# UNIVERSITAT JAUME I

# INCOME INEQUALITY OF THE SPANISH FAMILIES; AN EMPIRICAL ANALYSIS

**Marta Martinez Monterroso**

**Grado de Economía, group C**

**al185378@uji.es**


**Tutor: Simone Alfarano**

Castellón de la Plana, June 2016

# INCOME INEQUALITY OF THE SPANISH FAMILIES; AN EMPIRICAL ANALYSIS

**Abstract**

This paper analysed the evolution in inequality in income of the Spanish Households between 2002 and 2011. For that, it has been used Financial Survey of Household data. The aim of this paper is to analyse how some variables such a level of education of household head, the age of household head and the number of family members (family) affect the income of household. For that, we compute the variance of income, the Gini coefficient, the P50/25 ratio, the P90/50 ratio and the variance of explanatory variables (education, family and age). Our results show that education variable has the highest variance. Education variable is which explains a highest percentage of income variance.

**Keywords:** EFF, income, inequality, Spain, families.

**INDEX**

# FIGURES AND TABLES INDEX

## I. Introduction

The distribution of income is one of the keys to know the existence of a lack in the economic development of a country. It is important to understand the dynamics of economic growth as well as the welfare of the population. This is critical for the implementation of public policies to promote economic development.

There is an increasing interest in equity in income distribution and the role that public policies have about it, which have been recovered from different areas and organizations around the world. The critical economic situation which currently spends our country, the fall in aggregate demand, deterioration of family income and the difficulties of survival of businesses are some reasons that makes that researchers are more interested in income distribution and economic growth.

The process of economic globalization along with the severe and prolonged recession in many developed countries is leading some deep change patterns of income distribution. Therefore, the Bank of Spain decided to carry out the realization of the "Financial Survey of Households" (EFF) for the purpose of analyse the investment and finance of Spanish households in order to know the distribution of family income.

With a view to understand the inequality evolution of income distribution, the aim of this paper is to analyse how some variables such as the level of education of the household head, the age of household head and the number of family members affect the income of Spanish households. The study is based on the period from 2002 to 2011. Throughout this work it has been shown that the most predictive variable first of the three variables studied have the greatest impact when having a higher or lower household income is the level of education that has the head of household. Both the age of the household head as the members of each family are less important variables when a family has a higher or lower level of income.

The income inequality is a problem which affects to the Spanish population and it causes discomfort in society. The issue is really important for the country, since the existence of inequality is an important and significant problem for the Spanish population.

In this study it has been not possible to find comparable data sets, (surveys of different households were started but were interrupted over the years), and as a consequence

of that, the period investigated is relatively small and it covers four years from 2002 to 2011. But despite having a single data set, it has been possible the analysis of income distribution using raw income and normalized income according to the scale of the OECD income, which is explained in the results section.

Another problem is that exist a limitation in finding a high rate of answers of the surveyed households, because people answer to that kind of surveys according to their interests, or sometimes because they have not enough knowledge to answer so optimal. And this means that the information can be biased. There is also a problem in surveys to families with higher incomes, since there are very few families of them that respond to such surveys. For this, it is included an oversampling in order to have a representative sample of the population.

This research is divided into six sections, besides this first introductory section. The second section is a brief theoretical framework about the income distribution. In the third section it is explained a brief review of the literature about income distribution. The fourth section contains the analysis of the development of the Spanish Financial Survey of Families (EFF) and also a general explanation of the data set as well as the definition of the variables used. The fifth section presents the results obtained and the evolution of inequality in the period studied. The following section explains the study done by Piojan-Mas and Marcos Sanchez (2009), which has been used to carry out this paper. Finally, in the last section the conclusions of this paper are presented.

## II. Conceptual framework

Atkinson (1997) was one of the economists who started to be interested in this distribution. He created an index to measure income inequality which can be measurable through the ε coefficient in order to weigh the incomes. It is used the following equation:

$$A = 1 - \frac{1}{\mu} \left( \frac{1}{N} \sum_{i=1}^{N} y_i^{1-e} \right)^{1/(1-\varepsilon)} \quad \forall \epsilon : \varepsilon \in (0, \infty)$$

$$\boldsymbol{A = 1 - \frac{1}{\mu} \left( \prod_{i=\_1}^{N} y_i \right)^{1/N} \; para \; \varepsilon = 1}$$

Where Yi is income per capita (i = 1, 2, ..., N) and $\mu$ is the average income.

Other authors who have written about the history of inequality have been Cowell (2000) and Foster (1997). They thought that revenues can be seen as an indicator of economic welfare or control over resources. But there are cases where this variable is not the best measure of inequality, since consumer spending and wealth are also appropriate variables for this. The economic analysis of income distribution is interpreted in two different ways, one is the functional distribution of income and the other one is the distribution of income among people.

The functional distribution of income is an integral part of the economic analysis of relative prices, production and employment. The analysis focuses on basic economic concepts. In other words, it is based on the use of production factors of land, labor and capital and rates of remuneration for their services rent, salaries and benefits. Prices can be adjusted in order to ensure balance in all markets.

The distribution per capita is also a part of the prices analysis, since the household savings, investment in human capital and children education are determined by the price. In order to determine the household incomes is necessary to know the owner of the natural resources, the capital assets and corporate profits as well. Family budgets are determined by market prices and property rights, both of them are necessary taken into account in order to determine the revenue in the long run.

Many economists have questioned the importance of different aspects of the standard account in the income distribution, as explained above. You can focus on two points,

the first one is the role of prices, in which the economists are influenced by the approach of Keynesian theory. They have developed a range of alternative theories to the functional income distribution using some components of Keynes's theory[1]. The main features of these theories are saving decisions of capitalists and workers in order to achieve a combination between labor and capital. The second is the power of monopoly, restrictions of competition in the form of market segmentation and price discrimination are important for analysis of the size distribution of profits. It has also been considered for monopoly pricing functional income distribution[2].

There are different tools to represent and interpret the income distribution of population. One of them is the Lorenz curve which represents a function through a diagram. The function is constructed by a classification of the income level of the individuals in descending order.

**Figure 1:** *Lorenz Diagram. US income before tax 1987 and 2003*
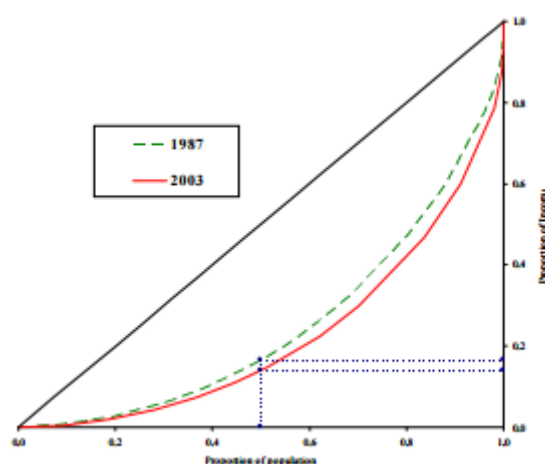


Source: Cowell (1997) Income Distribution an inequality.

Figure 1 shows an example of Lorenz curve of UE. X-axis represents the proportion of the population and axis Y represents the proportion of income. The figure shows the curve for two different years: 1987 and 2003. It is possible to observe that in year 1987, a 50% of the population own more than a 15% of the total income.

---

[1] Economic theory of Keynes (1936), based on analyzing the consequences and causes of variations in aggregate demand and the relationship with the level of income and employment. By formulating the multiplier intended to explain why an economic system can remain in balance in times of crisis or recession.

[2] Klecki (1939), has an important role in the Marxist concept of exploitation, where capitalists exploited workers by appropriating part of their work, and the theory of distribution based on the fight enter classes representing different factors production.

The Lorenz curve have two main different characteristics, it can be exist perfect equity, i.e., all the families have the same level of income. In this case, the Lorenz curve runs into the middle of the diagonal. The other characteristic is the perfect inequality, in that situation, only one person is the owner of total income. In this situation the Lorenz curve runs into the X-axis ascending vertically to (1, 1) point. Consequently, Lorenz curves closer to 45º diagonal mean more equitable distribution of income.

Another tool that measures inequality of income is the Gini coefficient. Through the calculation of this coefficient it is obtained a value from 0 to 1. Values closer to 1 mean that the income distribution is more unequal, in other words, there are households which own higher income and households with lower levels of income. On the other hand, when Gini coefficient approximates to 0, the income distribution is more equitable. In the extreme case in which the coefficient is equal to 0, the interpretation is that exist perfect equality. The Gini index is calculated as a proportion of the areas of the Lorenz curve diagram. The area between the perfect line of equity and the Lorenz curve is called a, and the area below the curve is called b. Hence, the Gini index is a/(a+b).

*III. Literature review*

Inequality covers the differences in the distribution of assets or income among individuals or groups. In recent years, many economists have been interested in studying how inequality is generated and how it follows reproducing over time. For this reason some researchers focus on the study of income inequality with regard to economic variables and observing how these variables affects to income inequality over time.

In the literature about the study of income inequality it can be found the research of Budría and Diaz-Gimenez (2006), who used in their study the panel data of European families from 1998 about the distribution of income in Spain, the income from work[3], capital income[4], transfers[5] and income[6] using economic variables such as age, employment, education and marital status. In the study conducted also they took into account the income inequality in other European countries. They concluded that income inequality capital in Spain was far from the European average. For the data analysis they used the EUROPANEL[7].

There are some authors, for instance, Heathcote, Perri et Vilanten (2009) who also consider the inequality in their paper, in which they use the following variables: wages, hours, income, consumption and wealth. Their research is based on data from the CPS (Current Population Survey), PSID (Panel Study of Income Dynamics), CEX (Consumer Expenditure Survey), SCF (Survey of Consumer Finances), NIPA (National Income and Produce Accounts Personal Income) in order to understand how the different measures of inequality are related through markets and institutions. This study is concluded saying that inequality in wages was increased in 1970. In respect of income inequality, hours worked were reduced as well as unemployment was increased, hence greater inequality. Household income increases slower than inequality of income of the individual who receives it. By PSID they have noted the importance of permanent and transitory shocks. Where detect an increase in permanent variation and inequality in consumption is maximized. To varying measure

---

[3] Defined as the sum of net income from both wage employment and self-employment.
[4] Defined as the sum of net capital inflows and net property income.
[5] Transfers defined as the sum of private and public transfers (pensions, benefits, disease, and subsidies).
[6] Income defined as the sum of labor income, capital income, and transfers.
[7] It is a survey that is conducted in the European Union and allows obtaining comparable information from all member states on income, labor, employment, poverty, health, and other social indicators. Defines a household as a group of people sharing the same house.

wealth inequality by the Gini coefficient[8] according to data from SCF and note that increases between 1983-2007.

There are also authors who measure income inequality in Spain between 1998-2010 such as Bonhome and Hospido (2012). They made the estimation with social security data. In their work they have evaluated the effect of cyclical variations[9] that Spain has experienced in recent decades on the evolution of income. To relate the evolution of employment and income they use a multisectoral model in which an increase in construction demand leads to an increase in employment and income in this sector, while a drop in demand has the opposite effect. Due to the situation in Spain during 1997-2007, the increase in employment and incomes in that sector caused the decline in global inequality.

This study was based on the study of D. H. L. F. Katz, and M. S. Kearney (2005), to measure income inequality using the workforce and prices. They found that the effects of both variables contributed to the decline in inequality during the boom years (2007-2010). Bonhome and Hospido have used records of individual earnings. Since 1998, the database used records of permanent, temporary and fixed contracts. Based on this, they note that the income gap of workers with these types of contracts decreased in the period of growth.

---

[8] It is a measure of inequality, is used to measure income inequality, this coefficient is a number between 0 and 1, where 0 corresponds to perfect equality (everyone has the same income) and where 1 corresponds to perfect inequality (one person has all income and others none).
[9] Income inequality and unemployment from 1990 to 2010.

## IV. Data set

The Bank of Spain, in order to carry out various studies about the investment and financing of households and in order to obtain direct information on the situation of Spanish families has launched a survey "Financial Survey of Households" (EFF).
This kind of surveys is used for to analyse the income evolution and to know the information about the financial decisions related to the expenditure, consumption, debts and assets of the Spanish families. In this paper, it has been used the households' incomes of the years available in the Bank of Spain: 2002, 2005, 2008, 2011.

Following these studies is possible to obtain better information about the current economic situation of the households. Based on this data it is possible to understand how our Economy works and as a result of that, it is workable to improve the design of the public policies. A minor problem which involves this type of surveys is that they are relatively new. Consequently, the data obtained are not from a long period which makes difficult a further investigation.

It has to be emphasized the importance of a big characteristic which differences this kind of survey from others which is the necessity of a large sample in order to obtain a result as representative as possible. For that reason, in the survey used in this study, there is an oversampling of the Spanish households with the greatest wealth since its distribution is very asymmetric (not all population have the same fraction of assets).

It should be noted that an important drawback of these surveys is the fact that people are not always honest with their personal information. There are many different kinds of families with different purchase power. Therefore, most of them could answer attending to their own interests or perhaps they do not tell the truth in their answers because they do not really know this information about themselves. The oversampling was possible thanks to the collaboration of National Institute of Statistics (INE).

The aim of this study is to analyse how a group of social-demographic variables affect to the household labor income of the Spanish families. The dependent variable studied in this paper is income which is composed by labor and non-labor incomes. Firstly, the mean and median of income variable are analysed. It is important to note that income is deflected with Consumer Price Index (CPI), 2011 used as the base year.

**Figure 2: *Comparison between evolution of the real mean income and evolution of the real median income***
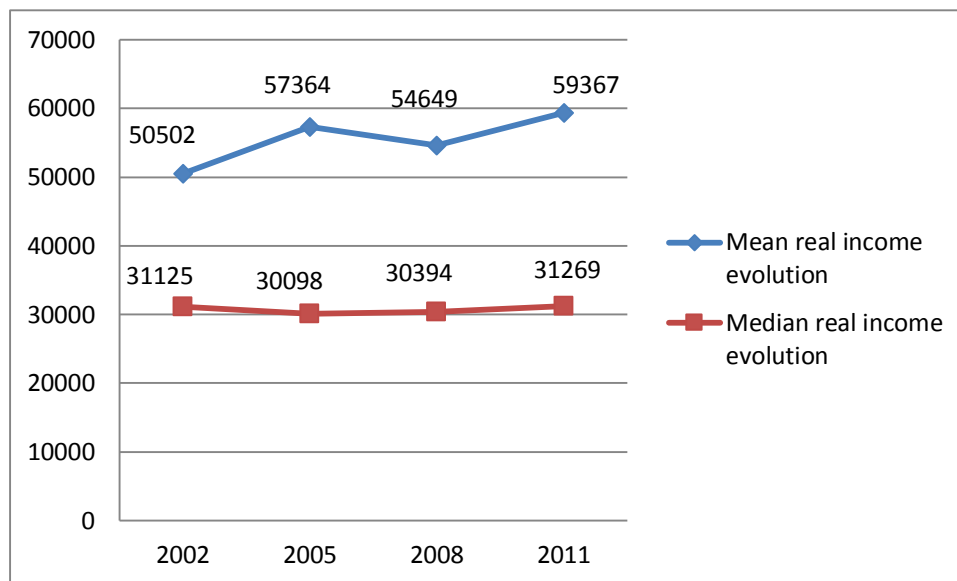
Figure 2 shows the evolution of income mean of the Spanish families according to the date obtained in the Financial Survey of Households (EFF) and taking into account the CPI in each year. And also shows the evolution of the median of the income in function of that data. In Spain the mean of income in year 2002 is 50,502€ and this value increase in year 2005 up to approximately 7000. Even though in 2008 this value decreases, it reaches its highest value in 2011 which is 59,367€. The median income means that a half of the households own an income lower than this median value, whereas the other half has a higher income than the median value. The median value in year 2002 is 31,125€ and it increases to 31,269€ in year 2011, while in years 2005 and 2008 this value is approximately 30000€ (€ is a real value to 2011as base year).

Referring to the sociodemographic variables, in this paper it has been studied the education variable which refers to the level of education of the household head in each Spanish family. The variable takes values from 1 to 3: the value 1 refers to the case where the household head has less than high school level. When it takes value 2 it means that the household head has high school studies. Lastly, the value 3 means university studies.

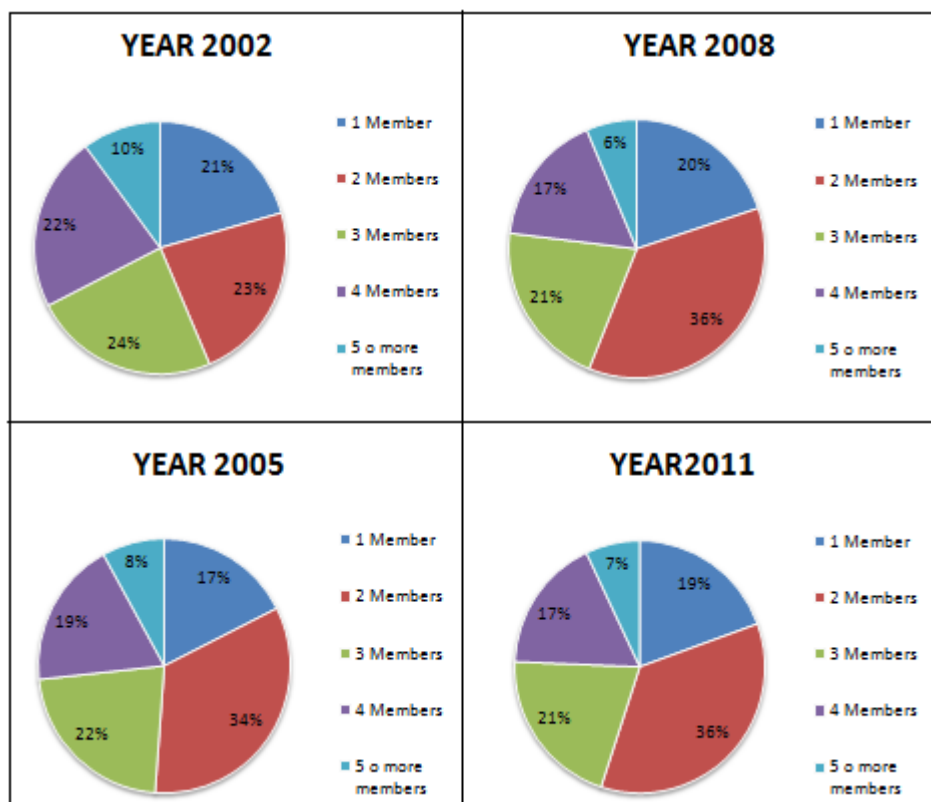**Figure 3: *Representation of education level of household head in Spanish families during the period 2002-2011***
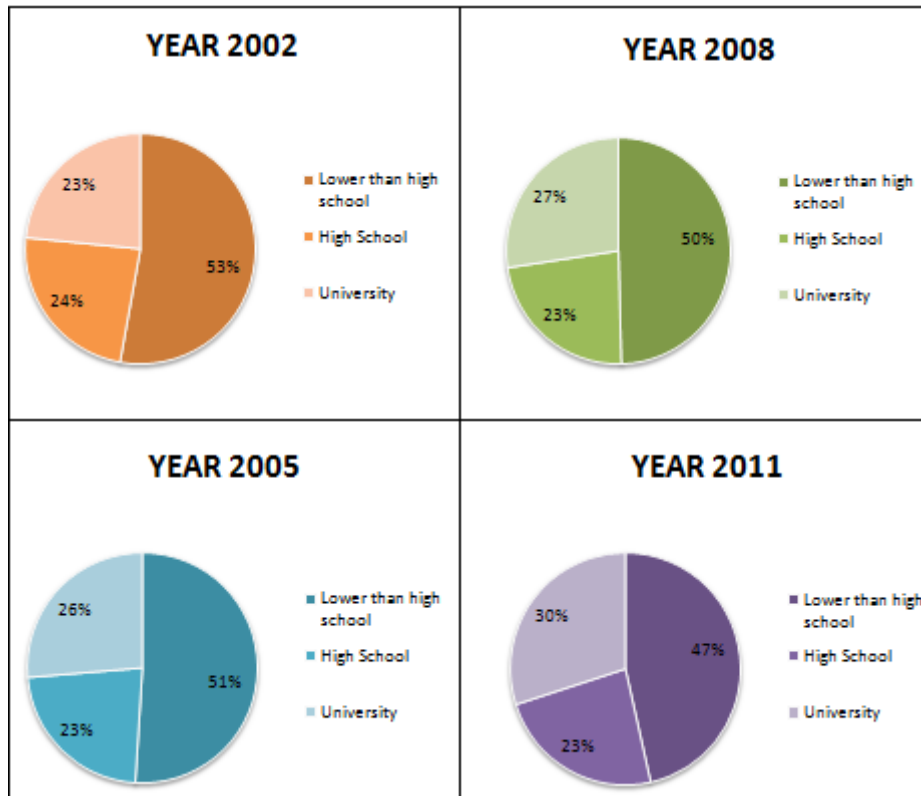
Figure 3 shows the percentage of different levels of education mentioned previously referent to the household head. Along the period studied, the highest percentage is obtained for head households that have highest education (university). On the other hand, attending to studies lower than high school, the percentages obtained are more equitable between the first three years. It is possible to note that the highest difference happens in year 2011, where there are less people who have university studies, and there are more people with studies lower than high school.

Following with the study of family variable, households are divided attending to the number of family members. This variable takes value from 1 to 5 depending on this number. For example; in a household composed by 1 member, this variable would take value 1, in another one composed by 2 members, this variable would take value 2, and so on. In this data set it is not taken into account the relationship between the family members.
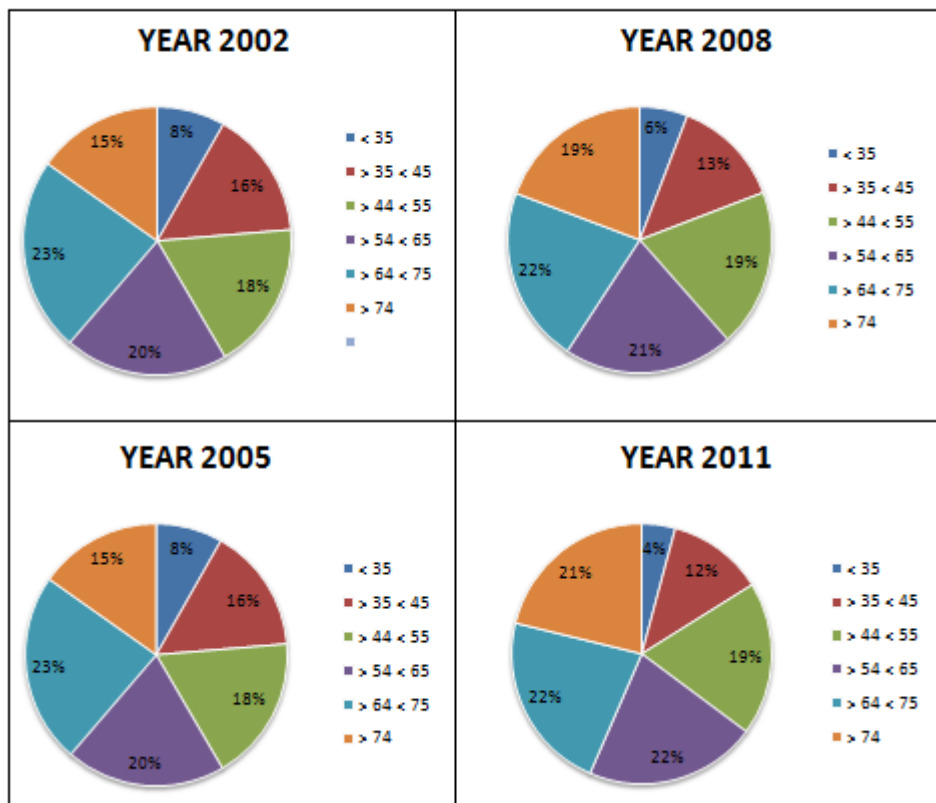
Source: own elaboration data from EFF.

Figure 4 shows the percentage of members in different families during the years analysed. In 2005, it can be seen an increase of 11% in the families with two members compared to 2002. Although, there are no major differences in the other cases in these two years. In 2008, still a small increase of 2% in families with two members and 3% in families with one member compared to 2005. In the last year remains the same percentage in families with two members and the rest of numbers of members remains the same as previous years.

Finally, the last sociodemographic variable that is studied in this paper refers to the age that has the household head in every home. It takes values from 1 to 6. The value 1 represents if age is less than 35, and it takes the value 2 when the age of the household head is more than 35 years but less than 45, it takes the value 3 when the household head has more than 44 years but less than 55, it takes the value 4 when the age is more than 54 years but less than 65 years, it takes the value 5 when the age is more than 64 years and less than 75 years and it takes a value bigger than 5 when the age of head of household is more than 74 years.

Figure 5 shows the percentages depending on the age of household head of every household in Spain. When the household head has less than 35 years old, there is a decrease in 2002 and 2005 compared to other years. In 2002 and 2005, when the age variable takes value 2, the percentage of 16% is maintained; the head of household has more than 34 years but less than 45 years. Whereas, in years 2008 and 2011 this percentage decreases. Another value that stands out is when is taken a value greater than five. In 2002 and 2005 there are fewer family members who are over 74 years compared to later years. In 2008 and 2011 there is an increase in this percentage, there are more household heads that are older than 74 years.

## V. Results

This section describes the income evolution of Spanish households from 2002 to 2011. For that, it has been used Financial Survey of Household data (EFF). In order to describe the evolution of inequality, it has been used a number of sociodemographic variables and a dependent variable which is income. To analyse this evolution, first of all it has used raw data. Secondly, labor income of households has been normalized according to the scale of equivalence of the OECD[10]. Income has been scaled depending on the number of family members of each household. It should be noted again that income is deflated using the Consumer Price Index (CPI).

It is necessary to do that since a household grows with each additional family member not in a proportional way, but using equivalence scales is possible to assign a value for each member of the population in proportion to their needs. The factors taken into account to allocate scale values are household size and age of family members.

It is used a logarithmic regression of household income $y_{i,t}$, Referent to the explanatory variables are used two dummies; one of them for education variable $D^e_{i,t}$, and the second one for the variable family (number of family members) $D^f{}_{i,t}$ , and finally it is used a polynomial[11] of age variable $f_t\left(A_{i,t}\right)$.

$$\ln y_{i,t} = D_t + \beta^e_t\, D^e{}_{i,t} + \beta^f_t D^f{}_{i,t} + f_t\left(A_{i,t}\right) + \varepsilon^y{}_{i,t}$$

Finally, it is discomposed the variance of each of the variables in the regression. Then, it is studied the residual variance, $var_t\left(\varepsilon^y_{i,t}\right)$, the variance of the education variable $var_t\left(\beta^e_t\, D^e_{i,t}\right)$, the variance of family variable, $var_t\left(\beta^f_t\, D^f_{i,t}\right)$ and the variance of age variable , $var_t\left[f_t\left(A_{i,t}\right)\right]$.

As it has been mentioned previously, firstly the results were analysed using raw data, using a regression for each year. The equation (1) is the regression for 2002. The equation (2) represents the data for year 2005. In equation (3) it is represented for year

---

[10] Organization for Economic Co-operation and Development is an economic organization founded to stimulate economic progress and world trade among countries that compose it.
[11] Is used because the variable age can be considered a parabola, age has an impact on income, it is natural that the rent increase with age, but there comes a point where age is high but it cannot work, so with retirement is natural that income back down. And so the quadratic term is put into variable age dependency makes it a parabola if the second term is negative.

2005. For last, in equation (4) it is represented the regression for year 2011. Then, it is going to analyse these four regressions jointly. In all regressions the same variables are estimated in order to see the effect on income.

$$(1) \ log \ \hat{Y} = 9.5136 + .8536 \ D^e + .5001 \ D^f + .5290 \ A - .0838 \ A^2$$
$$(.0613) \quad (.0263) \quad (.0389) \quad (.0366) \quad (.0049)$$
$$n = 5.129 , R^2 = .275, F \ (4, 5124) = \ 486.34$$

$$(2) \ log \ \hat{\hat{Y}} = 9.3934 + .9551 \ D^e + .4034 \ D^f + .5555 \ A - .0854 \ A^2$$
$$(.0643) \quad (.0267) \quad (.0439) \quad (.0384) \quad (.0052)$$
$$n = 5.910 , R^2 = .245, F \ (4, 5905) = \ 480.62$$

$$(3) \ log \ \hat{Y} = 9.5431 + .9293 \ D^e + .3504 \ D^f + .4590 \ A - .0713 \ A^2$$
$$(.0627) \quad (.0230) \quad (.0424) \quad (.0357) \quad (.0046)$$
$$n = 6.162 , R^2 = .271, F \ (4, 6157) = \ 573.79$$

$$(4) \ log \ \hat{Y} = 9.3901 + .9864 \ D^e + .3887 \ D^f + .4784 \ A - .0672 \ A^2$$
$$(.0726) \quad (.0235) \quad (.0428) \quad (.0402) \quad (.0051)$$
$$n = 6.075 , R^2 = .271, F \ (4, 6070) = \ 564.19$$

Referring to R-square coefficient, it is observed that in years 2002, 2008 and 2011 it is obtained 27%. This means that the explanatory variables explain 27% of the variance of the dependent variable income, and the rest is explained by the residue, i.e., 73%. In year 2005, the R-square value is 24%, therefore the explanatory variables explain 24% of the variance of income, and 76% is explained by the residue.

In order to prove if explanatory variables are signifinative predicting household income, it is used a multiple hypothesis test, i.e., hypothesis which have more than one restriction. In this paper it is studied F-statistic. To prove if null hypothesis is rejected, it is used a table of critical values of the distribution with a 5% of significance level.

Equation (1) has a F-statistic (4, 5124) = 486.34 and its critical value is 2.37, if F (4, 5124) > 2.37 at a significance level of 5% the null hypothesis is rejected and at least one of the explanatory variables is nonzero . Therefore it can be concluded that at least one of the explanatory variables is significant for the dependent income variable.

Equation (2) is a F-statistic (4, 5905) = 480.62 and its critical value is 2.37, so if F (4 , 5905 ) > 2.37 at a significance level of 5%, the null hypothesis is rejected in favor of the alternative hypothesis, where at least one of the explanatory variables is nonzero. It can be concluded again that at least one of the explanatory variables is significant for the dependent income variable.

Equation (3) has a F-statistic (4, 6157) = 573.79 and its critical value is 2.37 , so if F (4 , 6157 ) > 2.37 at a significance level of 5% the null hypothesis is rejected in favor of the alternative one, thus at least one of the explanatory variables is nonzero. It is concluded that at least one of the explanatory variables is significant for the dependent income variable.

Equation (4) is a F-statistic (4, 6070) = 564.19 and its critical value is 2.37 , so if F (4 , 6070) > 2.37 at a significance level of 5% the null hypothesis is rejected in favor of the alternative hypothesis, as a consequence of that, at least one of the explanatory variables is nonzero. It is concluded that at least one of the explanatory variables is significant for the dependent income variable.

To continue, it is analysed the regressions with normalized labor income according to the scale of equivalence. In equation (5), the regression for 2002 is shown. In equation (6) data for 2005 are presented. In equation (7) it is presented for year 2008. Finally, in equation (8) it is possible to see the regression for year 2011. To continue, a joint analysis of the four regressions is presented. In all the regressions are estimated the same variables in order to see the effect on income.

$$(5) \ log \ \hat{y} = 9.1767 + .8418 \ D^e + .1554 \ D^f + .4049 \ A - .0597 \ A^2$$
$$(.0570) \quad (.0244) \quad (.0362) \quad (.0340) \quad (.0046)$$
$$n = 5.129 \ , R^2 = .230, F \ (4,5124) = 384.66$$

$$(6) \ log \ \hat{y} = 9.0643 + .9433 \ D^e + .0610 \ D^f + .4311 \ A - .0617 \ A^2$$
$$(.0609) \quad (.0253) \quad (.0415) \quad (.0342) \quad (.0049)$$
$$n = 5.910 \ , R^2 = .219, F \ (4, 5905) = 414.53$$

$$(7) \ log \ \hat{y} = 9.1545 + .9165 \ D^e - .0128 \ D^f + .3820 \ A - .0539 \ A^2$$
$$(.0589) \quad (.0216) \quad (.0399) \quad (.0003) \quad (.0043)$$
$$n = 6.162 \ , R^2 = .253, F \ (4, 6157) = 521.69$$

$$(8)\ log\ \hat{y} = 8.967 + .9718\ D^e + .0333\ D^f + .4115\ A - .0507\ A^2$$
$$(.0688)\quad (.0223)\quad\ \ (.0405)\quad\ \ (.0381)\quad\ (.0048)$$
$$n = 6.075\,, R^2 = .260, F\,(4, 6070) = 533.46$$

The R-square in the case of normalized income is lower than in the case of raw income. For 2002 the explanatory variables explain 23% of the variance of the dependent variable income, and a 77% is explained by the residue. In years 2005 and 2008 the R-squared is the same, for both cases the explanatory variables explain 25% of the income variance, and the rest it is explained by the residue, i.e. 75%. For last, in year 2011 it gets the highest R-squared, 26% of the income variance is explained by the explanatory variables. And a74% is explained by the residue.

Furthermore, it is checked if the explanatory variables are significant in having the Spanish families a certain income. For that it is done a contrast of multiple hypothesis by observing the statistical F. Thus check if the null hypothesis is rejected or not using values critics of the distribution with a level of significance of 5%.

Equation ( 5) is a statistical $F\,(4\,, 5124\,) = 384.66$ and its critical value is 2.37 , so if $F\,(4\,, 5124\,) > 2.37$ at a significance level of 5% the null hypothesis is rejected in favor of the alternative hypothesis , where at least one of the explanatory variables is nonzero . It is concluded that at least one of the explanatory variables is significant for the dependent variable income.

Equation (6) has a statistical $F\,(4, 5905) = 414.53$ and its critical value is 2.37, yes $F\,(4, 5905) > 2.37$ at a significance level of 5% the null hypothesis is rejected, therefore the least one of the explanatory variables is nonzero. We conclude again that at least one of the explanatory variables is significant for equities.

Equation (7) has a statistical $F\,(4, 6157) = 521.69$ and its critical value is 2.37, so if $F\,(4, 6157) > 2.37$ at a significance level of 5% again reject the null hypothesis in favor of the alternative where at least one of the explanatory variables is nonzero. It is concluded that at least one of the explanatory variables is nonzero.

Equation (8) has a statistical $F\,(4, 6070) = 533.46$ and its critical value is 2.37, if $F\,(4, 6070) > 2.37$ significance level of 5% the null hypothesis is rejected. We can conclude

that at least one of the explanatory variables is significant for the dependent variable income.

Following with the study, it is represented graphically the different measures to represent income inequality; variance, Ginni coefficient and 50th- 25th and 90th- 50th ratio. The purpose of that is to observe how using this statistics it is possible to measure the evolution of inequity in raw and normalized form over time.

**Figure 6:** *Representation of variance to log raw and normalized income during the period 2002-2011*
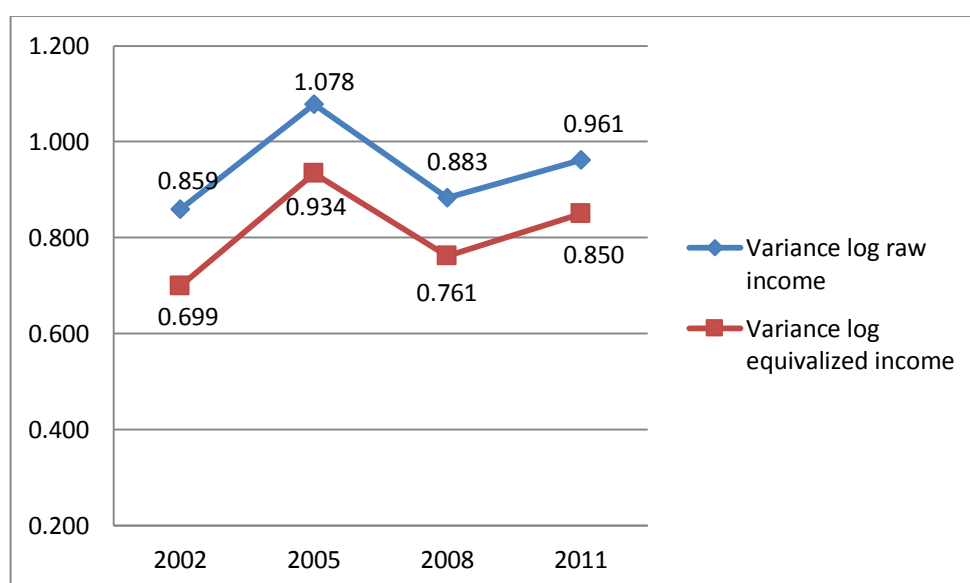
In figure 6 both as the variance of the logarithm of the dependent variable raw income as the logarithm of the normalized income is shown. The variance of raw income is higher than the standard income. But both follow the same distribution, from 2002 to 2005 there is an increase in inequality approximately 0.22. From 2005 to 2008 a decrease of approximately 0.18 inequality is observed. However, in the last two periods inequality rises again, but this increase reaches a lower value than in 2005.

**Figure 7: *Representation of Gini coefficient to raw and normalized income during the period 2002-2011***
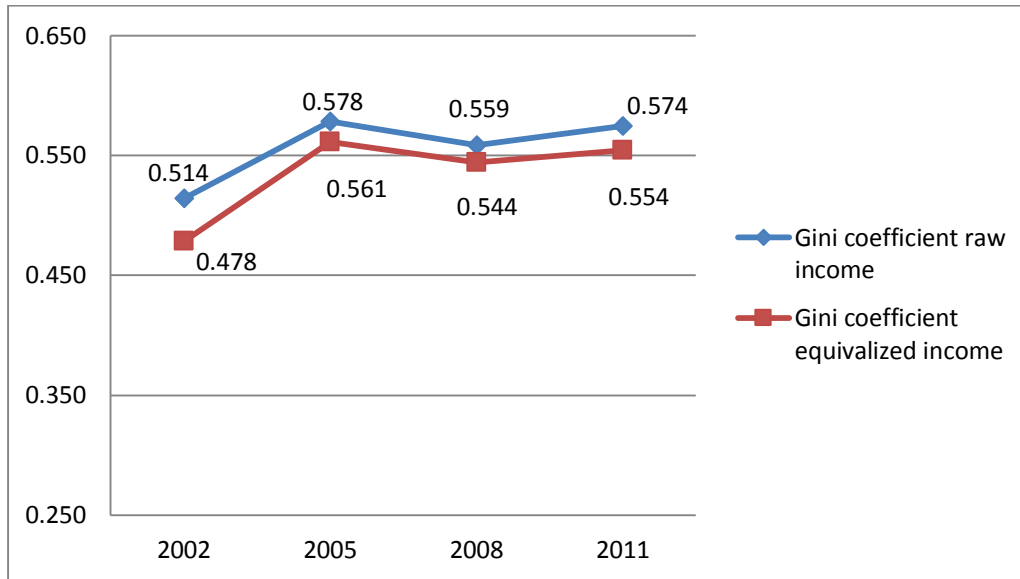
Figure 7 shows the evolution of Gini coefficient, which is represented for raw income and also for normalized income. It is obtained a higher coefficient for raw labour income. The highest levels of inequality were reached in years 2005 and 2011. With the study of raw income, inequality decreased in 2008. In year 2002 it reached the lower level. This value increased the remaining years, reaching in 2011 to almost the same level of inequality than in 2005.

In the case of measuring inequality of normalized income, the coefficients are lower but they have had a similar evolution to the raw income one, previously commented. In other words, the highest levels of inequality are reached in years 2005 and 2011. Inequality decreases from year 2005 to year 2008 and it reaches in 2011 almost the same coefficient than in year 2005.

**Figure 8: *Representation of the ratio between 50/25 percentile to raw and normalized income during the period 2002-2011***
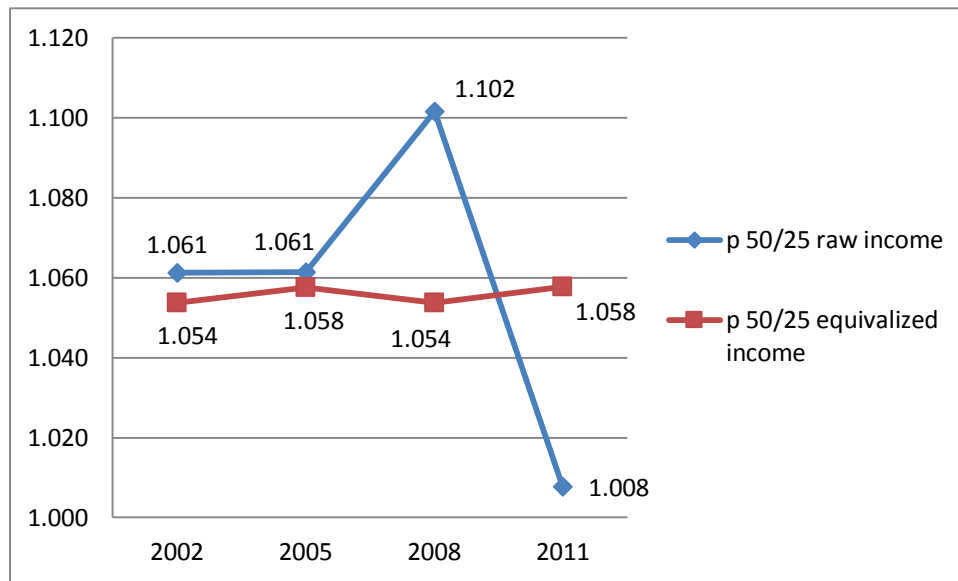
Figure 8 shows the evolution of 50th-25th percentile which is a statistical measure of inequality of income. The 50/25 percentile ratio means that half of the families have a higher value of income compared to the other half that have a value lower income. Thos ratio takes the lowest value with respect to the median (the P25 take the lower income). With the analysis of raw income, the percentile increases a 4.3% for the period from 2005 to 2008, in this moment in Spain the bubble bursts. But it decreases almost a 9% from 2008, due to the financial crisis. Therefore, this decrease of the value of the ratio means that there is less difference between the incomes that has the 50% of the population against income that has a 25% of the population.

Instead, when income is normalized the 50/25 ratio increases from 2002-2005 and during the period 2008-2011 a 0.4%. In the period 2005-2008 the ratio decreases. But in 2011 the ratio has a value greater therefore it can be concluded that there is a greater difference between the income that has 50% of the population compared to the income that has 25% of the population.

**Figure 9: *Representation of the ratio between 90/50 percentile to raw and normalized income during the period 2002-2011***
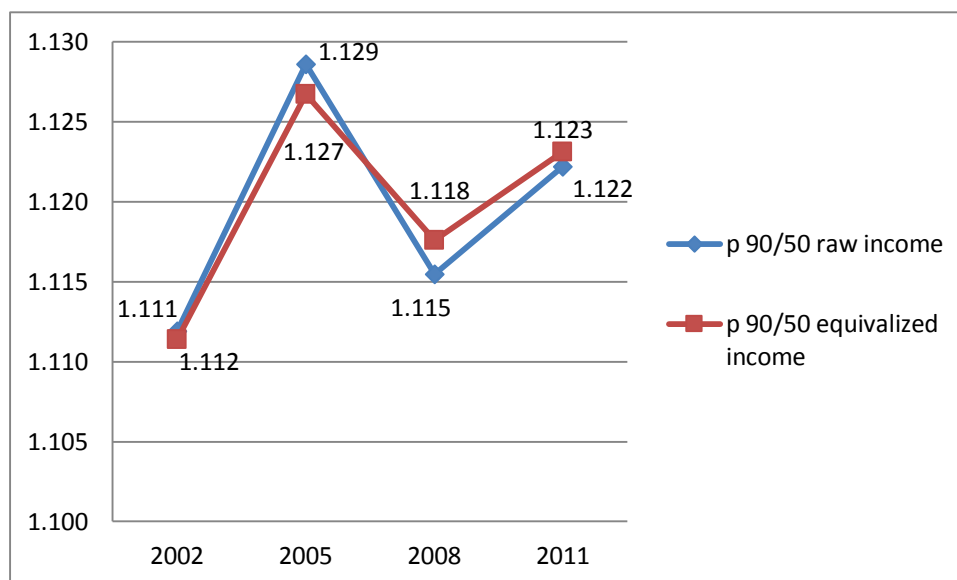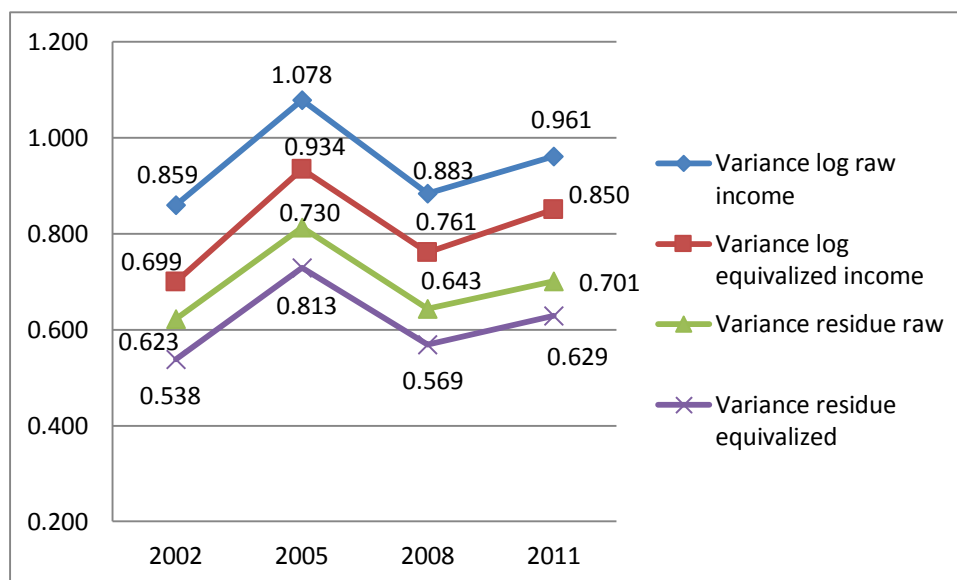


Source: own elaboration data from EFF.

Figure 9 shows the evolution of 90/50 percentile. It is another statistic to measure income inequality. The ratio of 90/50 means that 90% of the population have a higher income value compared to 50% of the population that has a lower income. Thos ratio takes the median with respect the highest incomes, (10% missing). With the analysis of raw data, it is possible to see that the ratio increases during the period 2002-2005 and decreases in the following one (2005-2008). Whereas, it reaches 1,122 points in 2011, thus an increase in the ratio means that there is a greater difference between the income that has a 90% of the population compared to income that own a 50% of the population.

However, the 90/50 ratio is analyzed taking normalized income, the 90/50 ratio change in a similar way to the ratio for raw income. In this case the ratio also increases in the period 2002-2005 and decreases in the next one. In the last year studied it reaches 1,123 points, so it can be concluded that this increase of the ratio means that there is a greater difference between the income that has a 90% of the population compared to which own a 50% of the population.

**Figure 10:** *Representation of the variance household net earnings during the period 2002-2011*

Figure 10 represents the evolution of the logarithm variance of raw income and normalized income, as well as the variance of the residue of both. The variance with raw income is higher than the variance with the normalized income, although both have the same evolution pattern: both increase in 2005 and 2011 and they decrease in 2008. Therefore, there is a decrease in inequality during the period 2005-2008 but in 2008 it begins to rise again reaching in 2011 a variance of 0.96 for the variance of the logarithm of raw income and 0.85 for the variance of the normalized income.

**Figure 11:** *Representation of the variance observable components household normalized income during the period 2002-2011*
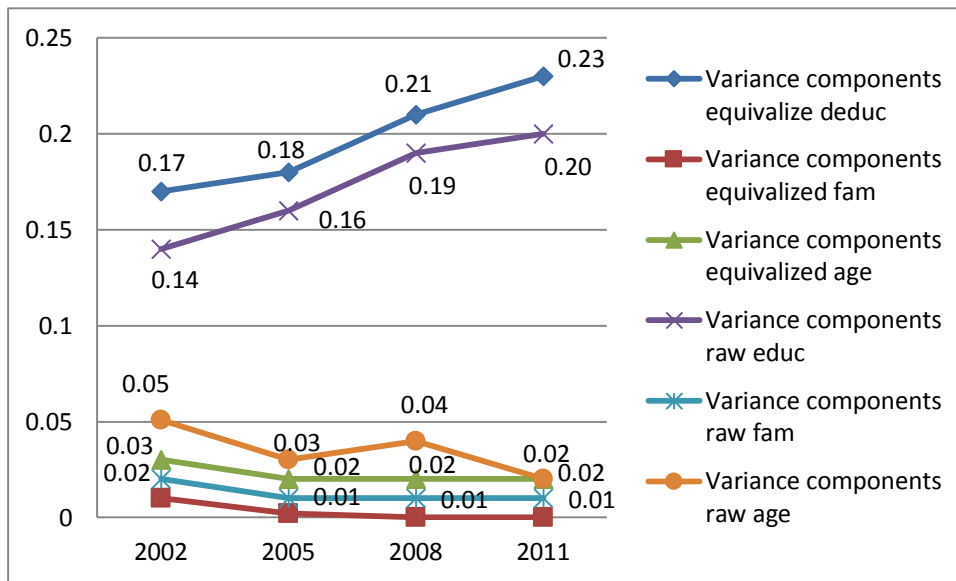
In figure 11 it is shown the variance of each of the explanatory variables in this work. The variance of each variable for raw income and normalized income is studied. It is possible to observe how in both cases the evolution is similar. Education variable is the most important since it explains 20% of the income variance; it is shown as over the years studied, the variance increases reaching a 20% in the case of raw income and 23% for scaled income. Referring to age and family variable, it is observed that both are less important, since they represent a smaller percentage of the total variance. They have also had a similar evolution in both cases. The age variable represents about a 2.5% of variance in raw income and 3.5% of the normalized income. The family variable is which represents the least proportion of the total variance, being almost 0.

## VI. Results of paper Spain is different

This section explains the results obtained by Piojan -Mas and Marcos Sanchez (2009)[12]. In their work they conducted the study of the evolution over time of inequality in hourly wages, in working hours, in labor income of households, in disposable income and consumption of households in Spain. We will focus on income, since in this research is studied the evolution of income in Spanish households.

Piojan and Sanchez (2009) used two different types of surveys and they studied the period from 1985 to 2000. The first survey used was the Continuous Household Budget Survey (HBS), this is the old expenditure survey of Spain and collects data during the period 1985 to 1996. The survey collects the expenditure of 215 different types of goods, it is a panel that collects 3600 households for 8 consecutive quarters and provides after-tax income. The use of (HBS) has allowed them to see the evolution of inequality in labor income households, disposable income of households and household consumption. The second data set (data) used is the European Community Household Panel (ECPH). This is a panel of European households which covers the period 1993-2000, and provides income before and after tax. These data have allowed them to see the inequality of labor income, disposable income, hours and each hour of wages both as at household and individual levels.

In this paper as it has been mentioned previously, it is studied the evolution of inequality in labor income of households. They have used the raw data and have scaled the income according to the OECD scale, which has been explained in the previous section. They used the following regression:

$$9) \ \ln y_{i,t} = D_t + \beta_t^e \, D^e_{\ i,t} + \beta_t^f D^f_{\ i,t} + f_t\left(A_{i,t}\right) + \varepsilon^y_{\ i,t} \ ^{13}$$

They analyzed with different measures the evolution of income inequality. They used the variance measures, Gini coefficient, and percentiles 50th-25th and 90th-50th for both databases during the period 1985 to 2000. In Figure 12 such measures of inequality of labor income are observed. It can be observed that variance and the Gini coefficient have an evolution of inequality very similar attending to labor income of the household head, it has a lower rate of inequality in the period 1986-1989, increasing in

---

[12] Spain is different: Falling the trends of inequality
[13] The regression is as dependent variable labor income, and as explanatory variables have two dummies, education and family, and has the polynomial of the age variable.

1992 and decreasing again in the period 1993 to 1997, but at a lower rate than in 1986-1989 period. Instead, there is a greater inequality in household income, the change in inequality in the variance and the Gini is also similar, having the least inequality in the period 1996 to 1999. In the 50th-25th percentile and during the period 90th-50th, the income of household head also has a similar variance and Gini coefficient evolution. Inequality has been declining for the period 1985-1991, with small increases over that period. However, there is a higher rate of inequality from 1992 until the following year, but the increase has been greater for 90th-50th percentile. From that year, inequality has gone down and up year after year to have less inequality in the last two years, this occurs in 50th-25th percentile. And for the 90th-50th percentile, the evolution of inequality decreases over the remaining years. In the case of households, it shows that in the case of the 50th-25th percentile there is a decrease in inequality throughout the period. Instead, in the 90th-50th percentile there is an evolution of the different inequality, which increases during period 1994 to 1996 becoming a lower rate in the last two years studied.

**Figure 12:** *Inequality in equivalized household labour earnings*
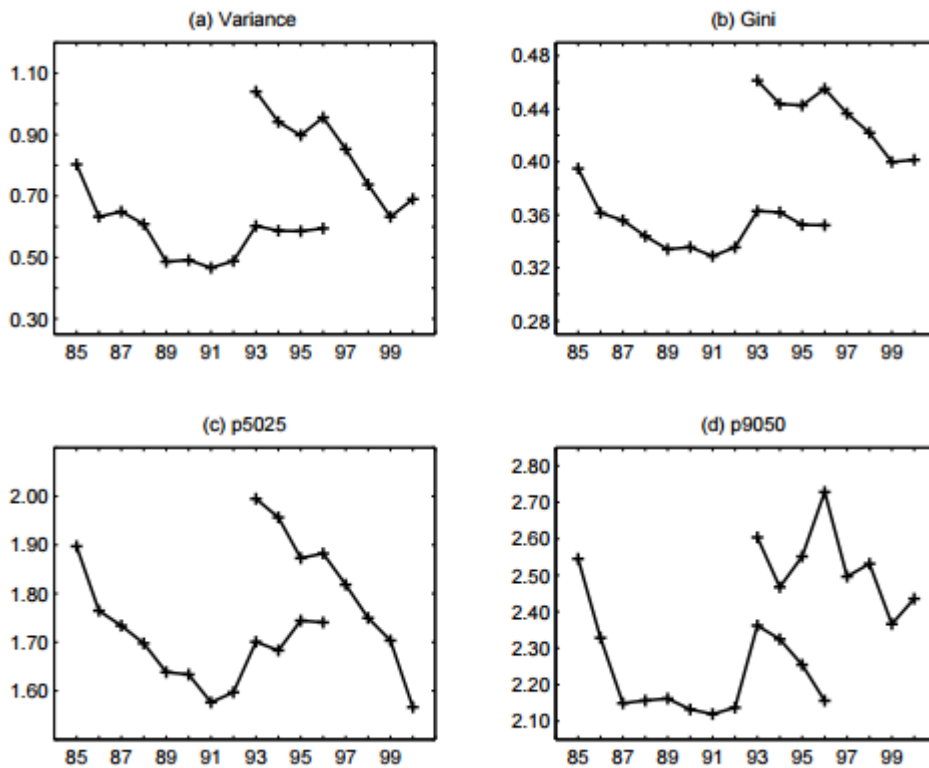


Source: Piojan y Sanchez (2009): Spain is different: falling trends of inequality

Figure 13 shows the evolution of raw income variance and the variance of the logarithm of the scaled income data sets. The evolution is very similar for both cases, being slightly lower the level of inequality for raw income. It is also noted in Figure 17 the variance of each explanatory variable for equation 9. Both as the case of raw income and for scaled income from shows that the education variable is the most important, representing nearly 16 percent of the variance total. In the case of family and age variables are less important, represents a less value of the total variance, representing about 5 percent and 2 percent respectively. It can be conclude that income inequality of households had a minor fall. Inequality of labor income of households is greater than the individual level.

**Figure 13: *Earnings inequality and its decomposition***
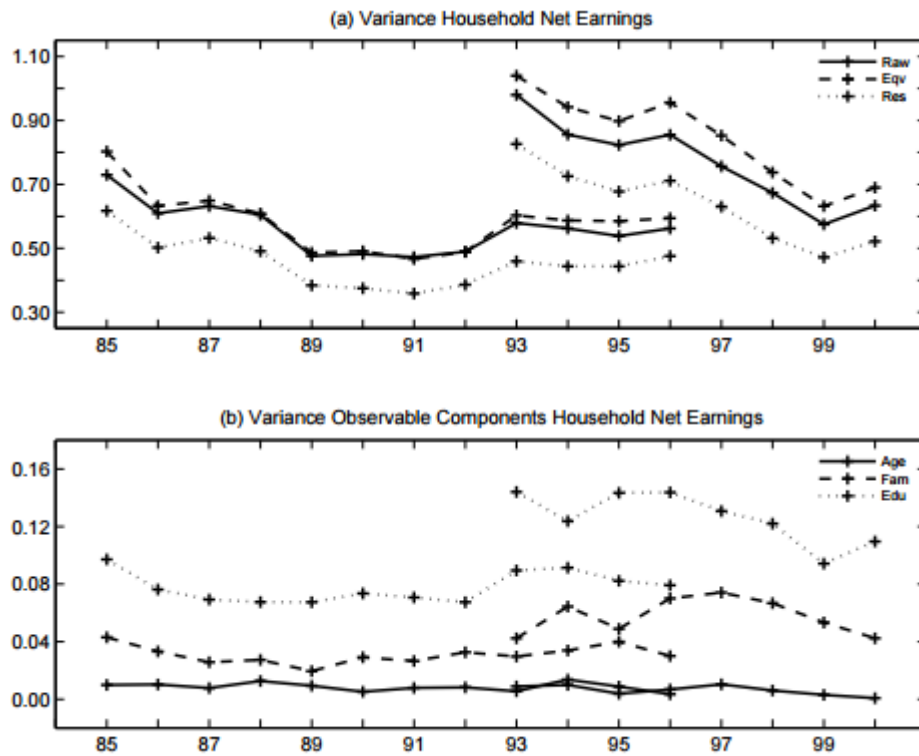


Source: Piojan y Sanchez (2009): Spain is different: falling trends of inequality

## VII. Conclusions

Taking into account the objectives given at the beginning, with this research it has been tried to show how income inequality in Spanish has evolved. The first step has been to analyse the evolution of income distribution from 2002 to 2011. The clearest result is that the normalized income median of income has evolved unevenly during this period. Which decreases in 2005, then, it has a small increase in 2008, reaching higher levels in 2002.

In order to know how the income inequality has evolved from 2002 to 2011, it has been used different measures of inequality. It has been estimated the variance for both as the logarithm of raw income and the logarithm of scaled income. Both follow the same pattern of distribution, but the variance of raw income is larger than the variance of normalized income. The result obtained is that there is an increase in inequality in the first period (2002-2005), then, this value decreases in the following period, but from 2008, with the existence of the financial crisis, the level of inequality increased again.

Another tool for measuring inequality that it has been used is the Gini coefficient, which follows the same pattern as the variance. In this case also it has calculated the index for raw income and normalized income. The coefficients are higher for raw income. The highest inequality arises in 2005 and 2011 for both cases. Therefore, the result coincides with the variance, inequality increases from the year 2008. This evolution of the Gini coefficient can be explained because in the period of the crisis there were major economic changes in Spanish families.

With respect to the ratio 50 / 25th and 90 / 50th, inequality follows the same evolution for both types of income. The 50/25 ratio is greater than the proportion of 90/50 for years, which means that levels of inequality are higher among the 90% of the population with higher income and a 50% of the population with lower income. In the period 2005-2008 the 90/50 ratio decreases, so there is less difference between the income level that has a 90% of the population with respect to the income that own a 50% of the population. However, the ratio increases from the financial crisis, there is greater inequality.

For last, it has been analysed the importance that the studied variables have on the income of Spanish households. In order to study that, it has been calculated the variance of all of them. The result obtained is that education variable has the highest

variance with respect to the others variables. In other words, education variable is which explains a highest percentage of income variance. Referring to age and family variables are less important since they explain a smaller percentage of the variance of income. It is the variable family the least important, becoming virtually 0%.

## References

Atkinson, A. B. (1970). *On the measurement of inequality. Journal of Economic Theory.* [Online].  Available at: *<https://core.ac.uk/download/files/153/7119470.pdf>*

Bank of Spain, 2007. *Encuesta Financiera de las Familias (EFF) 2005: métodos, resultados y cambios entre 2002 y 2005.* Boletín Económico, diciembre, Banco de España. [Online]. Available at: <http://www.bde.es/f/webbde/SES/estadis/eff/Se-parata_EFF_2007.pdf>

Bank of Spain, 2008. *Encuesta Financiera de las Familias (Survey of Household Finances)* *2008.* [Online]. Available at: <www.bde.es/bde/en/areas/estadis/Otras_esta- distic/En-cuesta_Financi/>

Bonhome,. S. and Hospido,. L.(2012). *Earnings Inequality in Spain: Evidence from Social Security Data.* [Online]. Available at: <https://www.cemfi.es/ftp/pdf/papers/pew/InequalitySpainMar2012.pdf>

Bover, O., 2004. *The Spanish survey of household finances (EFF): description and methods of the 2002 wave.* Documentos Ocasionales, nº 0409, Banco de España. [Online]. Available at: <http://www.bde.es/f/webbde/SES/Secciones/Publica-ciones/PublicacionesSeriadas/DocumentosOcasionales/04/Fic/do0409.pdf>

Bover, O., 2008. *The Spanish survey of household finances (EFF): description and methods of the 2005 wave.* Documentos Ocasionales, nº 0803, Banco de España. [Online].   Available at: <http://www.bde.es/f/webbde/SES/Secciones/Publica-ciones/Publicaciones-Seriadas/DocumentosOcasionales/08/Fic/do0803e.pdf>

Bover, O., 2008. *Wealth inequality and household structure: US vs. Spain.* Documentos Ocasionales, nº 0804, Banco de España. [Online]. Available   at: <http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/Publica-cionesSeriadas/DocumentosTrabajo/08/Fic/dt0804e.pdf>

Bover, O., 2008. *Dinámica de la renta y la riqueza de las familias españolas: resultados del panel de la Encuesta Financiera de las Familias (EFF) 2002-2005.* Documentos Ocasionales, nº 0810, Banco de España. [Online]. Available at:http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas /Do cumentosOcasionales/08/Fic/do0810.pdf

Bover, O., 2011. *The Spanish survey of household finances (EFF): description and methods of the 2008 wave.* Documentos Ocasionales, nº 1103, Banco de España. [Online]. Available at: <http://www.bde.es/f/webbde/SES/Secciones/Publica-ciones/Publicaciones-Seriadas/DocumentosOcasionales/11/Fich/do1103e.pdf>

Budría, S., Díaz-Gimenez, J.,(2006). *Economicinequality in Spain:*
*The european community household panel dataset,* [Online] Available at <https://www.etla.fi/edwin/publ/papers/spdf141.pdf>

Cowell, F. A. (2000). *Measurement of inequality. In A. B. Atkinson and F. Bourguignon (Eds.), Handbook of Income Distribution*, [Online] Available at <http://darp.lse.ac.uk/papersdb/Cowell_(DARP36).pdfZ>

Cowell, F. A. (2007). *Income Distribution and Inequality*, *Handbook of Income Distribution*, [Online] Available at <https://core.ac.uk/download/files/153/7119470.pdf>

D. H., L. F. Katz, and M. S. Kearney (2005), "*Residual Wage Inequality: The Role of Composition and Prices",* NBER working paper 11628.

Heathcote, J., Perri, F., and Vilanten G*.,(2006). An Empirical Analysis of Economic Inequality in the United States,* [Online] Available at <https://www.mpls.frb.org/research/SR/SR436.pdf>

Pijoan, J., Sánchez, V., 2009. *Spain is different: Falling trends of inequality*. CEMFI Working paper 0910. [Online] Available at <http://www.econ.umn.edu/~fperri/papers/spain2.pdf>