveys, different studies provide results that cover the period from the middle eighties up to present, allowing us to summarize the status and patterns of mobility in Spain. For the eighties, Cantó (2000) makes use of the Spanish Household Panel Survey (1982-1992), detecting a high income mobility, decreasing by the end of the period. Besides, the middle incomes turn out to be the more mobile. For the same period, Gradín, Cantó and del Río (2008) study how some methodological decisions may affect the results of mobility. Ayala and Sastre (2005) cover the period from 1994 to 2000 with the European Community Household Panel, and they also find out high mobility. But instead, their results show a higher mobility for the low and middle incomes. Moreover, the mobility is mainly due to re-rankings more than to economic growth, and more usual in wage earners. Finally, the paper by Bárcena and Moro (2013) refers to the 2001-2010 period, being remarkable as it analyses the beginning of the current economic crisis. Using EU-SILC, they also obtain high mobility patterns

---

[4] For further details, see http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t25/p458&file=inebase
[5] For further details, see http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/echp
[6] For further details, see
http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/metho
dology

though this mobility is short-haul, both up and down. Their results are close to those by Ayala and Sastre (2005), although some changes in the patterns of mobility are detected with the crisis: mobility becomes lower, descents in income distribution are more common and income values are smaller than in previous years.

Finally, a special reference has to be made to the work by Ayala and Onrubia (2001), as it is, as far as we know, the only one that have used tax data for Spain. The authors use a first Spanish PIT Return Panel (Panel de Declarantes por IRPF) that covers the period from 1982 to 1994. In this pioneering study in Spain both inequality and mobility issues related to the Spanish income distribution for these years are analysed. At the same time, a significant effort is made in order to explain specific methodological decisions and their implications in interpretation when using tax data instead of survey data. Their results can be slightly different in comparison to previous works, due to the nature of data (for example, the Hart, Theil and Shorrocks indices). The authors justify these results referring to the sample selection and the range of the considered period of analysis as the source of the low indices. In general terms, they conclude that mobility is higher for low income earners than for high income earners, and that mobility is higher when considering more years and for self-employed.

3. Data and methodology

3.1. The 1999-2008 Spanish PIT Return Panel

In this work we use the 1999-2008 Spanish PIT Return Panel[7] (hereinafter Panel), which is an expanded panel representative of each year's tax filers. Table 1 shows its main specifications.

---

[7] Detailed information about the 1999-2008 Panel can be found in Onrubia et al (2011 and 2012) [in Spanish]. This panel is an improved follow-up of the 1982-1995 Panel used by Ayala and Onrubia (2001), which was later extended to 1998. Both panels are disseminated by the Spanish Institute for Fiscal Studies (Instituto de Estudios Fiscales, IEF) and available free of charge for researchers at http://www.ief.es/recursos/estadisticas/fuentes_tributarias.aspx [retrieved 1st November 2013].

Table 1. Design specifications of the Spanish 1999-2008 PIT Return Panel

		Panel
Period		1999-2008
Type of microdata		Expanded panel
Scope	Reference population	Personal Income Tax filers
	Geographic scope	All the Spanish Territory except the chartered regions (the Basque Country and Navarre)
Observation unit		Tax return (individual or joint)
Sampling	Type	Minimum variance stratification under Neyman allocation
	Sample income	Sum of gross labour income, net income from other sources and income imputations
	Stratification variables	• Income level (10 levels) • Autonomous Communities (17: 15 of the Common Fiscal Regime, plus the Autonomous Cities of Ceuta and Melilla) • Source of income (2: labour income represents more or less than half of the total income of the return)

Source: own elaboration

The advantages and disadvantages of using tax data are well-known (Carroll et al, 2006; Auten and Gee, 2009), and all of them are applicable to the Panel. In particular the usual benefits are:

- The low attrition, which allows following the same individuals for long periods. In the Panel, more than 75% of the original taxpayers are followed throughout ten years.

- The high statistical representativity due to the large size of the samples. The Panel has information on 650.000-850.000 tax filers per year.

- The good representativity of high-income observations, usually due to stratified sampling. In our case there is a 10-level income stratification, and the last three levels include taxpayers earning more than 60,000 Euros.

- The accurate measures of all kinds of income compared to survey data. There is very detailed information in the Panel on more than ten categories of real and imputed income.

The main drawbacks of tax data are:

- The low representativity of low-income individuals, who in general are not required to file a return. Currently in Spain people are not required to file if they earn less than 22,000 of labour income and small quantities of other kinds of income; however we also have their information in the Panel if they choose to file in order to recover withholding taxes

- The personal, demographic and social information, usually less rich than in surveys. The Panel offers some direct information on personal characteristics (age and marital status) but only indirect information on family characteristics (via personal and family allowances).

- The dependence on the definition of the tax unit: it is not always possible to have information on individuals and/or economic households. The limitation in our case is that tax returns in Spain may correspond to individuals (unmarried persons, and married persons who decide to file individually) or married couples (when they choose to file jointly, which is optional). In the former case the Panel provides the sampled tax return and, if the tax filer is married, the tax return of his/her spouse (even if it had not been originally sampled); in the latter case the Panel provides only one tax return without separate information for the two spouses, since all incomes are summed up. As a consequence the only possibility to homogenize observations is to also sum um up the incomes of married couples that file individually. The Panel does this by creating a new observation unit called "tax household", which consist of a married couple or an unmarried individual.

The Panel was designed to keep cross-sectional representativity throughout the years, so the observations sampled in the base year (2003) are complemented with new observations that compensate the dropouts, both for the following and the previous years. This implies that not all observations are available for all the years; however the attrition is low: 88% of the taxpayers originally sampled in 2003 are still present in 2008, and 78% of them in 1999.

Regarding income information, the Panel offers detailed information on the categories of income that are taxable in the Spanish PIT: labour income, capital income, income

from self-employment, imputed income and capital gains and losses. For most of them it is possible to distinguish gross income, deductible expenses and other tax deductions, so it is feasible to build a variable corresponding to real income and not biased by the tax code.

3.2. Data and variable selection

Several decisions had to be taken about the observations and variables that should be used in the analysis. Three main decisions were taken regarding the selection of observations:

- **Observation unit**. The usual observation unit is the individual (Auten and Gee; 2007 and 2009; Splinter et al., 2009), so we decide to follow individuals. In the case of married couples we divide their household income by two, since it seems a better approach than keeping the individual allocation of income[8]. Another decision has to do with marital status. One possibility is keeping all the individuals, irrespective of their marital status (Auten and Gee, 2007). An alternative is to follow only individuals with constant filing status, as a way to remove income changes due to marriage and divorce (Splinter et al., 2009). Since we are interested in analysing all kinds of mobility, no matter what the cause is, we choose the first option, but we also make some calculations using the second option, as a way to measure the effect introduced by marriages and divorces in the period.

- **Age of tax filers**. The two tails of the age distribution are not very useful in mobility analysis. Old people usually have quite constant income due to retirement; while young people may face high increases caused only buy their entry into the labour market. Therefore we restrict the sample to individuals above 25 years old in 1999

---

[8] Furthermore, we do not have individual information for couples who file jointly, so this option homogenizes all types of couples.

and below 65 in 2008[9], so that as a general rule we avoid analysing students and pensioners[10].

- **Years.** We compare the first and the final period of the Panel (1999 vs. 2008), as in Auten and Gee (2007 and 2009) and Carrol et al. (2006). For Spain, Ayala and Onrubia (2001) also compared several intermediate periods.

Regarding the measurement of income, three questions have to be discussed:

- **Definition of income.** We take the usual approach of summing up all reported cash income, including realized capital gains, but not imputed income or income in kind (Carrol et al., 2006; Auten and Gee, 2007). All this information is included in the Panel, except exempted income (e.g. dividends up to 1,500 euro in 2008 or some pensions for disabled people), which is not reported by tax filers[11].

- **Inflation.** We take the usual solution of adjusting to Euros of the final period using the official consumer price index.

- **Income adjustment.** We do not have information on the economic household (number of members, age, etc.), so we may only apply equivalence scales for splitting household income between the two spouses. As already said, we have decided to use a per capita scale, dividing household income by two. However Gittleman and Joyce (1999) say that mobility analyses are less sensitive to equivalence scales than inequality analysis, so results may not change much.

Table 2 shows the number and type of observations along with the main income figures for 1999 and 2008. The first two columns show the data for all the observations for each year, while the third and fourth columns show the results for the observations that are common to both years and were chosen for our analysis.

---

[9] For married couples who file jointly we only have the age of one of the spouses, so we assume the same age for both.

[10] This is the option taken by Gittleman and Joyce (1999). The lower limit of 25 is present in most works, while there are more variations in the upper limit. See Sawhill and Condon (1992), Splinter et al (2009), Auten and Gee (2007 and 2009) and Carrol et al (2006) for alternative choices.

[11] The variable for 1999 is calculated as $c901+c902+c6-(c2+c3+c23+c24+c25+c26+c27)$, and for 2008 as $c455+c465+c17+c18+c19+c20+c76-(c5+c6+c7+c69+c220+c221+c222+c223+c224-c225+c226-c227+c245+c255+c265+c275)$. For the meaning of each variable see Onrubia et al (2012). The definition of the variables is similar to the one used by Ayala and Onrubia (2001).

Table 2. Number of observations and average income

		All observations		Common observations	
		1999	2008	1999	2008
Number of observations	Unmarried	108,042	215,204	79,607	70,452
	Married, joint file	283,615	282,135	199,250	159,699
	Married, individual file	226,363	335,741	162,842	211,548
	Total	618,020	833,080	441,699	441,699
Average income (Euros)	Nominal	12,705.94	19,642.07	12,963.80	23.413,32
	Adjusted	16,822.67	19,642.07	17,164.07	23.413,32

Source: own elaboration

We can see that 441,699 observations (more than 70% of the 1999 observations and more than half of all the 2008 observations) are present in both years. For 1999 the distribution of the selected observations in terms of marital status and file type is very similar to the total population of that year. In contrast, for 2008 the selection implies a smaller proportion of unmarried observations (70,452 out of 215,204, less than one third) and a higher proportion of married observations that file individually (211,548 out of 335,741, 63%). This is probably due to young people that got married in this period of time and kept on working, being individual filing the best option when both spouses work.

Regarding incomes, the picture is analogous. Average income for the selected observations is very similar to average income of the whole population in 1999 (12,963.80 vs. 12,705.94), but higher for 2008 (23,413.22 vs. 19,642.07). This is probably caused by a bias against low income in tax panels, since low income observations tend to disappear for different reasons: incomes that move up and down the minimum threshold, low-income pensioners who pass away, etc.

3.3. Indices

Once we have determined the data we are going to use and the period we want to cover, we have to decide on the mobility measures we want to calculate. Since our aim is offering a general overview of mobility in Spain, we have decided to calculate several well-known indices that can be found in the literature, try to cover three different

aspects: relative mobility, absolute mobility and correlation of incomes. Table 3 shows the selected indices and their formulae.

All the relative indices are based on the transition matrix, which in our case is based on quintiles of income. We chose quintiles (instead of deciles, like other researchers in Spain have done) because some deciles can be too narrow, so the mobility between is not relevant. However, by choosing quintiles we lose information on mobility of high incomes, because the top quintile (and even the last decile) is extremely broad; i.e. mobility within that quintile may be very relevant. The solution we adopt is building specific transition matrices for high income, which we do for the top 10%, 5% and 1% of the population.

Regarding absolute and correlation measures we offer two results for each index: one uses all the selected observations, and another removes the top 1% income observations to avoid distortionary effects or extremely high incomes.

Table 3. Some mobility indices: definitions and formulae

Type	Name	Formula	Definition and interpretation
Relative	Transition matrices (TM)	-	Percentage of movements from each of the five 1999 quintiles to the five 2008 quintiles. Specific matrices for constant marital status and for top income earners.
	Shorrocks (1978)	$M_S = \dfrac{q - trace(TM)}{q-1}$ where q is the number of quantiles, i.e. the number of rows and columns of TM	It takes value $\dfrac{q}{q-1}$ for total mobility and 0 for total immobility
	Immobility Index	$II = \dfrac{trace(TM)}{\sum\limits_{i=1,j=1}^{i=q,j=q} a_{i,j}}$ where $a_{i,j}$ is the element in row i and column j of the transition matrix	Sum of individuals that do not move (trace of the TM) as a proportion of the whole population (sum of all the elements of the TM). It takes value 1 for total mobility and 0 for total immobility
	Mobility Ratio	$MR = \dfrac{\sum\limits_{i=1,j=1}^{i=q,j=q} a_{i,j}, j < i}{\sum\limits_{i=1,j=1}^{i=q,j=q} a_{i,j}, j > i}$	Sum of individuals that move upward (sum of the elements above the main diagonal of the TM) divided by the sum of individuals that move downward (sum of the elements below the main diagonal of the TM). Values above (below) 1 indicate more upward (downward) than downward (upward) mobility
Absolute	Fields and Ok (1996)	$M_{F-Ok} = \dfrac{\sum\limits_{i=1}^{N} \left| \ln(x_{it}) - \ln(x_{i0}^{adj}) \right|}{N}$ where x_{i0}^{adj} is the inflation-adjusted income of the same individual in year 0, x_{it} is the income of individual i in year t and N is the number of individuals in the population	Population average of the differences (in absolute value) between the natural logarithms of the individual income in the two periods
Correlation	Correlation coefficient	$CC = 1 - r(x_{it}, x_{i0}^{adj})$	Correlation coefficient between income in year t and inflation-adjusted income in year 0. 1 indicates complete correlation, 0 no correlation
	Hart Index (Shorrocks, 1993)	$M_{Hart} = 1 - r(\ln(x_{it}), \ln(x_{i0}^{adj}))$	One minus the correlation coefficient between the natural logarithms of income in year t and inflation-adjusted income in year 0. 1 indicates no correlation, 0 complete correlation.
	Regression slope	$RS = \beta$	Coefficient β of the OLS regression between the natural logarithm of income in year t (dependent variable) and natural logarithm of the inflation-adjusted income in year 0 (independent variable). Higher values indicate higher correlation.

Source: own elaboration

4. Results

In this section we provide the results derived from the calculations of the previous indices. First, we focus on the global measures for the main indices, that is, they are calculated for the whole population defined in the previous section. Then, we replicate some of these calculations restricted to some subgroups of interest: the top income earners (in order to analyse if the upper tail of the income distribution exhibits a differentiated pattern in terms of mobility) and the subgroup of observations with unchanged marital status. (to isolate the mobility effect caused by getting married, widowed or divorced).

Using the indices shown in Table 3, the first measure we provide is the transition matrix for income quintiles (1999 and 2008). This is a usual way to analyse the sense of changes in income along the years, from a disaggregated point of view. In particular, we report a relative transition matrix using quintiles of income: the main diagonal shows the percentage of individuals remaining in the same quintile of income in both years; the values above the main diagonal reflect the proportion of people improving their relative income position, and the values below the main diagonal exhibit those individuals worsening their relative income position from 1999 to 2008.

Table 4. Transition matrix (%), all individuals

		Quintiles of income 2008					
		1	2	3	4	5	Total
	1	46.1	23.5	14.5	9.1	6.8	100
	2	22.4	30.0	25.1	15.6	6.9	100
Quintiles of income 1999	3	11.5	16.7	30.6	28.2	13.0	100
	4	5.4	7.0	15.5	40.0	32.2	100
	5	2.2	1.7	3.5	11.0	81.6	100
	Total	**15.6**	**14.6**	**17.4**	**21.5**	**30.9**	**100**

Source: own elaboration

There are several conclusions we may obtain from Table 4. First, we observe that the most common case is remaining in the same economic position from 1999 to 2008. But there are important differences: more than 80% of the richest individuals maintain their position between these ten years, but this proportion declines to half this value

for the rest of groups, being the middle income earners the most mobile. The most usual movement is upwards, increasing as quintiles do. The selected period is an increasing growth phase, so this is an expected result. The path of descents is the complete opposite: falling from a better economic position to a worse one is more common for the worse-off individuals.

If we compare these results with the ones obtained for the USA with tax data, they are not very different except for the top quintile: compared to the 81.6% immobility in Spain, Carrol et al. (2006) find a 53% immobility (for 1979-1995), Auten and Gee (2007) 67% (1986-1996) and Auten and Gee (2009) 69% (1996-2005). Although the periods differ and the methodologies are not identical, it seems that mobility for the top quintile is lower in Spain than in the USA[12].

If we now restrict the transition matrix to those individuals who did not change their marital status between 1999 and 2008, we may analyse the impact of marriages, divorces, separations and widowing in the overall mobility. We expect that the income mobility will be lower now, since it will be caused only by changes in individual income[13]. Table 5 show the results for this subgroup.

Table 5. Transition matrix (%), individuals with unchanged marital status

		Quintiles of income 2008					
		1	2	3	4	5	Total
	1	49.54	24.23	13.76	7.40	5.07	100
	2	22.75	31.56	25.71	14.47	5.52	100
Quintiles of income 1999	3	10.74	16.32	32.56	28.92	11.46	100
	4	5.16	5.79	14.96	42.8	31.29	100
	5	2.10	1.49	3.06	10.18	83.18	100
	Total	**16.17**	**14.69**	**17.46**	**21.3**	**30.37**	**100**

Source: own elaboration

Comparison between Table 4 and Table 5 can help us to isolate the marital status issue from the mobility matter. We confirm that mobility is lower, that is, the values of the

---

[12] Since Ayala and Onrubia (2001) did not use quintiles, we cannot compare here the results for Spain, but we do it for the top decile in the discussion of Table 8.
[13] When an individual marries another individual with different income, they are both likely to change their relative position. This effect is captured in Table 4, but removed in Table 5.

main diagonal are slightly higher[14]. Both movements up and down are lower for all quintiles. This means that some of the movements in Table 4 are due to changes in the marital status, and not changes in individual incomes. Both kinds of movements have effects on people's welfare, but if we are interested in the determinants of mobility, the reasons are obviously different.

Derived from the transition matrices, the relative measures in Table 6 summarize the previous information.

Table 6. Relative measures*

Index	All observations	Unchanged marital status
Shorrocks (Ms)	0.66460941	0.63427985
Immobility Index (II)	0.46831247	0.49257612
Mobility Ratio (MR)	1.62726940	1.65354520

* In order to calculate the indices the values for each row of the transition matrix have been weighted to account for the different size of each 1999 quintile.
Source: own elaboration

The Shorrocks index in our case is slightly low, according to precedent works for Spain (with values generally in a range from 0.70 to 0.80). This mobility index reports a lower value (0.66 and 0.63, respectively for the general case and the case of unchanged marital status) than most of the results obtained in previous studies: 0.63 (in 1985) to 0.71 (in 1992) (Cantó, 2000), 0.75 in the short term and 0.60 in the medium term (Ayala and Sastre, 2005), 0.729 to 0.753 in a year to year comparison (Bárcena and Moro, 2013), 0.47 in the short term and 0.81 in the long term (Ayala and Onrubia, 2001). The degree of mobility (complementary of the Inmobility Index) is also slightly smaller in our case (0.53 and 0.50, respectively) to those values reported in other works: 0.66 (Bárcena and Moro, 2013), 0.60 (Cantó, 2000). Finally, the mobility ratio shows values above 1. It means that upward movements are higher than downward transitions, specifically more than 60% higher.

If we remove the effect of changes in marital status (right column of Table 6), we observe that the Shorrocks Index is lower than in the general case. This implies less mobility when marital status remains unchanged, as it was explained before. The

---

[14] However, these differences seem to be much lower than in the US (see Auten and Gee, 2007).

Mobility Ratio is higher, showing more upward mobility in relation to downward mobility when changes in marital status are not taken into account.

Finally, Table 7 show some usual absolute and correlation measures. In order to check possible distortions caused by very high income earners, we make the calculations twice: one for all the observations and another one removing the top 1%. Trimming[15] is a usual practice in this kind of studies, to avoid the influence of extreme values. In general, trimming is done in both tails of the distribution, but in our case the problem only arises in the top income due to the characteristics inherent to tax data. Table 7 shows the results.

Table 7. Absolute and correlation measures

Index	All observations	1% top trimming
Fields and Ok (F-Ok)	0.57472094	0.55856079
Correlation	0.16705007	0.66638920
Hart Index (MHart)	0.47726294	0.50535158
Regression slope	0.51799787	0.62137178

Source: own elaboration

The higher effect of the trimming can be seen in the correlation coefficient, which increases from a very low correlation of 0.167 to 0.666. This is due to the large variance increase introduced by the top 1% incomes, which in some cases are thousands of times higher than the average income. Differences in the other indices are lower due to the smoothing induced by logarithms. Compared to the already cited Spanish works, the result (with trimming) of Fields and Ok is higher (around 0.30 in other works), the correlation is in the bottom range (0.66 vs. 0.60 to 0.80), the Hart index is higher (0.50 vs. 0.29 to 0.40) and the regression slope is inside the expected interval of results (0.60 to 0.75).

With all these results, we may conclude that mobility in the considered period is lower than the one reported in other studies for the same period using surveys. Although additional analysis are needed to determine de causes, the use of tax records may recude mobility because of the bias towards individuals with stable incomes.

---

[15] We perform the trimming just for the absolute indices, as the effect of the top 1% is not remarkable for the relative indices.

To close this section we make use of one of the main advantages of tax data: the availability of good data for the upper tail of the income distribution. Table 8 shows transitions for top income earners. We divide de population in two (first 4 quintile vs. top quintile) and offer a further breakdown of the movements within the top quintile (for the top 10%, 5% and 1%).

Table 8. Transition matrix (%), top income earners

		2008				
		First 4 quintiles	Top quintile (top 20%)	Top 10%	Top 5%	Top 1%
1999	First 4 quintiles	84.2%	15.8%	5.5%	2.4%	0.5%
	Top quintile (top 20%)	18.4%	81.6%	55.3%	30.4%	7.1%
	Top 10%	11.1%	88.9%	74.5%	48.9%	12.9%
	Top 5%	8.7%	91.3%	83.0%	69.3%	23.0%
	Top 1%	7.7%	92.3%	87.8%	81.9%	57.0%

Source: own elaboration

If we analyse the main diagonal we see that immobility is very high. More than 80% of the observations starting in the top quintile are still there at the end of the period, but the figures are also very high for the three sublevels analysed (even in the top 1% the immobility is higher than 50%). Although Ayala and Onrubia (2001) do not show the transition matrices, they indicate that the immobility for the top decile is 59% for 1984-1994, clearly a lower value than the 74.5% in Table 8[16]. This value is also lower for the USA in Auten and Gee (2007) (60%) and Auten and Gee (2009) (61.7%). Regarding the top 1%, immobility is also higher in Spain (57%) than in the USA: 30% in Carrol et al (2006), 46% in Auten and Gee (2007) and 41% in Auten and Gee (2009).

5. Conclusions

---

[16] However we must be careful with this comparison, because the deciles in Ayala and Onrubia where constructed taking into account only the observations present in all years (pure Panel), not all the observations for each year.

This paper is the first attempt of using the Spanish PIT Return Panel (1999-2008) for analysing income mobility. As one could expect from data coming from tax records, it is a good dataset for analysing mobility across middle and high income earners, but not as good for low income earners. We calculate a common set of indices to measure the mobility experienced in Spain from 1999 to 2008, taking the usual methodological choices. We provide three types of results: for the whole population, restricted to those individuals who do not change their marital status, and for the top income earners.

As a general rule, our results show less mobility than other studies that used survey data, even lower than the sole study that used tax data previously in Spain, and closer to the results of other studies using tax data. Further research should be carried out to determine the precise causes of the differences, which may related to real changes or just to differences related to the type of data, methodological issues or the exact definitions of the matrices and indices used. Besides, our results show that mobility can report different values if we isolate the effect of changes due to marital status. Finally, including the top income earners may have effects in the mobility results. Isolating them can be a good strategy to study their behaviour, usually different to the rest of the population.

References

Auten, G.E.; Gee, G. (2007). "Income mobility in the U.S.: evidence from income tax returns for 1987 and 1996", *U.S. Department of the Treasury OTA Papers,* 99.

Auten, G.E.; Gee, G. (2009). "Income Mobility in the United States: New Evidence from Income Tax Data", National Tax Journal, LXII(2):301-328

Ayala, L.; Onrubia, J. (2001). "La distribución de la renta en España según datos fiscales", Papeles de Economía Española, 88:89-112.

Ayala. L.; Sastre, M. (2005). "La movilidad de ingresos en España", Revista de Economía Aplicada, 38(XIII), 123-158.

Bárcena-Martín, E.; Moro Egido, A.I. (2013). "Movilidad de ingresos en España: el efecto de la crisis", Papeles de Economía Española, 135.

Cantó, O. (2000). "Income mobility in Spain: how much is there?", Review of Income and Wealth, 46:85-102.

Carroll, R.; Joulfaian, D.; Rider, M. (2006). "Income Mobility: The Recent American Experience", Andrew Young School of Policy Studies Research Paper Series 07-18.

Cervini-Plá, M. (2011) Intergenerational earnings and income mobility in Spain. Munich Personal RePEc Archive, nº 34942

Fields, G.S.; Ok, E.A. (1996), "The meaning and measurement of income mobility", Journal of Economic Theory, 71(2): 349-377.

Fields, G.S.; Ok, E.A. (1999), "Measuring movements of income", Economica, 66(264): 455-471.

Gil, M., De Pablos, L., Martínez, M. (2010) **"**Los determinantes socioeconómicos de la demanda de educación superior en España y la movilidad educativa intergeneracional", *Hacienda Pública Española/ Revista de Economía Pública,* nº 193 (2/2010) pp. 75-108.

Gittleman, M.; Joyce, M. (1999). "Have Family Income Mobility Patterns Changed?" Demography, 36(3): 299–314.

Gradín, C., Cantó, O., Del Río, C. (2008) "Inequality, poverty and mobility". Investigaciones económicas, vol xxxii (2)

Lillard, L.A., and R.J. Willis. 1978. "Dynamic Aspects of Earnings Mobility." *Econometrica* 46: 985-1012.

Onrubia, J.; Picos, F.; Pérez, C. (2011). Panel de Declarantes de IRPF 1999-2007: diseño, metodología y guía de utilización, Instituto de Estudios Fiscales.

Onrubia, J.; Picos, F.; Pérez, C.; Gallego, C.; González, M.C.; Huete, S. (2012). "Panel de declarantes de IRPF 1999-2008: Metodología, estructura y variables", Documentos de Trabajo del Instituto de Estudios Fiscales, 12/2012.

Pascual, M. (2009) "Intergenerational income mobility: The transmission of socio-economic status in Spain". Journal of Policy Modeling 31, pp. 835–846.

Prais, S.J. (1955) "Measuring Social Mobility." *Journal of the Royal Statistical Society,* Series A, 118:56-66.

Sawhill, I; Condon, M. (1992). "Is U.S. Income Inequality Really Growing?: Sorting Out the Fairness Question", Policy Bites. Washington, D.C.: Urban Institute, 1992.

Schiller, B. (1977) "Relative Earnings Mobility in the U.S.." *American Economic Review* December: 926-941.

Shorrocks, A.F. (1978b): "The measurement of mobility", *Econometrica*, vol. 46, págs. 1013-1024.

Shorrocks, A.F. 1981. "Income Stability in the United States." In N.A. Klevmarken and J.A. Lybeck (eds.), *The Statics and Dynamics of Income.* Tieto: Clevedon.

Shorrocks, A.F. (1993). "On the Hart measure of income mobility", in Casson, M. abd Creedy, J. (eds.). Industrial Concentration and Economic Inequality, Edward Elgar.

Splinter, D.; Diamond, J.; Bryant, V. (2009). "Income Volatility and Mobility: U.S. Income Tax Data, 1999-2007" Proceedings of the 102nd Annual Conference of the National Tax Association.

US Department of the Treasury, Office of Tax Analysis (1992a). "Household Income Changes over Time: Some Basic Questions and Facts", Tax Notes, 56:1065–1074.

US Department of the Treasury, Office of Tax Analysis (1992b). "Household Income Mobility During the 1980s: A Statistical Assessment Based on Tax Return Data", Tax Notes, 55 (special supplement)