Assessing the correlation between labour income risk and household portfolio investment in risky assets: evidence from Italian longitudinal data.

Stefano Alderighi
Department of Economics, University of Essex
saldera@essex.ac.uk

October 2014

Abstract

The aim of the present paper is to assess the correlation between labour income volatility and household investment in risky assets, using a representative sample of Italian households (SHIW, Bank of Italy). The paper firstly analyses the covariance structure of earnings in Italy, at a household and at an individual level. The evidence found is used to construct a measure for labour income risk. The variable so created is then introduced as a regressor in a reduced form estimation modelling the determinants of household portfolio allocation in risky assets. The regression analyses shows that labour income volatility is negatively correlated with household investment in risky assets. This finding is consistent with theoretical predictions, and is robust to different specifications and different econometric techniques. The regression results also show that disentangling labour income risk for household heads ("husband and wife") does not add explicative power to the model, whilst a joint measure of labour income risk is negatively correlated with the choice of investing in risky assets. When incomes of household heads are positively correlated, and labour income risk is consequently perceived as overall higher, household portfolio allocation is switched towards safer investment decisions.

1 Introduction and literature review

The present paper investigates the correlation between labour income risk and household portfolio decisions. More in detail, this study shows how labour income volatility is negatively correlated with household investment in risky assets.

The standard Arrow-Debreu portfolio choice model predicts that the optimal share of wealth (w_i) invested in risky assets is approximately equal to:

$$\mathrm{w}_i \approx R\frac{\tilde{R}}{\tilde{\sigma^2}}\frac{1}{\theta i} \quad (1)$$

where w_i is the share of risky assets, R are the returns on the risk - free asset, $\tilde{R}$ and $\tilde{\sigma^2}$ are respectively the expected return premium and the expected volatility of the risky asset, and θ is the individual relative risk aversion coefficient (Gollier, 2002)[1]. The baseline model result predicts that the choice of investing in risky assets depends on a limited set of variables, and does not admit a corner solution. As Cooper & Zhu (2013) indicate, "standard portfolio models predict that every household should participate in the stock market, and that the share of stocks over financial wealth should be high". Calibrations of the standard model show indeed that, given the characteristics of risk-free and risky assets, risk averse households should on average invest a relevant fraction of their wealth on the latter ones (Merton, 1971; Heaton & Lucas, 1997). However, empirical evidence does not support the theoretical conclusions. Firstly, household do not invest as much as predicted in risky assets. Secondly, most of the households actually do not invest at all in them, which is at odds with the predicted absence of corner solutions. This result is known as the stockholding puzzle (Haliassos & Bertaut, 1995).[2]

Reconciling the theory with the empirical evidence presents henceforth a twofold problem:

1. the above equation does not include several indicators that have been found important in explaining the choice of investing in risky assets. The baseline model needs a richer specification;

2. households on average choose not to invest in risky assets. This issue (the corner solution problem) has been extensively investigated both from a theoretical and an empirical perspective. The debate of the reasons why households do not invest in risky assets is still open to question.

Starting with the first point, equation (1) predicts that the optimal choice of risky asset share depends **only** on asset market characteristics and on the individual measure for risk aversion. As anticipated, the empirical evidence is not in line with this conclusion. Haliassos & Michaelidis (2002) for example show that optimal the share of risky assets increases with income, wealth and age. This problem has been more recently addressed by Chang, Hong & Karabarbounis (2012). Calvet & Sodini (2012) use a sample of Swedish twins to show that financial wealth has "a strong positive impact on the share of liquid portfolio invested in risky assets". Cooper & Zhu (2013) study the dependence between household portfolio choices and the educational level of household members, finding a significant correlation. A positive and significant correlation between education and investment in risky assets has been found before by Cambpell (2006) and Vissing-Jorgensen (2002).

Moving to the second point, the literature proposes several explanations to address the stockholding puzzle. Perraudin & Sorensen (2000) argue that the presence of monitoring costs "either of time and money, of holding various stocks" prevents households to hold all possible kinds of assets. Arrondel & Calvo-Pardo (2009) emphasize the role of borrowing constraints and labour income uncertainty as a possible cause for the corner solution problem. Chang, Hong & Karabarbounis (2012) stress the role of labour income risk in explaining the stock-holding puzzle. Sanromar (2013) focuses her attention on the existence of participation costs in the asset market. Knupfer, Rantapuska & Sarvimaki (2013) use an exogenous unpredictable event to show that labour income shocks have a strong impact on household portfolio allocation in risky assets.

[1] "Household Portfolio", edited by Guiso, Haliassos & Japelli, MIT Press, 2002, page 35

[2] Miniaci & Weber (2002), on the stockholding puzzle: " How do households decide whether to invest in risky financial assets? Here the key open question is why so many households do not have direct holdings of risky assets (stocks, equity funds, and long term bonds). This is known as the stockholding puzzle (Haliassos & Bertaut, 1995) and is the micro analogue of the equity premium puzzle" Miniaci & Weber in "Household Portfolio", edited by Guiso, Haliassos & Japelli, MIT Press, 2002, page 145.

This paper empirically assesses household allocation choices, explaining the model specification problem and the corner solution criticism through labour income risk. When labour income is more volatile, households tend to invest in safer assets to diversify their risk. As Chang, Hong & Karabarbounis (2012) point out, when income is "skewed to labour earnings, (household tend to) entice investments towards safe assets to protect against risky human capital". The reduced form estimation proposed relies on this simple assumption.

As the aim of this research is to concentrate on labour income risk, this paper firstly focuses its attention on labour income dynamics and its variance decomposition. The purpose of this analysis is the creation of a labour income risk score that can be used as a regressor in a reduced form estimation. The first step for the generation of this variable is the analysis of the covariance structure of earnings in Italy, and the consequent identification of permanent and transitory components of labour income. The literature on the topic dates back to the seventies. Lillard & Weiss (1979), MaCurdy (1982), Abowd and Card (1989) are seminal contributions on the topic. In the cited studies the authors pool together micro data and analyse the covariance structure of earnings assuming homogeneity of the income process across individuals. Baker (1997) deals with the problem of individual heterogeneity, comparing the "profile heterogeneity model", in which "the earnings profile varies across individuals" with a unit root model that does not allow any individual heterogeneity. The author concludes that the former "provides a more consistent representation of the data". Alvarez, Browning & Ejrnaes (2010) decompose labour income allowing complete individual heterogeneity, that is by letting each individual in the sample have their own income process[3]. The approach herein pursued is hybrid, mainly due to data unavailability. Individual heterogeneity is allowed, but among the boundaries of a generic specification derived from a standard analysis of the covariance structure of earnings.

The analysis of the covariance structure of earnings is finalized to the creation of a labour income volatility measure, to be later introduced as a regressor in a reduced form estimation. This variable is calculated both at a household and at an individual level. As already stated, the aim of the contribution is to assess how labour income risk affects household portfolio allocation. It is henceforth logical to focus on households the income of which is skewed towards labour earnings. Econometric estimations will consequently be performed on a sub-sample of households where the second and the main earner of the household ("husband and wife") are both employees. Two reduced-form models will be estimated: one in which labour income is calculated at a household level, another one in which labour income is disentangled for the main and the second earner[4]. The results found show that what matters for household portfolio allocation is the joint labour income risk of the main and the second earners in the household.

The paper is structured as follows: section 2 describes the database chosen. Section 3 describes the covariance structure of earnings at a household level and at an individual level. Section 4 describes the construction of the labour income risk score. Section 5 describes the estimation process. Section 6 concludes.

---

[3]Recent work on the covariance structure of earnings by Altonji, Smith & Vidango (2013) has remarkably contributed to the literature on the topic by developing a multivariate analysis of individual earnings. The cited seminal contribution by Abowd & Card (1989) already moved towards this direction. The authors indeed analyse how earnings co-vary with working hours. However, and as the Altonji et alii underline, "with only one indicator (that is, earnings themselves) even richly specified models cannot specify the various sources of earnings fluctuations". The authors develop a complex structural model describing the behaviour of wages and earnings in the lights of labour market transitions, experience accumulation, worked hours changes. The model combines the literature on earnings decomposition with the search-and-matching one. Albeit this work represents the research frontier on the topic, at least to my knowledge, as this paper focuses on the relationship between labour market volatility and portfolio allocation, such a complicated analysis of earnings is out of the purposes of this paper. See also Guvenen & Smith (2013) for an analysis of the volatility of earnings in the light of consumption decisions

[4]To the knowledge of the author, there are no empirical works assessing household portfolio allocation using a set of regressors that disentangles for "husband and wife" characteristics

2 Database Description

The aim of this section is to describe the database chosen to investigate the proposed research question.

The data used are taken from the Survey of Household Income and Wealth of the Bank of Italy (SHIW from now on). "The SHIW survey is based on a random sample of approximately 8,000 households per year, and is available from 1977 annually and at odd years after 1987. It contains information both on households (family composition) and on individuals" (Brunello, Comi & Lucifora, 2001). However, the database has an actual panel structure only from 1989 onwards. For this reason, the analysis proposed pertain to the 1989 - 2010 time span, using eleven waves in total.

The dataset contains information about individual characteristics, individual employment status, individual income and wealth, household income and wealth, house tenancy. In particular, the information about household wealth is structured in detail. This makes the dataset particularly suitable for the research question investigated in this paper. The SHIW dataset has indeed already been used in the literature to study household portfolio choices, albeit with different techniques (Sanroman, 2014; Pelizzon and Weber, 2009; Guiso and Japelli, 2002; Guiso *et alii*,1996). The archive FAMIxx (where xx represents the corresponding wave) is dedicated to financial wealth. It contains several dummies describing whether the individual holds different sorts of assets. Government bonds are for example divided by type (zero coupon, short, mid and long run). Current accounts are divided into bank and postal accounts. Shares are divided by type of company (small business, limited responsibility). The archive contains also information about the amount of money individuals have invested in each of the assets taken into account. Information about liabilities is detailed as well, and spans from mortgage status to informal debts towards family and friends. The archive IMMPxx contains instead information on real wealth and housing. The variables enclosed describe several house characteristics, as house market value, house owning status, physical features (surface, number of toilets, neighbourhood).

Variables about labour market characteristics and incomes are spread across all the dataset. COMPxx archives contain information about the sector of the main activity, non-occupation outcomes (unemployed, looking for a position, student, retired) and occupation status (blue collar, white collar, manager). The archive LDIPxx contains instead a limited number of variables referring to the working activity (number of worked hours, size of the firm the employee is working in, amount of fringe benefits, hours of overtime work). Unfortunately, information about individual education and human capital is relatively limited. The only variable describing educational choices is the maximum educational level attained by the individual. Moreover, the survey does not contain any proxy for individual skills, nor variables related to labour market experience or on-the-job training.

2.1 The SHIW structure

This subsection describes the general structure of the SHIW database.

The database is a household survey. Hence, each individual is observed in a specific household. Some variables are collected at a household level, some others at an individual one. Each household corresponds to a questionnaire number, labelled NQUEST (literally, "number of questionnaire"). Each individual in the household is then identified by an order number, labelled NORD (literally, "order number"). The order of the individual depends on her level of earnings. In the survey, the main earner in the family is called the **reference person** (**RP** from now on). Her order number is always 1. An individual with order number equal to two is supposed to be the second earner in the household. So it happens for all the other numbers, depending of course on household size.

The variable PAR describes instead the role of each individual in the household. For the reference person (NORD=1), this variable is always equal to one as well. Hence, for the reference person, PAR and NORD coincide by definition. PAR is equal to two if the person is married to the reference person. PAR is equal to 3 if the individual is the son or the daughter of the reference person. PAR is equal to 4 if the individual has some other role (i.e. grandparent, uncle). Albeit tightly correlated, PAR and NORD obviously don't perfectly match. It may happen indeed that the second earner is not the household member married to the reference person, but, say, RP's daughter. If an individual is married to the reference person **and** is also the second main earner in the household, in this paper she is labelled "reference person two" (**RP2** from now on). The analysis is then restricted to households where both RP and RP2 are in employees. Appendix one describes the sub-sample selection more in detail.

Income is observed at a household level (key variables NQUEST and ANNO) and at an individual one (key variables NQUEST, NORD, ANNO). The following identity holds for both households and individuals:

$$Y = YL + YT + YM + YCA + YCF$$

where Y is total income, YL is labour income, YT is income from other transfers (retirement), YM is self employment income, YCA is real assets income, YCF is financial income. All variables can be disaggregated more.

A similar structure applies to **household** wealth (key variables NQUEST, ANNO), as shown in the following:

$$W = AR + AF1 + AF2 + AF3 - PF$$

where W is total net wealth, AR is real assets, $AF1$ is current accounts, $AF2$ is government bonds, $AF3$ is other securities, PF is financial liabilities. All variables can be disaggregated more. Gross Wealth (GW) is finally defined in this paper as real wealth plus financial wealth. Using the SHIW notation,

$$GW = AR + AF1 + AF2 + AF3$$

All analyses of this paper are performed on a subsample of households where

- the main earner (**reference person**, RP) is employed;
- the main earner is married;
- the person married with the reference person is the second earner in the household and is employed as well (**reference person two**, RP2).

Appendix 1 contains all the details of the sub-sample selection. Section 3 instead concentrates on the covariance structure of labour income.

3 The covariance structure of earnings

The present section analyses the covariance structure of earnings in Italy. The analysis is performed both at a household level and at an individual level for the reference person and for reference person two. The present section displays the results found for the reference person. Section 3.1 discusses summary statistics in the SHIW database, mainly regarding labour income. Section 3.2 describes the procedure used to determine the covariance structure of earnings. Section 3.3 studies the covariance structure of earnings for the reference person. Section 3.4 draws conclusions about the generating process.

3.1 Descriptive statistics

This subsection provides descriptive statistics regarding real labour income and other characteristics in the sub-sample chosen. The aim of this analysis is to preliminary study the covariance structure of earnings.

The following table displays the mean and the standard deviation of real labour income for all individuals earning a positive amount, for the reference person, and for reference person two[5]. By definition, the means show a physiological gap, as the reference person is the main household earner.

Table 1: Summary statistics / real labour income

	Mean	Std dev
Real Labour Income	16795.58	8821.38
Real Labour Income (RP)	19272.42	9730.06
Real Labour Income (RP2)	15786.17	7410.56

The following graphs plots the year means of the three variables displayed above along the panel time span.

[5] RP2 being defined, as stated in subsection 2.2 as a person married with the reference person, being the second earner in the household, and earning a positive labour income

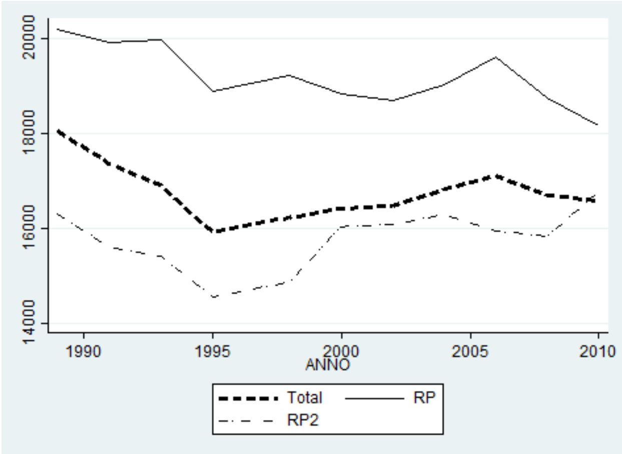

Figure 1: Real earnings along time

Real earnings show a stable pattern along time, without sharp peaks of throats. As Sanroman (2014) underlines commenting on SHIW data, "(real) labour income remained almost invariant over the last two decades".[6] The following graph displays how labour income volatility evolves along time for the three categories studied so far. The statistics shown are standard deviations from year means.

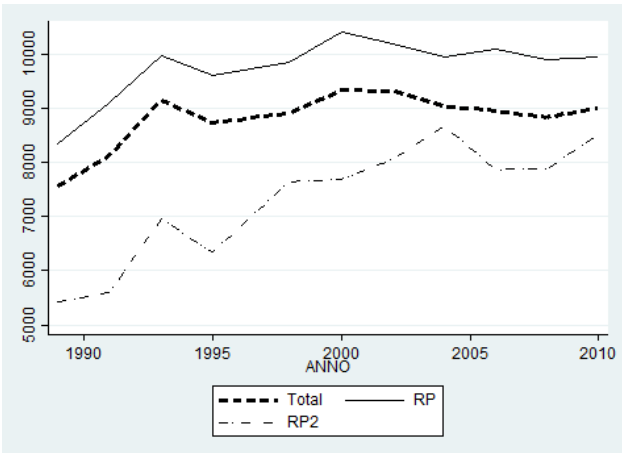

Figure 2: Real earnings volatility along time

[6]The difference between RP and RP2 is deep. It should be considered that the reference person is by definition the main earner and, as already stated, the gap between the two values is physiological. However this difference can be also attributed to gender inequality in the Italian labour market (Mussida and Picchio, 2011). Indeed 81.33 % of RP2 are women. Graph 1 in Appendix 1 plots real earnings divided by gender (regardless of being a reference person or not). The evidence suggests that gender actually represents one of the driving forces for this wage gap. Males earn averagely more, and are as a consequence more likely to be reference persons in the SHIW sample.

Labour income is particularly stable for household heads, whilst it shows an increasing volatility for reference person two.

The following table displays summary statistics referred to household level income:

Table 2: Summary statistics for household real labour income

	Mean	Std dev
Real Labour Income (RP working)	28169.04	15280.12
Real Labour Income (RP and RP2 working)	37916.35	14865.44

The following graphs display instead how labour income mean and volatility evolve long time. Household labour income volatility increases over the time span chosen:

Figure 3: Household real earnings along time

Figure 4: Household real earnings volatility along time

This preliminary analysis shows that, for the three categories chosen, real earnings are characterized by a stable mean. Household level income and RP2 real earnings show a growing variance, whilst for the reference person labour income volatility is overall stable. From the graphs above it is difficult to draw conclusions about the persistency of the process. The following subsections will deepen the analysis performed so far.

3.2 Determining the covariance structure of earnings

The procedure used to determine the covariance structure of earnings is well established in the literature on the topic. The first step of this analysis involves the estimation of a first stage regression of log - earnings on a set of covariates. What regressors to introduce into the first stage is still open to question. Many scholars (see for example Baker, 1997; Abowd & Card, 1989) regress the logarithm of labour income on potential experience, usually defined as individual age minus years of education minus 5. Year effects are generally introduced as well. Baker (1997) for example regresses the logarithm of labour income on potential experience, using a system of simultaneous equations in which each equation corresponds to an year. Meghir & Pistaferri (2004) and Borella (2001) regress log of labour income on a more involved set of covariates. Browning et alii (2012) regress log of labour income on "on age and experience variables and time dummies".

The specification proposed in the paper is partially constrained by data availability. The dataset used contains indeed the educational level of the individual, but does not contain the actual years of education. In Italy it is possible to repeat high school years if the individual is not prepared enough for the following year. Moreover, the university system potentially allows students to take as much time as they need to obtain their degree. Finally, different disciplines took different periods to be completed, at least before the 2000 reform.[7] As a consequence, measuring the actual years of education of an individual without knowing the year in which she obtained the degree becomes pretty much impossible. This in turn prevents any estimation of potential experience. The specification chosen simply involves the introduction of age dummies and year dummies. The reason for this choice is simple. The first stage regression residuals will represent a proxy for individual earnings. It is henceforth worth controlling for exogenous variation in the level of earnings (business cycle, age profile) leaving the residual as individual specific as it could be. The target of this analysis is indeed the creation of an individual labour income risk score. Controlling for personal characteristics (education or family background, and, in the opinion of the author, even experience) would lead to misleading conclusions when analyzing the individual specific generating process. The choice is henceforth to leave the first stage regression as simple as possible[8]. Formally:

$$lny_{it} = \hat{\alpha} + \sum_{i=1}^{11} \hat{\beta}_i Dyear_i + \sum_{j=1}^{40} \hat{\gamma}_j Dage_j + u_{it} \tag{2}$$

After running the first stage regression, the aim of the analysis is to study the covariance structure of the regression residuals. This involves the creation of a variance-covariance matrix, estimated after pooling together all individuals across waves. The underlying hypothesis is that all individuals are characterized by the same, homogeneous, income generating process. Calling $\bar{u_i}$ the between-group mean of the residuals, the deviations from the mean are defined by $\tilde{u_i} = u_{it} - \bar{u_i}$. The estimated variance-covariance matrix is consequently: $CL = \frac{1}{n_t} \sum_{i=1}^{n_t} \tilde{u_i}\tilde{u_i}'$, where CL stands for "covariance of levels". CD (covariance of deviations) is the name of the variance covariance matrix for Δu_{it}, estimated using the same formula.[9] The variance covariance matrices are informative about the nature of the income generating process.[10]

The following section describes the procedure in detail for household heads. As the analysis follows exactly the same steps for labour income at a household level and for reference person two, results pertaining to these two categories are displayed in appendix 3.

[7] In the year 2000 a massive reform radically changed the Italian university system. The most noticeable change pertained the duration of University degrees. Before the year 2000, indeed, the Italian university system was characterized by a one-tier path, without distinction between undergraduate and postgraduate studies. Each faculty was characterized by a different number of years for complexion. Economics and psychology, for example, took 4 years to be completed. Law and engineering, 5 years. Medicine, 6 years. At a certain moment, 3 years degree were introduced and overlapped with the one-tier path. After the reform, the three years plus two system has been introduced smoothly, as those who were enrolled with the old rules were not forced to adopt the new system. Henceforth, and as an example the degree of an individual that took 10 years to complete a bachelor degree in physics cannot be compared with the degree of an individual taking an economics one-tier degree on time.

[8] To run the first stage regression t is necessary to understand what is the working life (in terms of age) of both reference person one and reference person two. Graphs 2 and 3 in appendix 1 plot the age distributions of both. The age span chosen for household heads is 25 to 65. The age span chosen for RP2 is instead 20 to 60. For the household level first-stage regression age dummies refer to the average age between reference person one and two.

[9] The notation follows Abowd & Card's statistical appendix (1994), to which I refer for further details.

[10] Indeed, and as an example, if the covariance of the process in levels shows persistence across time, this would represent the hint for a non - stationary generating process. However, a covariance matrix displaying high persistence could also be due to an error term characterized by an MA(q) representation, which if inverted would have an AR(∞) representation.

3.3 The covariance structure of earnings for the reference person

First stage regressions are performed on the male heads of the sub-sample described in appendix 2.

3.3.1 The process in levels

The following table displays the variance covariance matrix for the level of earnings (CL) of the reference person.

Table 3: Correlations - RP labour income

	2010	2008	2006	2004	2002	2000	1998	1995	1993	1991
2010	**0.198****	0.785*	0.680*	0.749*	0.676*	0.621*	0.359*	0.422*	0.035*	0.018*
2008	0.150*	**0.185****	0.731*	0.707*	0.668*	0.687*	0.412*	0.384*	0.422*	0.031*
2006	0.156*	0.162*	**0.266****	0.697*	0.660*	0.584*	0.401*	0.336*	0.393*	0.216*
2004	0.133*	0.121*	0.143*	**0.159****	0.743*	0.655*	0.455*	0.427*	0.418*	0.364*
2002	0.121*	0.115*	0.137*	0.119*	**0.161****	0.680*	0.541*	0.458*	0.436*	0.375*
2000	0.108*	0.116*	0.118*	0.102*	0.107*	**0.153****	0.700*	0.535*	0.527*	0.368*
1998	0.059*	0.065*	0.076*	0.067*	0.080*	0.101*	**0.136****	0.581*	0.499*	0.461*
1995	0.057*	0.050*	0.053*	0.052*	0.056*	0.064*	0.065*	**0.092****	0.584*	0.395*
1993	0.040*	0.070*	0.078*	0.064*	0.068*	0.079*	0.071*	0.068*	**0.148****	0.486*
1991	0.003*	0.038*	0.038*	0.050*	0.052*	0.049*	0.058*	0.041*	0.064*	**0.118****

Autocovariances below diagonal, autocorrelations above diagonal, cross sectional volatility on diagonal. * : significant at the 5% level.

All the correlations displayed are significant at the 5% level. The generating process is consequently characterized by a high persistence. This finding can be attributable to different reasons. The process can be characterized by non stationarity. Alternatively, the generating process can be stationary, and simply show a high persistence that does not die out in the time span considered. Moreover, if the process is stationary and invertible, and if the error term follows an MA(q) process, the Wold representation theorem predicts that the autoregressive part can be expressed as an AR(∞) model. This would lead to a highly persistent correlogram as well. More detailed analysis of the process is consequently needed to draw accurate conclusions about the nature of the generating process. In particular, the very same correlation analysis should be performed on the process in first differences to understand what the nature of the error term is (Baker, 1997; Meghir & Pistaferri, 2004).

The following table shows how cross sectional volatility evolves along time. The table displays how the sample size evolves, as the longitudinal data used are not balanced. Indeed, as the sample contains individuals observed more than five times, at the beginning of the time span the number of individuals has to be lower than in the middle of the period considered. Moreover, displaying how volatility evolves is informative about the stationary properties of the process. In a covariance stationary process, indeed, the variance should be constant along time. However, if the variance increases or decreases with time, this represent a hint for the presence of non stationarity.

Table 4: Cross sectional volatility across time - RP

Year	Variance	Frequency
1991	0.118	138
1993	0.148	170
1995	0.092	177
1998	0.136	228
2000	0.153	255
2002	0.161	256
2004	0.159	224
2006	0.266	210
2008	0.185	183
2010	0.198	140

The following graph plots cross sectional variation against time:

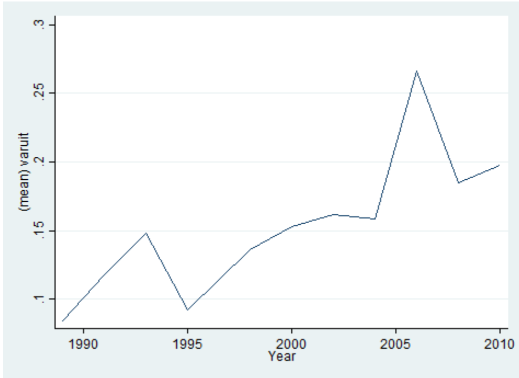

Figure 5: Cross sectional variation against time

A covariance stationary process is characterized by a constant variance along time. The graph shows instead a growing variance, or, in other words, a growing inequality. However, despite at an aggregate level the process seems to show non stationary properties, it is still early to conclude that individual generating processes are non stationary. This pattern can be indeed due to individual heterogeneity.

In a covariance stationary process, the covariance merely depends on the lag order considered, and not on time. Hence the first order autocorrelation of a stationary process should oscillate around a certain mean if plotted against time. The following graph shows the described plot. The observation for 2010 represents the correlation between 2010 and 2008. The observation for 2008 represents the correlation between 2008 and 2006, and so on.

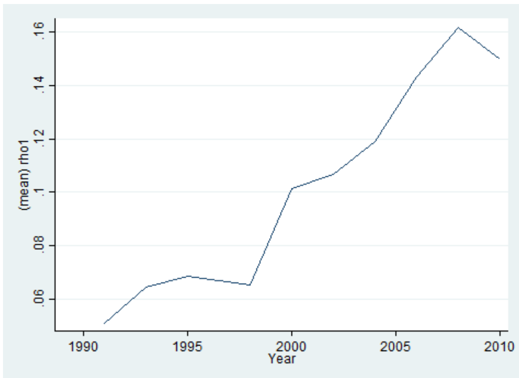

Figure 6: First order autocovariance plotted against time

As for cross sectional dispersion, first order autocorrelation shows a growing pattern. That is, the covariance does not seem to depend on the lag itself, but also on the time period considered. According to this graph, the process has non stationary properties.

The following two graphs plot second and third order autocorrelations against time. The conclusions drawn accordingly are similar to the ones drawn for the variance and the first order autocorrelation.

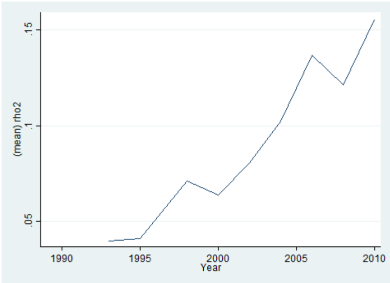

Figure 7: Second order autocovariances

Figure 8: Third order autocovariances

In line with the time series plots of the variance and the autocovariances, the process shows not stationary properties. The variance depends on time, and the autocovariances as well. The following table contains OLS regression results where the dependent variables are the second moments, and year represents a regressor. In the following, and using the notation commonly used in time series analysis, the autocorrelation for the j^{th} lag is indicated as γ_j.

Table 5: Second moments regressed on time - RP

VARIABLES	(1) γ_0	(2) γ_1	(3) γ_2	(4) γ_3	(5) γ_4	γ_5
Year	0.00574***	0.00600***	0.00685***	0.00556***	0.00566**	0.00324
	(4.200)	(11.42)	(10.21)	(7.328)	(3.781)	(1.233)
Constant	-11.32***	-11.89***	-13.62***	-11.05***	-11.27**	-6.424
	(-4.150)	(-11.33)	(-10.15)	(-7.274)	(-3.754)	(-1.221)
Observations	11	10	9	8	7	6
R-squared	0.639	0.933	0.925	0.771	0.703	0.347

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The regression results provides evidence that the second moments of the process are time dependent. This finding is consistent with the literature on the topic. Baker (1997) for example points out that "there is evidence of non stationarity as variances rise briefly at the beginning of the panel, and then trend rather randomly before rising sharply in the final years". Abowd & Card (1989) notice that "variances ... vary over time. Cross sectional dispersion in earnings was relatively small in 1972-73 and relatively large in 1975-1976". Variance dependence on time as a hint for non stationarity is an established result in the literature. The analysis displayed so far suggest that the generating process for earnings is non stationary.

3.3.2 The process in differences

The present subsection discusses the analysis of the first difference of the earnings process. This represents a well established procedure in the literature, as studying the correlation of a process in differences is informative about the time series properties of the error term (Baker (1997), Meghir & Pistaferri, 2004). Supposing indeed that the generating process has an AR(1) representation with a unit root and the error term shows an MA(q) representation, as indicated in the following equation:

$$y_t = c + y_{t-1} + \Psi(\varepsilon; q)$$

the correlogram of Δy_t is informative about the time series properties of the error term. Indeed, if the correlogram dies, say, after one lag, the error term has an MA(1) representation. If the correlogram dies abruptly after 3 lags, the error term has an MA(3) representation. Citing Baker, (1997): "assuming (the presence of a unit root) ... all the autocovariance above the (MA) order should be equal to zero; that is , in contrast to (the autocovariances of the process in levels) there is no persistent serial correlation in the earnings growth rates."[11]

[11]The reasoning is valid under stationarity as well, albeit it is more difficult to prove it. Suppose indeed that the process is stationary, and it is characterized by an ARMA(1,1) representation as in the following:

$$y_t = c + \phi y_{t-1} + \Psi(\varepsilon; q), \text{ where } |\phi| < 1$$

it is common knowledge that $E[y_t] = \frac{c}{1-\phi}$. As the process is stationary this is valid $\forall t$. The expected value of Δy_t would be consequently equal to zero, as $c - (1-\phi)E[y_{t-1}]$ would be equal to zero. The first order autocorrelation would then be given by:

$$E[c^2 - (1-\phi)cy_{t-1} - (1-\phi)cy_{t-2} + (1-\phi)^2 y_{t-1}y_{t-2} + \theta\varepsilon^2 + \rho] = (c^2 - c^2 - c^2 + c^2 + c^2 cov(y_{t-1}y_{t-2}) + \sigma^2\theta)$$

The following table displays the variance covariance matrix for the process expressed in first differences (CD).

if the term $c^2 cov(y_{t-1} y_{t-2})$ is small enough, the term $\sigma^2\theta$ should make the first order autocorrelation show a higher magnitude than the higher order ones. However, it is important to notice that in case of a stationary process, the first difference of an ARMA(1,1) process shows a small persistent correlation.

Table 6: Correlations

	2010	2008	2006	2004	2002	2000	1998	1995	1993	1991
2010	**0.079**	-0.239*	-0.114	-0.066	-0.077	-0.098	-0.155	0.099	0.035	1.521
2008	-0.026*	**0.156**	-0.696*	-0.008	-0.061	-0.024	-0.033	0.153	-0.080	-0.011
2006	-0.012	-0.103*	**0.141**	-0.275*	0.117	0.096	-0.035	-0.093	0.155	0.193
2004	-0.005	-0.001	-0.028*	**0.074**	-0.449*	-0.112	-0.034	0.168	-0.140	-0.048
2002	-0.005	-0.006	0.011	-0.030*	**0.062**	-0.380*	-0.050	0.074	-0.012	-0.057
2000	-0.007	-0.002	0.009	-0.008	-0.024*	**0.065**	-0.161*	-0.089	0.175	0.051
1998	-0.011	-0.003	-0.003	-0.002	-0.003	-0.011*	**0.067**	-0.284*	-0.093	-0.033
1995	0.008	0.017	-0.010	0.013	0.005	-0.006	-0.020*	**0.076**	-0.575*	-0.156
1993	0.026	-0.012	0.022	-0.014	-0.001	0.017	-0.009	-0.059*	**0.140**	0.784*
1991	0.094	-0.020	0.016	-0.003	-0.003	0.003	-0.002	-0.009	0.064*	**0.048**

Autocovariances below diagonal, autocorrelations above diagonal, cross sectional volatility on diagonal. * = significant at the 5% level.

The table shows how the correlogram dies after one lag, suggesting that the error term has an MA(1) representation. An MA(1) process in the error term represents a plausible explanation for the high persistency of the process in level, in case of stationarity and invertibility.

The following graphs plot cross sectional variation, first and higher order autocorrelations against time:

Figure 9: Cross sectional variation

Figure 10: First order autocovariance

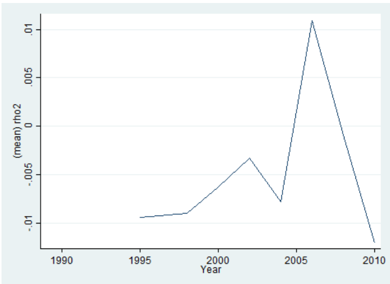

Figure 11: Second order autocovariances

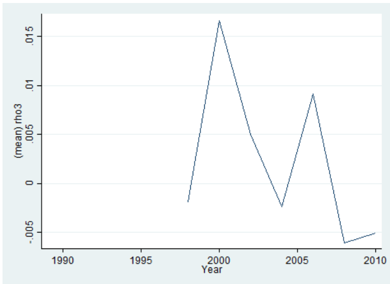

Figure 12: Third order autocovariances

The process expressed in first differences shows stationary properties, despite cross sectional volatility seems to be growing during in the last years of the sample. The following regression shows how the second moments do not actually depend on time.

Table 7: Second moments regressed on time - RP

VARIABLES	(1) γ_0	(2) γ_1	(3) γ_2	(4) γ_3
Year	0.00197	-0.00207	0.000404	-0.000908
	(0.873)	(-0.962)	(0.731)	(-1.117)
Constant	-3.848	4.110	-0.814	1.821
	(-0.853)	(0.955)	(-0.737)	(1.117)
Observations	10	9	8	7
R-squared	0.106	0.153	0.081	0.218

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To conclude, the process in differences shows stationary properties, with an error term characterized by an MA(1) representation.

3.4 Generating process and individual specific estimation

The present subsection draws conclusions about the underlying generating process, based on the evidence found in the previous subsection. It then proposes the estimation of individual specific earning processes.

The behaviour of the second moments of the process in levels represents a clue for non stationarity. However, pooled AR(p) models of the first stage regression residuals provide coefficients that characterize stationary processes. The following table displays the results for the pooled AR(p) estimations.

Table 8: Autoregressive Models - RP

VARIABLES	(1) u_{it}	(2) u_{it}	(3) u_{it}
u_{it-1}	0.714***	0.500***	0.431***
	(16.95)	(7.923)	(6.119)
u_{it-2}		0.313***	0.237***
		(5.171)	(3.335)
u_{it-3}			0.231***
			(5.556)
Constant	0.0150**	0.0177**	0.0216**
	(2.174)	(2.316)	(2.432)
Observations	1,663	1,316	980
R-squared	0.490	0.533	0.562

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As Alvarez *et alii* (2012) underline, the inconsistency of the data findings can be attributed to individual heterogeneity. As the authors point out: "mixing from different populations with different processes is known to lead to more complicated processes if the pooled data is treated as homogeneous". It is indeed logical to think that each individual in the sample is characterized by her own generating process[12]. The literature has investigated this problem using different procedures. Baker (1997) for example allows the presence of individual heterogeneity in the earnings profile by regressing the first stage residuals on actual individual experience (a sort of "second stage regression"). However, this choice is rather arbitrary in the opinion of the author. Individual specific earnings profiles can be indeed attributed to different educational levels, or to different family background characteristics. As already underlined in section 1, the Alvarez *et alii* (2012) allow for complete individual heterogeneity. Despite this choice looks perfectly logical, the parameters distributions the authors choose in their paper appears rather arbitrary. Moreover an earnings process estimation as complicated as the one the authors propose lies out of the purposes of the present contribution. The approach followed will consequently be hybrid. Individual heterogeneity is allowed among the boundaries of the specification suggested by the analysis of the pooled earnings process [13].

[12]In their contribution, Alvarez *et alii* (2010)state that they are "sceptical that everyone has the same process with much the same parameters. Rather it may be that different workers have different processes, some with a unit root, some with a stationary AR(1) model, and others with a MA(1) model, for example"

[13]Alvarez, Browning & Ejrnaes (2010) actually take into account the possibility of estimating individual specific processes in their paper. As the authors state: "one (possible) option is to first conduct an analysis of time series on each person and then to use this to generate a model of unobserved heterogeneity using parametric distributions for the unknown parameters - a 'bottom-up' approach. The problem with following this strategy is that the individual estimates suffer from considerable small sample and endogeneity biases. It might be possible to implement analytic or simulation based small sample corrections to the estimators properties but these corrections

The specification proposed in this paper for the individual income generating process is the following:

$$u_{it} = \alpha_i + \phi_i u_{it-1} + \varepsilon_{it} + \Theta \varepsilon_{it-1} \tag{3}$$

where ϕ_i can be lower, equal or bigger than one, α_i can be equal or different from zero. It is worth noticing that Browning et alii also use an MA(1) specification for the error term. This functional form is consistent with previous studies in the literature (Baker (1997), MaCurdy (1982), Abowd & Card (1989), in which actually the error term follows a MA(2) process, Brwoning et alii (2004))

Both the constant term and the AR(1) coefficient are left individual specific, to allow for individual heterogeneity.

Altough the database used contains relevant information for the purposes of this paper, it is characterized by a short time dimension. Individuals can be observed 11 times at most. Moreover, the number of individuals observed for a time span this long is particularly limited (11 individuals in my sub-sample). This represents a limit for the estimation of the variance of the generating process. Indeed, maximum likelihood estimation of individual ARMA(1,1) models is not feasible. The only possible option for individual specific estimations is OLS estimation of autoregressive processes. The results displayed in the following refer to individual specific OLS estimation of a simple AR(1) model. The following equation describes the individual regressions performed:

$$u_{it_i} = \hat{\rho} u_{it-1_i} + e_{it_i} \tag{4}$$

where the sub-index i underlines the fact that the panel is not balanced.

The following graph plots the distribution of the AR(1) coefficients for the 327 individual estimations performed:

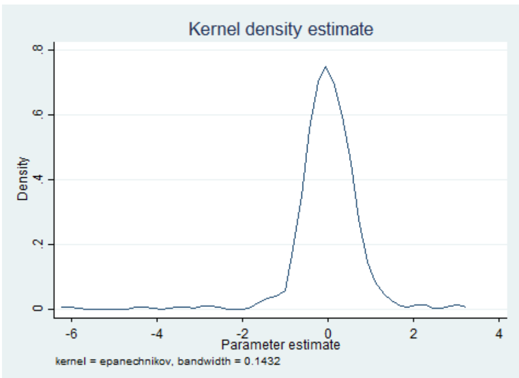

Figure 13: Autoregressive coefficients - 327 individual estimations

impose stronger assumptions on the distributional properties of the errors than those we would like to impose a priori". It is worth noticing that the authors actually underline the degree of arbitrariness of their own approach.

The distributions looks roughly symmetric around a value of zero (despite the long tails). Among all coefficients, only a small portion of them (25) are significantly different from zero at the 10% level. Only 4 individuals show non-stationarity, and 10 of them display a negative AR(1) coefficient. The regression results show massive individual heterogeneity. However, it should be taken into account that for most observations the sample size is particularly small.

4 Labour income risk score

The aim of this paper is to assess whether and how labour income risk affects household portfolio choices. The aim of the analysis proposed in section 3 is the creation of an indicator measuring labour income volatility. This will be later introduced as a regressor in a reduced form estimation studying household investment in risky assets. Recalling from equation (4) that e_{it} represents the residual from the individual specific regression, the measure proposed is the following:

$$s^2 = \frac{1}{t_i - 1} \sum_{t_i} e_{it_i}^2 \tag{5}$$

that is, the unbiased sample variance estimator, calculated for each individual. This is however an imprecise point estimate. A simple way to partially overcome this problem is the following. Knowing that:

$$E[s^2] = \sigma^2 \text{ and } var[s^2] = \tfrac{2\sigma^4}{n-1}$$

it is possible to re-sample with replacement individual observations to obtain an expected value for the sample variance. It is worth noticing that consistency is not exploitable in this case, as it is not possible to observe the individual more than they actually have been. However, this bootstrap-like procedure makes it possible to calculate the expected value of the variance sample estimator. Formally:

$$\frac{1}{k} \sum_{j}^{k} [\frac{1}{n-1} \sum_{i}^{n} (x_1 - \bar{x})^2] = Risk_i \tag{6}$$

provided that k re-samplings with replacement are taken. The following table displays estimations of labour income risk at a household level, given the number of replications:

Table 9: Labour Income Risk Score

Number of resamplings	Expected Value	Variance
Point estimate	0.199	0.165
10 replications	0.167	0.141
20 replications	0.166	0.135
30 replications	0.166	0.134
40 replications	0.165	0.129
100 replications	0.166	0.132
500 replications	0.166	0.132

The table shows that 20 replications are enough to converge to the expected values. The procedure corrects the bias of the point estimate, providing a more precise and reliable measure than the one obtained without re-sampling.

As already stated, the analysis proposed is performed on households in which both the reference person and the second reference person are employees. When estimating individual labour income risk scores, the covariance of RP and RP2's labour incomes is introduced as a measure for their joint risk. The very same procedure described in the present and in the previous sections is then replicated introducing unemployment as a source of risk. Unemployment is the main source of volatility for most employee positions. Not taking unemployment into account would consequently represent, if not a mistake, a severe carelessness. The following vector of risk scores is henceforth estimated:

$$\begin{bmatrix} \sigma_i^{HH} & \sigma_i^{RP} & \sigma_i^{RP2} & cov_i & \tilde{\sigma}_i^{RP} & \tilde{\sigma}_i^{RP2} & c\tilde{o}v_i \end{bmatrix} \quad (7)$$

where the measures with tilde take into account unemployment as a source of labour income risk.

5 Estimations

The aim of the present section is to discuss the results of the econometric estimations performed. Subsection 5.1 delineates the estimation problem, framing it into economic theory and the literature on the topic. Subsection 5.2 describes the estimation technique chosen. Subsection 5.3 describes the model specification. Section 5.4 contains descriptive statistics of the regressors. Section 5.5 comments on the estimation results.

5.1 Theoretical framework

This paper provides empirical evidence through a reduced form estimation strategy[14]. As already underlined above, most of the households (or individuals) **do not invest at all in risky assets**.

[14]Several scholars approached this research question using structural equation modelling. This is usually followed by a calibration of the result obtained (as in Haliassos & Michaelidis, 2003) or by an econometric estimation based the Indirect Inference Approach (see Sanromar, 2014; Alan, 2012; Alan, 2006). This latter technique involves the minimization of the distance between an estimated vector of parameter and a simulated vector of structural ones. The estimated model represents the restricted one. Keane and Smith (2004) describe this procedure using the triade of classical maximum likelihood tests. Sanromar (2014) for example, who estimates both

This implies that the dependent variable, in this kind of problem, is generally characterized by a mass point at zero, and a long tail. The literature has approached reduced form estimation of household investment in risky assets using three different interpretations:

1. one possibility is to study the mere choice of participating in the risky assets market. This first alternative would lead to the estimation of a binary choice model;

2. a second possibility is to treat the choice of not-investing as a possible investment choice (the household invests a "zero amount"). An estimation of this kind is usually described as a "data censoring" problem (Miniaci & Weber, 2002). The techniques used are generally Tobit or negative binomial regressions;

3. a third possibility is to divide the participation choice from the allocation one, under the assumption that those who decided to participate in the stock market are not a random sample. This third alternative would require the implementation of a Heckman sample selection approach.

All strategies has been followed in the literature. Haliassos & Bertaut (1995), Bertaut (1998), King & Leape (1998), Arrondel & Calvo Pardo (2010) follow the first one. Guiso, Japelli & Terlizzese (1996) and Chang, Hong & Karabarbounis (2012) follow instead the second approach. The third methodology is becoming increasingly more popular. Perraudin & Sorensen (2000) implement this methodology[15]. Arrondel & Calvo Pardo follow this approach as well in the cited contribution.

In the opinion of the author the choice of investing can not be disentangled from the choice of how much to invest. As a consequence, the present paper mainly concentrates on the second alternative, although in Appendix 4 it provides estimation results for a mere participation choice and for a Heckit model. The following subsection describes the proposed estimation technique.

5.2 The estimation technique

As stated in the previous subsection, a dependent variable with a mass point at zero and a long right tail calls for a data censoring estimation. Given the non-linearity of the data, OLS regression would indeed lead to unreliable results. As Santos Silva, Tenreyro & Wei (2013) mention, given the nature of the dependent variable "the partial effect of the regressors on the conditional mean of the dependent variable cannot be constant and must approach zero as the conditional mean approaches its bounds. Therefore, ignoring the nature of the data and simply using OLS, is likely to lead to erroneous conclusions because the linear model assumes that the partial effects are constant". A likelihood-based estimation taking into account of the non-linearity of the dependent variable is henceforth needed. A technique typically used to estimate this kind of models is TOBIT (see for example Guiso, Japelli & Terlizzese, 1996). However, TOBIT is suitable for corner solution problems, that is when the dependent variable takes,

the restricted model and the unrestricted one and minimizes the distance between the two vectors, implements a LR approach. Alan (2012) implements a Wald - like one. The theoretical model typically involves an intertemporal utility maximization problem in which individuals earn a positive labour income, consume, and invest their savings into different assets (usually the risk-free asset and a risky one, plus liquidity). Cash-in-hand, defined as the sum of labour income and asset returns, represents the state variable of the intertemporal maximization problem. Labour income is usually modelled as a process with a deterministic (permanent) and stochastic (transitory) part. Depending on the model considered, different assumption are added to the model specification, as probability for disastrous events (Alan, 2012), participation costs (Sanromar, 2014; Alan, 2006), transaction costs (Bonaparte, Cooper & Zhu, 2012), participation *and* transaction costs (Vissing-Jorgensen, 2002), borrowing constraints (Haliassos & Michaelidis, 2003) monitoring costs (Perraudin & Sorensen, 2000)

[15] The authors underline how "one may decompose household portfolio choice into the choice of which assets to include in the portfolio and how much of those assets to hold", Perraudin & Sorensen, (2000).

say, negative values but the number observed is instead zero. However, the zero observed in the distribution of risky assets choice are **actual** zeros, as individuals literally choose not to invest in risky assets. Moreover, "standard count data models, such as Poisson and negative binomial regressions ... ignore the upper bound (of the dependent variable) and therefore are also unsatisfactory" (Santos-Silva & Tenreyro, 2013). Santos Silva, Tenreyro & Wei (2013) consequently suggest a different approach is needed. In their paper, the authors study the choice of two countries to trade on a certain number of sectors. In analysing this problem, the authors notice that:

1. the distribution of the dependent variable shows a mass point at zero (as countries do not trade in all sectors);

2. the dependent variable shows an upper bound, as the number of sector is not infinite.

The authors propose an alternative, flexible, estimation technique taking into account the peculiarities of the dependent variable. In particular:

$$E[A_{it} \mid x_{it}] = 1 - (1 - \omega e^{x_{it}\beta})^{\frac{-1}{\omega}} (8)$$

where A_{it} is a variable bounded between zero and one, and the ω parameter is a shape parameter that adjusts the skewness of the distribution. In particular, this parameter " allows the distribution to be symmetric ($\omega = 1$), left-skewed ($\omega < 1$), or right-skewed ($\omega > 1$)". The interpretation of the coefficients in not straightforward, in this technique[16]. Indeed "the estimates of β are not very informative. Therefore, inference should focus on the partial effects of the regressors of interest and not on the parameter estimates per se."[17]. What matters, in terms of interpretation, is the significance level, the sign, and the manitude of the ω paramater. The authors show that the technique they propose fits the model better than previously applied methods (OLS, Tobit, Pseudo Poisson Maximum Likelihood) and enhances the efficiency of the estimator.

The following graph shows how the dependent variable fits with the characteristics described above:

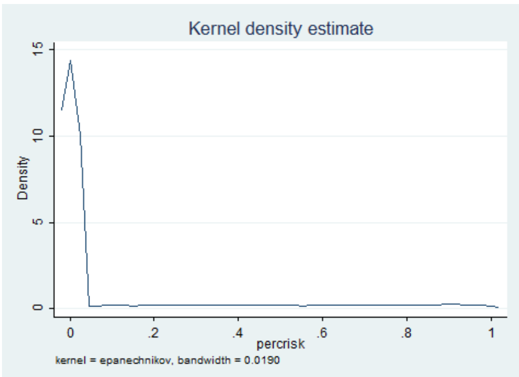

Figure 14: Percentage of risky assets over financial wealth

[16]Santos Silva et alii, 2013 , page 4
[17]Santos Silva et alii, 2013 , page 5

As the problem described in Santos Silva, Tenreyro & Wei recent contribution closely resembles the one approached in this paper, the estimation technique chosen will be the one proposed by the authors.

5.3 The model specification

The present subsection describes the specification proposed and the regressors choice. The specification proposed in this paper is the following:

$$\frac{A}{F}_{it} = g(\sigma_i\beta + x_{it}\gamma + x_{it}^{RP}\gamma^{RP} + x_{it}^{RP2}\gamma^{RP2} + z_i\delta + w_t\rho + \varepsilon_{it}) \quad (9)$$

where:

- $\frac{A}{F}_{it}$ is risky asset A divided by F, financial wealth;
- $g()$ is the functional form describe in the previous subsection;
- vector σ_i contains the labour income risk scores. When measured at a household level, the vector is a scalar and simply contains the standard deviation of the second stage regression residual. When measured at an individual level it contains the labour income risk score for both reference person one and reference person two, and the covariance between the two labour incomes. Individual labour income risk score can contain or not unemployment as a possible source of risk. Section 4 details about the construction of this vector. Labour income risk scores are overall expected to have a negative sign;
- vector x_{it}^{HH} contains household level variables, such as financial exposition, real wealth, financial wealth, number of children. Using the notation introduced in 2.1., financial exposition is defined as PF/GW, where PF represents financial liabilities and GW represents gross wealth. Financial wealth is expressed in relative terms, as $(AF1+AF2+AF3)/GW$. For collinearity reasons, real wealth is approximated using a dummy equal to one if the household owns the house in which it lives [18]. The literature on the topic underlines that investment in risky assets is positively correlated with the overall household level of "richness", and in particular with financial wealth (Haliassos & Michaelidis (2002); Guiso et alii (2002); Chang, Kong & Karabarbounis, 2012) and real wealth (**find a reference about the correlation between real wealth and risky assets)** and labour income (Campbell, (2006); Arrondel & Calvo Pardo (2013)). Financial exposition is instead expected to be negatively correlated, as as Arrondel & Calvo-Pardo (2013) notice;
- vectors $x_{it}^{RP}, x_{it}^{RP2}$ contain individual characteristics (age, sex, education) and labour market variables attributable respectively to reference person one and reference person two. In particular, they contain log of labour income level and log of labour income level squared, worked hours (per week), occupation status. Labour income is the natural logarithm of individual real labour income, deflated using an OECD CPI index. Labour income is expected to be positively correlated (Campbell, 2006; Arrondel & Calvo Pardo, 2013). The

[18] Housing on average represents 70% of household's real wealth

literature on the topic provides empirical evidence that risky assets investment is positively correlated with age (Chang, Kong & Karabarbounis,2012) and education (Campbell, 2006; Cooper & Zhu, 2013). Women are generally found to be more risk averse than men, and being female is expected to be negatively correlated with the investment decision (**put reference**);

- vector z_i contains geographical indicators (area of origin, size of the town of origin);
- vector w_t contains year dummies and regional unemployment rate (disentangled by gender).

Vector z_i is intended to capture geographical or cultural differences in investment behaviour. Italy is indeed characterized by deep north / south differences, the northern part being an more dynamic and overall richer, as opposed to a traditional, agricultural south. Italy still shows deep cultural differences between provinces and cities, captured by the town size dummies.[19]

Finally, based on the reasoning that the business cycle is likely to affect portfolio choices, the specification proposed introduces year dummies, and a measure for unemployment rate. This data, taken from the national statistics institute, are regional unemployment rates disentangled by gender. A negative correlation between unemployment rate and risky assets investment is expected.

A similar specification has been proposed in Knupfer, Rantapuska & Sarvimaki (2013), albeit using different estimation techniques.

5.4 Descriptive Statistics

The aim of the present subsection is to provide descriptive statistics of the variables chosen for the estimation model. The following table displays the estimated values:

[19] As maybe all countries, though.

[h!]

Table 10: Descriptive Statistics

Variable	Mean	Std. Deviation
Age (RP)	45.00	8.15
Education (RP)	3.78	0.81
Female (RP)	1.08	0.28
Age (RP2)	42.17	7.48
Education (RP2)	3.89	0.74
Woked Hours (RP)	38.40	6.98
Boss (RP)	0.16	0.36
Bluecollar (RP)	0.28	0.45
Worked Hours (RP2)	33.51	8.71
Boss (RP2)	0.07	0.26
Bluecollar (RP2)	0.26	0.44
Labour Income (RP)	20601.27	11116.14
Labour Income (RP2)	16884.84	6790.6
Number of Children	1.51	0.84
South	0.22	0.41
North	0.60	0.48
Big	0.07	0.25
Small	0.10	0.30
Own the House	0.77	0.42
Percentage of Financial Wealth	0.20	0.27
Financial Exposition	0.11	0.46
Percentage of Risky Asstes	0.19	0.30

The reference person is on average older than than reference person 2 (45 years against 42.17), and male in most of the cases (recall that female is equal to two, and male is equal to one). They generally work more hours per week (38.40 against 33.51) and they usually have managerial positions (16% for the reference person, against 7% for reference person 2). They earn sensibly more on average (as already shown above). The households considered have on average 1.5 children, mainly live in the north (60% of the sample) in a medium sized town. Their wealth is mainly concentrated on housing (77%). Their level of financial exposition is about 11% with respect of their total wealth. 19% of their financial wealth is invested in risky assets.

5.5 Estimation Results

The regressions are performed on a sub-sample of households where both the reference person and the second reference person are employed, earn a positive labour income, and are observed at least 5 times. Depending on the specification proposed, the sample size slightly changes. This is due to collinearity reasons and to the fact that the dependent variable shows missing values as well. As a baseline, the sub-sample herein described is the one showing non-missing covariance. The sub-sample so defined amounts to 327 households, for a total of 2244 observations, when unemployed is not taken into account as a labour income risk. When unemployment is introduced, 337 households are in the sub-sample for a total of 2313 observations.

The dependent variable, the percentage of risky assets over financial wealth, is defined as:

$$AF3_{it}/(AF1_{it} + AF2_{it} + AF3_{it})$$

where AF3 contains "bonds, mutual funds, equity, shares in private limited companies, foreign securities, loans to cooperatives" (Bank of Italy, (2013)).

The following tables in Appendix 3 display the main regression results. The estimation technique chosen is the one described in section 5.2 . The specification proposed is the one described in section 5.3 . In table 5 labour income risk is measured at a household level. In table 6 labour income risk is measured at an individual level, and does not contain unemployment as a possible source of labour income risk. In table 7 labour income risk is measured at an individual level, and contains unemployment as a possible source of labour income risk. Table 8 contains robustness checks. Namely, the models contained in tables 5, 6 and 7 are estimated using different econometric techniques OLS, pseudo poisson ML, Tobit).

The main result is that, overall, labour income risk shows, as expected, a negative and highly significant coefficient. The significance is robust to the different specifications proposed. This finding is consistent with the literature on "temperance", predicting that a higher exposition to labour income volatility drives the choice of investing less in risky financial assets (Kimball, (1993); Gollier & Pratt, (1996)). More in detail, table 5 shows how labour income risk, calculated at a household level, is negatively correlated with the choice of investing in risky assets. This result is robust to different specifications. Tables 6 and 7 show that, when labour income risk is disentangled for reference person 1, reference person 2 and joint risk (covariance) the only variable that maintains a significant, negative sign is the joint risk measure. Table 8 shows consistent results with respect to the ones contained in tables 5, 6 and 7. The omega parameter is greater than one in all regressions. This is the expected result, as the distribution of the dependent variable is clearly right-skewed.

The following comments apply to tables 5, 6, 7 and 8. In the regressions displayed, age and education show the expected, positive sign. However, the coefficients do not show significance. The variables are indeed likely to be correlated with labour market indicators, the introduction of which reduces the significance of the considered regressors. The coefficient of the female dummy does not show the predicted, negative sign. This results can be attributed to sampling reasons. On the contrary, it may indeed represent an actual characteristic of the population this paper wants to draw inference about. Indeed, in households where the head is female the percentage of financial wealth invested in risky assets is actually higher than in households where the head is male. In the sub-sample considered, when the head is male, the average percentage is 17.9%, whilst when the head is female this percentage is 22.1%. It is worth noticing that for collinearity reasons only one of the dummies is introduced in the model, as the sex of the reference person implicitly defines the sex of reference person two[20]. The number of children shows the predicted negative sign. More numerous households' perceived risk is likely to be higher. The coefficient, however, does not show significance. Households from the north invest more than households from the south (as expected), and households living in small towns invest overall less than household living in bigger centres. Labour market features do not seem to capture much of the model significance. Being a blue-collar (for both reference person one and two) has a negative significant correlation with the investment in risky assets, and this result is robust to all specifications. Labour income does not show significance Real wealth, financial wealth are significant with the expected signs. Unemployment rate displays a negative significant coefficient, as expected. Financial exposition is unexpectedly not significant, although it shoes the expected negative sign.

6 Conclusions

The aim of the present paper is to analyse the correlation between labour income risk and household portfolio allocation in risky assets. To study the proposed research question, the paper firstly investigates the covariance structure of earnings in Italy. The results found, in line with the literature on the topic, have been used to construct a labour income volatility measure to proxy labour income risk at a household and at an individual level. The variables so estimated have been in turn introduced as regressors in reduced form estimations modelling the household decision of investing in risky assets. The regression results are in line with the literature on the topic. In particular, household labour income risk is negatively correlated with household investment in risky assets, as expected. However, disentangling labour income risk by household heads ("husband and wife") does not add explicative power to the model, whilst a measure for joint risk shows a negative and significant correlation. The result is robust to different estimation techniques. Individual labour income risk does not help in explaining the household decision of investing in risky assets.

Appendix 1 - Subsample selection

The aim of the present appendix is to provide a statistical investiagtion of the SHIW database aimed at choosing the sub-sample on which to perform the estimations. The choice of concentrating on labour income is theory driven, as underlined in the introductory section. However, the analysis proposed in the present appendix provides a statistical justification for concentrating on household in which "husband and wife" are both employees. Starting from descriptive statistics on household size and composition, the analysis shows how Italian households' income is generally skewed towards labour earnings. Moreover, it shows that, if both individuals are employees, basically 99% of household income is in the hands of the two main earners ("husband and wife"). Although driven by economic reasons, the sub-sample selection has precise statistical legitimacy. The following analysis aims at furnishing it.

[20] As in Italy only heterosexual marriage is allowed, and reference person two is an individual "married with the reference person".

The following table displays household composition, disentangled by household size.

Table 1: Household composition by household size

Hhsize	RP	Married to RP	Child of RP	Other	Total
1	17,223 0.071	0 0.000	0 0.000	0 0.000	17,223 0.071
2	24,057 0.099	18,213 0.075	3,761 0.015	2,081 0.009	48,112 0.198
3	19,430 0.080	16,773 0.069	19,436 0.080	2,641 0.011	58,280 0.240
4	18,944 0.078	17,997 0.074	36,402 0.150	2,345 0.010	75,688 0.312
5	5,978 0.025	5,670 0.023	16,126 0.066	2,096 0.009	29,870 0.123
Total	87,766	60,642	83,180	11,122	242,710

Sample Size: 242,710

94% of the households in the sample have up to five members. 60% of the individuals in the sample are either reference persons (RP) or individuals married with the reference person. It is clear, from the data, that Italian household have a traditional composition (husband and wife, and a "couple of children"). Numerous families have a minor role. Given the purpose of concentrating on household portfolio decisions, it appears natural to concentrate on the "husband and wife" bundle. This rules out, for example, income pertaining to the children.

The following table displays the individual order number (that is, the individual ranking in household income contribution) if he or she is married to the reference person.

Table 2: Number of order when the individual is married to the reference person

Nord	Frequency	Percentage
2	60228	99.32%
3	320	0.53%
4	76	0.12%
5	8	0.01%
6	7	0.01%
8	3	0.00%
Total	60642	100.00%

The data show that an individual married to the reference person is also the second main earner in the household in 99% of the cases. This table strengthens the choice of concentrating on the "RP and RP2" nucleus in studying household portfolio choices.

The following table display the relative shares of household income by income types. The purpose of this analysis is understanding whether it is worth focusing the attention on labour income, from a mere statistical perspective. The following shares refer to the pooled mean of real values for the income types described in section 2.1. Nominal values are deflated using a 2010 CPI index taken from OECD.

Table 3: Shares of income sources

Labour Income	44.71%
Self Employment Income	15.59%
Financial Income	19.82%
Transfers (retirement)	19.87%
Total	100%

Table 3 shows how labour income represents the lion's share of household income (almost 45% on average). Studying labour income volatility appears consequently a relevant choice from a statistical point of view.

It if finally worth understanding whether concentrating on couples who are both in the labour market is a relevant statistical choice. The following table displays the relative importance of each family member in determining household labour income **when the reference person is employed**. Table 4 displays real values.

To conclude, in households where the reference person is earning a positive amount, if the person married to the reference person is employed as well, statistically speaking the whole household income comes from labour market activities. It is hence rather natural that wealth decision, and in particular portfolio choices, are in the hands of these two individuals. This paper

Table 4: Relative contribution of household members to household income - Pooled means

	Real Labour Income	Share of Household Income
RP	19243.62	0.684
Married to RP	13043.42	0.463
Children of RP	2904.01	0.103
Other	9136.78	0.325
	HH Labour Income	28147.29

consequently concentrates its attention on a sub-sample of households where both the reference person and the second reference person are employees. The sample so found has a numerosity of 327 households, observed 2032 times.

Appendix 2 - Graphs

Figure 1: Real earnings divided by gender

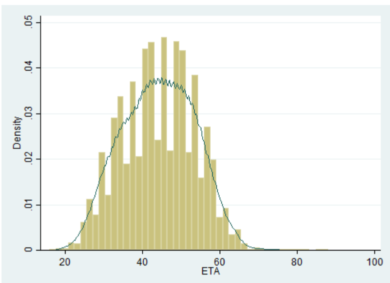

Figure 2: Age distribution - RP

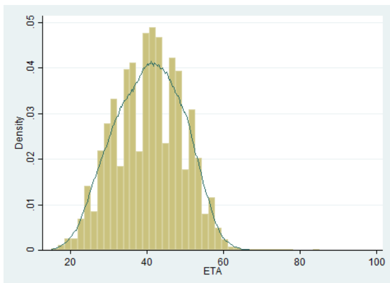

Figure 3: Age distribution - RP2

Appendix 3 - Estimations - Flexible Regression Models

The present appendix displays the results of the flexible regression models, as described in section 5.

Table 5: Flexible Regression Model - Household Risk
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk (HH)	**-5.168**	**-8.492****	**-2.863***	**-1.714***	**-3.132**	**-1.416****	**-1.290***
	(-0.570)	**(-2.036)**	**(-1.763)**	**(-1.716)**	**(-1.450)**	**(-2.282)**	**(-1.923)**
Age	-	0.132	0.041	0.032	0.065	0.024	0.014
		(1.179)	(1.470)	(1.128)	(1.258)	(1.115)	(0.902)
Education	-	0.388	0.352	0.121	0.253	0.096	0.126
		(1.237)	(1.435)	(0.800)	(0.861)	(0.862)	(1.100)
Sex	-	1.325		0.351	1.054	0.463	0.593**
		(1.220)		(0.805)	(1.103)	(1.355)	(2.113)
Age (Rp2)	-	0.027	0.035	0.013	0.020	0.015	0.003
		(0.529)	(1.014)	(0.967)	(0.724)	(1.204)	(0.211)
Education (RP2)	-	0.305	0.072	-0.064	-0.084	-0.074	-0.049
		(0.757)	(0.474)	(-0.840)	(-0.448)	(-0.904)	(-0.496)
Number of Children	-	-	-0.117	-0.073	-0.110	-0.057	-0.063
			(-1.242)	(-1.491)	(-1.015)	(-1.104)	(-0.997)
South	-	-	-1.213	-0.757*	-1.427	-0.729**	-0.142
			(-1.593)	(-1.871)	(-1.568)	(-2.201)	(-0.594)
North	-	-	0.379	0.237	0.321	0.216	0.134
			(1.365)	(1.216)	(1.308)	(1.449)	(0.995)
Small Town	-	-	-0.561	-0.367	-0.758	-0.377	-0.508**
			(-1.345)	(-1.213)	(-1.331)	(-1.541)	(-2.151)
Big City	-	-	-0.272	-0.110	-0.375	-0.172	-0.068
			(-0.653)	(-0.530)	(-0.831)	(-0.682)	(-0.369)
Worked Hours (RP)	-	-	-	0.004	0.002	-0.002	-0.002
				(0.462)	(0.125)	(-0.194)	(-0.243)
Boss (RP)	-	-	-	0.142	-0.014	-0.006	0.099
				(0.889)	(-0.045)	(-0.046)	(0.594)
Blue Collar (RP)	-	-	-	-0.110	0.075	-0.021	-0.130
				(-0.818)	(0.207)	(-0.155)	(-0.874)
Worked Hours (RP2)	-	-	-	-0.008	-0.012	-0.008	-0.012
				(-1.293)	(-0.872)	(-0.997)	(-1.384)
Boss (RP2)	-	-	-	-0.041	-0.248	-0.088	-0.091
				(-0.252)	(-0.600)	(-0.513)	(-0.453)
Blue Collar (RP2)	-	-	-	-0.313	-0.424	-0.319*	-0.274*
				(-1.640)	(-1.610)	(-1.888)	(-1.744)
Labour Income (RP)	-	-	-	-	-7.779	-1.831	-1.597
					(-0.971)	(-0.607)	(-0.572)
Labour Income (RP2)	-	-	-	-	-4.022	-0.931	-2.135
					(-0.953)	(-0.403)	(-0.774)
Labour Income Squared (RP)	-	-	-	-	0.445	0.107	0.096
					(0.998)	(0.690)	(0.681)
Labour Income Squared (Rp2)	-	-	-	-	0.219	0.049	0.123
					(0.936)	(0.389)	(0.806)
Own the House	-	-	-	-	-	1.206**	1.139***
						(2.386)	(3.656)
Financial Wealth	-	-	-	-	-	1.989*	2.101***
						(1.896)	(2.928)
Financial Exposition	-	-	-	-	-	-0.594	-0.823
						(-1.459)	(-1.641)
Unemployment Rate	-	-	-	-	-	-	-0.117**
							(-2.509)
Year Dummies	-	-	-	-	-	-	x
Observations	2,170	1,976	1,976	1,951	1,951	1,951	1,951

Table 6: Flexible Regression Method - Disentagled Risk
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk (RP)	**-0.168**	**-0.400**	**1.014**	**0.400**	**1.922**	**1.204**	**0.838**
	(-1.300)	**(-1.120)**	**(0.705)**	**(0.662)**	**(1.142)**	**(0.578)**	**(1.045)**
Risk (RP2)	**-0.516*****	**-0.416**	**-0.357**	**-0.336**	**-0.917**	**-0.386**	**-0.615**
	(-3.211)	**(-1.368)**	**(-0.676)**	**(-0.903)**	**(-0.937)**	**(-0.596)**	**(-1.326)**
Covariance	**-1.699*****	**-5.298*****	**-5.150****	**-3.556****	**-5.286****	**-3.965***	**-3.288****
	(-4.462)	**(-2.755)**	**(-1.978)**	**(-2.327)**	**(-2.205)**	**(-1.713)**	**(-1.967)**
Age	-	0.017	0.052	0.032*	0.057	0.032	0.011
		(1.524)	(1.041)	(1.781)	(1.311)	(0.921)	(0.643)
Education	-	0.103*	0.396	0.165	0.322	0.212	0.171
		(1.661)	(1.020)	(1.234)	(1.014)	(0.576)	(1.243)
Sex	-	0.213	0.504	0.417	1.096	0.724	0.654**
		(1.485)	(0.861)	(1.346)	(1.374)	(0.957)	(2.228)
Age (Rp2)	-	0.015	0.026	0.016	0.028	0.024	0.004
		(1.371)	(0.790)	(1.056)	(0.888)	(0.707)	(0.242)
Education (RP2)	-	0.034	0.097	-0.067	-0.110	-0.065	-0.055
		(0.622)	(0.474)	(-0.732)	(-0.554)	(-0.470)	(-0.470)
Number of Children	-	-	-0.104	-0.085	-0.140	-0.078	-0.074
			(-1.036)	(-1.564)	(-1.142)	(-0.913)	(-1.051)
South	-	-	-1.277	-0.833***	-1.625	-1.044	-0.184
			(-1.134)	(-2.734)	(-1.568)	(-0.995)	(-0.696)
North	-	-	0.458	0.294*	0.389	0.347	0.172
			(1.111)	(1.831)	(1.418)	(0.959)	(1.127)
Small Town	-	-	-0.666	-0.416*	-0.826	-0.550	-0.562**
			(-1.046)	(-1.768)	(-1.387)	(-0.972)	(-2.233)
Big City	-	-	-0.290	-0.119	-0.401	-0.352	-0.095
			(-0.546)	(-0.625)	(-0.832)	(-0.517)	(-0.463)
Worked Hours (RP)	-	-	-	0.003	-0.008	-0.004	-0.006
				(0.346)	(-0.504)	(-0.323)	(-0.606)
Boss (RP)	-	-	-	0.122	-0.177	-0.081	0.053
				(0.858)	(-0.504)	(-0.295)	(0.292)
Blue Collar (RP)	-	-	-	-0.107	0.081	0.034	-0.151
				(-0.765)	(0.232)	(0.110)	(-0.933)
Worked Hours (RP2)	-	-	-	-0.009	-0.015	-0.013	-0.015
				(-1.284)	(-0.960)	(-0.668)	(-1.433)
Boss (RP2)	-	-	-	-0.048	-0.281	-0.129	-0.094
				(-0.269)	(-0.623)	(-0.418)	(-0.412)
Blue Collar (RP2)	-	-	-	-0.336**	-0.437	-0.403	-0.272
				(-2.071)	(-1.501)	(-1.162)	(-1.575)
Labour Income (RP)	-	-	-	-	-6.525	-1.852	-0.374
					(-0.691)	(-0.242)	(-0.126)
Labour Income (RP2)	-	-	-	-	-5.202	-2.438	-3.062
					(-0.947)	(-0.487)	(-1.139)
Labour Income Squared (RP)	-	-	-	-	0.399	0.128	0.042
					(0.758)	(0.290)	(0.282)
Labour Income Squared (Rp2)	-	-	-	-	0.278	0.129	0.172
					(0.934)	(0.477)	(1.153)
Own the House	-	-	-	-	-	1.614	1.197***
						(1.356)	(3.708)
Financial Wealth	-	-	-	-	-	2.911	2.334***
						(1.039)	(3.270)
Financial Exposition	-	-	-	-	-	-0.679	-0.878
						(-1.404)	(-1.593)
Unemployment Rate	-	-	-	-	-	-	-0.127***
							(-2.657)
Year Dummies	-	-	-	-	-	-	x
Observations	2,170	1,976	1,976	1,951	1,951	1,951	1,951

Table 7: Flexible Regression Method - Disentagled Risk, Unemployment Included
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk (RP)	**-0.131**	**-0.875**	**-0.333**	**-0.403**	**-0.227**	**-0.239**	**-0.181**
	(-1.416)	**(-1.084)**	**(-0.871)**	**(-1.170)**	**(-0.493)**	**(-0.457)**	**(-0.486)**
Risk (RP2)	**-0.179****	**-0.139**	**-0.072**	**-0.055**	**-0.052**	**-0.104**	**-0.130**
	(-1.964)	**(-1.146)**	**(-0.519)**	**(-0.460)**	**(-0.248)**	**(-0.748)**	**(-0.938)**
Covariance	**-1.461*****	**-4.861**	**-4.385****	**-3.409****	**-4.346****	**-3.158**	**-3.103***
	(-5.864)	**(-1.487)**	**(-2.558)**	**(-2.208)**	**(-2.439)**	**(-1.280)**	**(-1.942)**
Age	-	0.024	0.039*	0.031**	0.042	0.020	0.011
		(1.019)	(1.652)	(2.051)	(0.995)	(0.772)	(0.692)
Education	-	0.148	0.278*	0.145	0.202	0.107	0.156
		(1.352)	(1.848)	(1.417)	(0.723)	(0.687)	(1.261)
Sex	-	0.278	0.340	0.389	0.746	0.445	0.606**
		(0.823)	(1.096)	(1.620)	(0.891)	(0.893)	(2.247)
Age (Rp2)	-	0.019	0.019	0.016	0.023	0.017	0.007
		(0.910)	(1.063)	(1.193)	(0.869)	(1.068)	(0.435)
Education (RP2)	-	0.027	0.056	-0.064	-0.099	-0.085	-0.061
		(0.262)	(0.472)	(-0.773)	(-0.660)	(-1.056)	(-0.596)
Number of Children	-	-	-0.089	-0.085*	-0.115	-0.068	-0.073
			(-1.412)	(-1.695)	(-1.155)	(-1.351)	(-1.122)
South	-	-	-0.923**	-0.772***	-1.131	-0.709	-0.129
			(-2.352)	(-3.615)	(-1.161)	(-1.572)	(-0.524)
North	-	-	0.306	0.252*	0.306	0.201	0.136
			(1.360)	(1.762)	(1.289)	(0.750)	(0.951)
Small Town	-	-	-0.514*	-0.411**	-0.625	-0.389	-0.550**
			(-1.864)	(-2.125)	(-1.060)	(-1.139)	(-2.292)
Big City	-	-	-0.199	-0.142	-0.287	-0.186	-0.095
			(-0.784)	(-0.854)	(-0.692)	(-0.551)	(-0.509)
Worked Hours (RP)	-	-	-	0.002	-0.004	-0.004	-0.005
				(0.266)	(-0.333)	(-0.445)	(-0.562)
Boss (RP)	-	-	-	0.144	-0.051	-0.007	0.091
				(1.113)	(-0.187)	(-0.053)	(0.547)
Blue Collar (RP)	-	-	-	-0.100	-0.008	-0.007	-0.127
				(-0.773)	(-0.029)	(-0.061)	(-0.802)
Worked Hours (RP2)	-	-	-	-0.008	-0.012	-0.008	-0.014
				(-1.352)	(-0.880)	(-0.747)	(-1.421)
Boss (RP2)	-	-	-	-0.053	-0.211	-0.087	-0.105
				(-0.327)	(-0.541)	(-0.453)	(-0.503)
Blue Collar (RP2)	-	-	-	-0.324**	-0.380	-0.295	-0.263
				(-2.245)	(-1.460)	(-1.322)	(-1.613)
Labour Income (RP)	-	-	-	-	-4.723	-1.487	-1.265
					(-0.377)	(-0.659)	(-0.452)
Labour Income (RP2)	-	-	-	-	-3.186	-0.518	-2.404
					(-0.661)	(-0.098)	(-0.781)
Labour Income Squared (RP)	-	-	-	-	0.278	0.090	0.081
					(0.407)	(0.759)	(0.573)
Labour Income Squared (Rp2)	-	-	-	-	0.174	0.028	0.138
					(0.665)	(0.100)	(0.817)
Own the House	-	-	-	-	-	1.196	1.149***
						(1.515)	(4.027)
Financial Wealth	-	-	-	-	-	1.928	2.154***
						(1.120)	(3.194)
Financial Exposition	-	-	-	-	-	-0.590	-0.841
						(-1.397)	(-1.599)
Unemployment Rate	-	-	-	-	-	-	-0.123**
							(-2.515)
Year Dummies	-	-	-	-	-	-	x
Observations	1,976	1,976	1,976	1,951	1,951	1,951	1,951

Table 8: Robustness Check - Different Estimation Techniques

	Bootsrapped OLS (250 replications)			Tobit			ppml		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Risk (Household)	**-0.175**** **(-2.337)**	-	-	**-0.368*** **(-1.708)**	-	-	**-0.949** **(-1.379)**	-	-
Risk (RP)	-	**0.098** **(1.239)**	-	-	**0.293** **(1.538)**	-	-	**0.354** **(0.858)**	-
Risk (RP2)	-	**-0.053** **(-1.142)**	-	-	**-0.176** **(-1.142)**	-	-	**-0.468** **(-1.007)**	-
Covariance	-	**-0.294***** **(-2.888)**	-	-	**-0.637**** **(-2.191)**	-	-	**-2.232**** **(-2.104)**	-
Risk (RP) - Unemployment	-	-	**-0.015** **(-0.348)**	-	-	**-0.061** **(-0.449)**	-	-	**-0.152** **(-0.422)**
Risk (RP2) - Unemployment	-	-	**-0.008** **(-0.460)**	-	-	**-0.047** **(-1.037)**	-	-	**-0.106** **(-0.590)**
Covariance - Unemployment	-	-	**-0.224***** **(-3.073)**	-	-	**-0.538**** **(-2.186)**	-	-	**-2.159**** **(-2.015)**
Personal Characteristics	x	x	x	x	x	x	x	x	x
Geographical Indicators	x	x	x	x	x	x	x	x	x
Labour Market Indicators	x	x	x	x	x	x	x	x	x
Wealth Indicators	x	x	x	x	x	x	x	x	x
Year Dummies	x	x	x	x	x	x	x	x	x
Unemployment	x	x	x	x	x	x	x	x	x
Observations	1,951	1,951	1,951	1,951	1,951	1,951	1,951	1,951	1,951

Appendix 4 - Additional Empirical Evidence

The present appendix provides additional empirical evidence supporting the results found so far. As underlined in section 5.1, this kind of research question can be approached folowing three different empirical strategies. The paper mainly focuses on one of them (treating the choice of not participating in the risky assets market as the choice of investing a zero amount). This appendix briefly investigates the other two possibilities therein discussed.

Participation Choice

To study the correlation between labour income risk and the choice of investing in risky assets, a natural way to proceed is to study the mere choice of participating in the risky assets market, using a binary choice model. The dependent variable would indeed be a binary indicator, equal to one if the household invests in risky assets, equal to zero otherwise. The functional form chosen for the probability density function is the standard normal (PROBIT model). The specification of the models is the following:

$$\Pr(\text{ Participation}_{it} = 1 \mid X) = \phi(\sigma_i\beta + x_{it}\gamma + x_{it}^{RP}\gamma^{RP} + x_{it}^{RP2}\gamma^{RP2} + z_i\delta + w_t\rho + \varepsilon_{it}) \quad (10)$$

where X is the whole set of regressors, ϕ is the standard normal density function, and the choice of regressors is the same described in section 5.3.

The results obtained show some differences with the ones displayed in Appendix 3. The risk variables show the expected negative sign. However, their significance is overall different. The household level risk is generally significant and negatively correlated. However in specification (6) and (7) (table 9) the variable loses significance. Table 10 displays an interesting result. When risk is disentangled between husband and wife, the only variable showing significance is the labour income volatility of reference person two. Recalling that on average the reference person two is female, it can be concluded that in households where the wife has a more volatile income, the choice is to invest less in risky assets. Table 11 does not show particularly interesting results on the risk side.

It is worth noticing that in all tables, the educational level of the reference person, the age of reference person two and the income of reference person two are positively correlated with the choice of investing. These findings are line with the literature on the topic. Financial exposition, house owning and financial wealth are significant, displaying the expected sign. Unemployment rate shows the expected, significant negative correlation.

Table 9: Participation Choice - Probit Models
Dependent Variable : Household Holding of Risky Assets (Binary)

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk (Household)	**-0.901****	**-1.386*****	**-1.009*****	**-0.896****	**-0.820****	**-0.577**	**-0.521**
	(-2.555)	**(-3.618)**	**(-2.584)**	**(-2.267)**	**(-2.010)**	**(-1.364)**	**(-1.141)**
Age	-	0.006	0.011	0.010	0.007	0.003	-0.009
		(0.620)	(1.261)	(0.985)	(0.700)	(0.253)	(-0.821)
Education	-	0.119**	0.193***	0.127**	0.099*	0.116**	0.120**
		(2.571)	(4.055)	(2.284)	(1.737)	(2.013)	(2.224)
Sex	-	0.046	0.081	0.079	0.078	0.108	0.172
		(0.410)	(0.615)	(0.627)	(0.593)	(0.800)	(1.362)
Age (Rp2)	-	0.022**	0.022***	0.023**	0.021**	0.025**	0.019**
		(2.391)	(2.652)	(2.384)	(2.204)	(2.516)	(1.964)
Education (RP2)	-	0.041	0.048	-0.043	-0.083	-0.071	-0.055
		(0.830)	(0.962)	(-0.736)	(-1.342)	(-1.124)	(-0.882)
Number of Children	-	-	-0.063*	-0.069*	-0.076**	-0.061	-0.058
			(-1.694)	(-1.799)	(-1.980)	(-1.553)	(-1.581)
South	-	-	-0.601***	-0.608***	-0.616***	-0.605***	-0.125
			(-5.914)	(-5.815)	(-5.846)	(-5.631)	(-0.847)
North	-	-	0.134	0.139	0.108	0.123	0.045
			(1.611)	(1.639)	(1.264)	(1.416)	(0.522)
Small Town	-	-	-0.169	-0.174	-0.175	-0.184*	-0.236**
			(-1.612)	(-1.595)	(-1.597)	(-1.651)	(-2.097)
Big City	-	-	-0.101	-0.090	-0.099	-0.147	-0.027
			(-0.885)	(-0.783)	(-0.859)	(-1.237)	(-0.214)
Worked Hours (RP)	-	-	-	0.002	-0.001	-0.001	-0.001
				(0.366)	(-0.130)	(-0.190)	(-0.140)
Boss (RP)	-	-	-	0.031	-0.065	-0.063	0.004
				(0.346)	(-0.671)	(-0.638)	(0.039)
Blue Collar (RP)	-	-	-	-0.173**	-0.143	-0.118	-0.201**
				(-1.966)	(-1.600)	(-1.290)	(-2.038)
Worked Hours (RP2)	-	-	-	-0.008**	-0.010**	-0.013***	-0.015***
				(-2.279)	(-2.383)	(-2.825)	(-3.546)
Boss (RP2)	-	-	-	-0.042	-0.118	-0.109	-0.088
				(-0.346)	(-0.940)	(-0.848)	(-0.646)
Blue Collar (RP2)	-	-	-	-0.219**	-0.181*	-0.207**	-0.169*
				(-2.308)	(-1.867)	(-2.090)	(-1.718)
Labour Income (RP)	-	-	-	-	-1.292	-0.923	-0.446
					(-0.980)	(-0.688)	(-0.204)
Labour Income (RP2)	-	-	-	-	-2.226**	-1.854*	-2.405*
					(-2.239)	(-1.808)	(-1.853)
Labour Income Squared (RP)	-	-	-	-	0.079	0.058	0.031
					(1.164)	(0.835)	(0.278)
Labour Income Squared (Rp2)	-	-	-	-	0.127**	0.107*	0.143**
					(2.290)	(1.869)	(1.997)
Own the House	-	-	-	-	-	0.930***	0.827***
						(7.147)	(5.085)
Financial Wealth	-	-	-	-	-	1.753***	1.697***
						(9.397)	(7.879)
Financial Exposition	-	-	-	-	-	-0.325***	-0.417*
						(-2.869)	(-1.805)
Unemployment Rate	-	-	-	-	-	-	-0.085***
							(-4.235)
Year Dummies	-	-	-	-	-	-	x
Observations	2,170	1,976	1,976	1,951	1,951	1,951	1,951

Table 10: Participation Choice - Probit Models
Dependent Variable : Household Holding of Risky Assets (Binary)

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk (RP)	**-0.022**	**-0.203**	**0.251**	**0.216**	**0.444**	**0.425**	**0.338**
	(-0.076)	**(-0.694)**	**(0.820)**	**(0.681)**	**(1.290)**	**(1.216)**	**(0.938)**
Risk (RP2)	**-0.781*** **	**-0.381**	**-0.419***	**-0.436***	**-0.539** **	**-0.453***	**-0.586** **
	(-3.421)	**(-1.621)**	**(-1.730)**	**(-1.739)**	**(-2.027)**	**(-1.661)**	**(-2.080)**
Covariance	**-0.139**	**-0.760**	**-0.766**	**-0.604**	**-0.592**	**-0.413**	**-0.364**
	(-0.275)	**(-1.478)**	**(-1.447)**	**(-1.134)**	**(-1.117)**	**(-0.767)**	**(-0.663)**
Age	-	0.004	0.009	0.008	0.005	0.002	-0.009
		(0.389)	(0.984)	(0.820)	(0.544)	(0.165)	(-0.924)
Education	-	0.124***	0.201***	0.137**	0.108*	0.124**	0.126**
		(2.662)	(4.187)	(2.451)	(1.902)	(2.143)	(2.101)
Sex	-	0.052	0.009	0.071	0.083	0.109	0.153
		(0.466)	(0.079)	(0.559)	(0.626)	(0.812)	(1.035)
Age (Rp2)	-	0.022**	0.022**	0.022**	0.021**	0.024**	0.018*
		(2.360)	(2.315)	(2.339)	(2.138)	(2.430)	(1.729)
Education (RP2)	-	0.040	0.040	-0.054	-0.092	-0.080	-0.062
		(0.798)	(0.789)	(-0.903)	(-1.485)	(-1.256)	(-0.947)
Number of Children	-	-	-0.067*	-0.074*	-0.081**	-0.065	-0.060
			(-1.788)	(-1.929)	(-2.108)	(-1.644)	(-1.456)
South	-	-	-0.607***	-0.622***	-0.632***	-0.619***	-0.157
			(-5.957)	(-5.923)	(-5.972)	(-5.737)	(-1.039)
North	-	-	0.148*	0.149*	0.119	0.133	0.054
			(1.778)	(1.749)	(1.385)	(1.519)	(0.586)
Small Town	-	-	-0.161	-0.164	-0.161	-0.170	-0.220*
			(-1.519)	(-1.499)	(-1.462)	(-1.519)	(-1.923)
Big City	-	-	-0.091	-0.079	-0.088	-0.135	-0.020
			(-0.797)	(-0.687)	(-0.758)	(-1.131)	(-0.161)
Worked Hours (RP)	-	-	-	0.001	-0.003	-0.003	-0.003
				(0.210)	(-0.489)	(-0.477)	(-0.473)
Boss (RP)	-	-	-	0.006	-0.104	-0.097	-0.029
				(0.066)	(-1.059)	(-0.964)	(-0.280)
Blue Collar (RP)	-	-	-	-0.177**	-0.142	-0.120	-0.209**
				(-1.998)	(-1.588)	(-1.317)	(-2.208)
Worked Hours (RP2)	-	-	-	-0.009**	-0.010**	-0.012***	-0.015***
				(-2.447)	(-2.324)	(-2.776)	(-3.309)
Boss (RP2)	-	-	-	-0.039	-0.115	-0.104	-0.084
				(-0.323)	(-0.912)	(-0.809)	(-0.628)
Blue Collar (RP2)	-	-	-	-0.219**	-0.177*	-0.201**	-0.160
				(-2.310)	(-1.833)	(-2.031)	(-1.574)
Labour Income (RP)	-	-	-	-	-0.696	-0.395	-0.060
					(-0.513)	(-0.286)	(-0.041)
Labour Income (RP2)	-	-	-	-	-2.424**	-2.041**	-2.709**
					(-2.435)	(-1.987)	(-2.527)
Labour Income Squared (RP)	-	-	-	-	0.051	0.033	0.013
					(0.741)	(0.470)	(0.176)
Labour Income Squared (Rp2)	-	-	-	-	0.136**	0.115**	0.157***
					(2.449)	(2.017)	(2.634)
Own the House	-	-	-	-	-	0.931***	0.817***
						(7.157)	(5.967)
Financial Wealth	-	-	-	-	-	1.752***	1.692***
						(9.390)	(8.703)
Financial Exposition	-	-	-	-	-	-0.314***	-0.401***
						(-2.799)	(-3.367)
Unemployment Rate	-	-	-	-	-	-	-0.082***
							(-4.092)
Year Dummies	-	-	-	-	-	-	x
Observations	1,976	1,976	1,976	1,951	1,951	1,951	1,951

Table 11: Participation Choice - Probit Models
Dependent Variable : Household Holding of Risky Assets (Binary)

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk (RP) (unemployment)	**-0.334**	**-0.418****	**-0.129**	**-0.158**	**-0.027**	**-0.000**	**-0.061**
	(-1.636)	**(-2.027)**	**(-0.617)**	**(-0.731)**	**(-0.115)**	**(-0.000)**	**(-0.250)**
Risk (RP2) (unemployment)	**-0.129***	**-0.100**	**-0.064**	**-0.050**	**-0.058**	**-0.077**	**-0.091**
	(-1.866)	**(-1.381)**	**(-0.863)**	**(-0.661)**	**(-0.764)**	**(-0.989)**	**(-1.130)**
Covariance (unemployment)	**-0.395**	**-0.734***	**-0.805***	**-0.687**	**-0.719***	**-0.523**	**-0.516**
	(-1.008)	**(-1.843)**	**(-1.941)**	**(-1.643)**	**(-1.724)**	**(-1.225)**	**(-1.188)**
Age	-	0.004	0.009	0.008	0.005	0.001	-0.010
		(0.443)	(0.994)	(0.822)	(0.518)	(0.112)	(-0.975)
Education	-	0.131***	0.204***	0.136**	0.108*	0.127**	0.132**
		(2.811)	(4.230)	(2.438)	(1.895)	(2.177)	(2.181)
Sex	-	0.048	0.013	0.078	0.075	0.099	0.157
		(0.432)	(0.110)	(0.613)	(0.568)	(0.732)	(1.055)
Age (Rp2)	-	0.023**	0.023**	0.024**	0.022**	0.026***	0.020**
		(2.475)	(2.467)	(2.489)	(2.317)	(2.608)	(1.986)
Education (RP2)	-	0.034	0.045	-0.048	-0.091	-0.082	-0.066
		(0.678)	(0.860)	(-0.807)	(-1.452)	(-1.273)	(-1.003)
Number of Children	-	-	-0.070*	-0.076**	-0.084**	-0.067*	-0.063
			(-1.874)	(-1.997)	(-2.177)	(-1.693)	(-1.522)
South	-	-	-0.596***	-0.608***	-0.620***	-0.610***	-0.132
			(-5.839)	(-5.785)	(-5.856)	(-5.645)	(-0.870)
North	-	-	0.137	0.140*	0.110	0.123	0.041
			(1.643)	(1.646)	(1.281)	(1.400)	(0.452)
Small Town	-	-	-0.178*	-0.180*	-0.179	-0.186*	-0.240**
			(-1.692)	(-1.651)	(-1.634)	(-1.665)	(-2.108)
Big City	-	-	-0.108	-0.096	-0.104	-0.152	-0.033
			(-0.951)	(-0.839)	(-0.901)	(-1.268)	(-0.260)
Worked Hours (RP)	-	-	-	0.001	-0.002	-0.002	-0.002
				(0.199)	(-0.334)	(-0.366)	(-0.329)
Boss (RP)	-	-	-	0.025	-0.076	-0.071	-0.002
				(0.272)	(-0.783)	(-0.714)	(-0.015)
Blue Collar (RP)	-	-	-	-0.174**	-0.143	-0.117	-0.201**
				(-1.967)	(-1.602)	(-1.284)	(-2.123)
Worked Hours (RP2)	-	-	-	-0.009**	-0.011**	-0.013***	-0.016***
				(-2.318)	(-2.440)	(-2.850)	(-3.394)
Boss (RP2)	-	-	-	-0.050	-0.125	-0.115	-0.095
				(-0.412)	(-0.992)	(-0.892)	(-0.713)
Blue Collar (RP2)	-	-	-	-0.228**	-0.185*	-0.207**	-0.169*
				(-2.411)	(-1.917)	(-2.093)	(-1.666)
Labour Income (RP)	-	-	-	-	-1.140	-0.799	-0.430
					(-0.839)	(-0.580)	(-0.292)
Labour Income (RP2)	-	-	-	-	-2.231**	-1.895*	-2.464**
					(-2.233)	(-1.835)	(-2.276)
Labour Income Squared (RP)	-	-	-	-	0.072	0.052	0.030
					(1.033)	(0.736)	(0.399)
Labour Income Squared (Rp2)	-	-	-	-	0.128**	0.109*	0.146**
					(2.288)	(1.894)	(2.418)
Own the House	-	-	-	-	-	0.939***	0.831***
						(7.218)	(6.068)
Financial Wealth	-	-	-	-	-	1.757***	1.700***
						(9.416)	(8.737)
Financial Exposition	-	-	-	-	-	-0.327***	-0.417***
						(-2.886)	(-3.469)
Unemployment Rate	-	-	-	-	-	-	-0.085***
							(-4.234)
Year Dummies	-	-	-	-	-	-	x
Observations	1,976	1,976	1,976	1,951	1,951	1,951	1,951

Sample Selection Approach

It can be argued that the choice of participating in the risky assets market can be divided by the choice of how much to invest, once participated. In other words, the empirical research question would present a sample selection problem. Those who decide how much to allocate, decided to participate in the market in first place. To overcome this problem, this paper implements a two-stages ample selection approach (Heckman, 1976).

Identification requires an exclusion restriction, that is a variable (serving as an instrument) that is correlated with the choice of participating, but not correlated with the amount allocated to risky assets. The regressions contained in Appendix 3 and Appendix 4 (above) suggest a possible exclusion restriction. The choice of participating is always correlated with the educational level of reference person one, whilst this variable does not show any correlation with the amount allocated (regressions in Appendix 2). Such an exclusion restriction has already been used by Arrondel & Calvo Pardo (2010). The following table displays the estimation results. Joint risk plays again th main role in explaining the choice of investing in risky assets.

Table 12: Heckman Selection Model - Exclusion Restriction: Education

VARIABLES	(1)	(2)	(3)
Risk (Household)	**-0.205**	-	-
	(-1.135)		
Risk (RP)	-	**-0.043**	-
		(-0.361)	
Risk (RP2)	-	**0.119**	-
		(0.833)	
Covariance	-	**-0.526****	-
		(-2.557)	
Risk (RP) - Unemployment	-	-	**0.009**
			(0.079)
Risk (RP2) - Unemployment	-	-	**0.015**
			(0.519)
Covariance - Unemployment	-	-	**-0.387***
			(-1.898)
Personal Characteristics	x	x	x
Geographical Indicators	x	x	x
Labour Market Indicators	x	x	x
Wealth Indicators	x	x	x
Year Dummies	x	x	x
Unemployment	x	x	x
Observations	2,003	2,003	2,003

References

AbowdCardAbowd, J.M. and D. Card (1989), “On the Covariance Structure of Earnings and Hours Changes”, Econometrica, 57, 411-445.

[0] [1] Alan, S. (2006), “Entry Costs and Stock Market Participation over the life cycle”, Review of Economic Dynamics, 9, 588-611.

[2] Altonji, J. G., Smith, A. A. Jr. Vidangos, I., (2013), “Modeling Earnings Dynamics”, Econometrica, Econometric Society, Econometric Society, vol. 81(4), pages 1395-1454, 07.

[3] Arrondel, L. Calvo Pardo, L. (2009), “Are the French Cautious? Assets and Labour-Market Risk”, Economie et Statistique, Institut National de la Statistique et des Etudes Economiques, vol. 417, pages 27-53., June.

[4] Baker, M. (1997),“Growth-Rate Heterogeneity and the Covariance Structure of Life- Cycle Earnings”, Journal of Labor Economics, 15(2): 338-375.

[5] Borella, M. (2001), “The Error Structure of Earnings: an Analysis on Italian Longitudinal Data”, CeRP Working Papers 07, Center for Research on Pensions and Welfare Policies, Turin (Italy).

[6] Browning, M., Ejrnaes, M. Alvarez, J. (2010), “Modelling Income Processes with Lots of Heterogeneity”, Review of Economic Studies, Oxford University Press, vol. 77(4), pages 1353-1381.

[7] Brunello, G., Comi, S., Brunello, G.(2000) ”The Returns to Education in Italy: A New Look at the Evidence,” IZA Discussion Papers 130, Institute for the Study of Labor (IZA).

[8] Calvet, L., E. Sodini, P. (2013), “Twin picks: Disentangling the determinants of risk-taking in household portfolios”, SAFE Working Paper Series 13, Research Center SAFE - Sustainable Architecture for Finance in Europe, Goethe University Frankfurt.

[9] Campbell J. Y. (2006),“Household Finance,” Journal of Finance, American Finance Association, vol. 61(4), pages 1553-1604, 08.

[10] Chang, Y. Hong, J. Karabarbounis, M., (2013), ”Life Cycle Uncertainty and Portfolio Choice Puzzles,” 2013 Meeting Papers 595, Society for Economic Dynamics.

[11] Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005), “Consumption and portfolio choice over the life cycle,” Review of financial Studies, 18(2), 491-533.

[12] Cooper, R., Zhu, G. (2013) “Household Finance: Education, Permanent Income and Portfolio Choice,” NBER Working Papers 19455, National Bureau of Economic Research, Inc.

[13] Guiso L., Haliassos, M. & Japelli, T. (editors) (2002) “Household Portfolio”, MIT Press.

[14] Guiso, L. ,Jappelli, T., (2002) “Private Transfers, Borrowing Constraints and the Timing of Homeownership,” Journal of Money, Credit and Banking, Blackwell Publishing, vol. 34(2), pages 315-39, May.

[15] Guiso, L., Jappelli, T., Terlizzese, D. (1996) “Income Risk, Borrowing Constraints, and Portfolio Choice,” American Economic Review, American Economic Association, vol. 86(1), pages 158-72, March.

[16] Guvenen, F., Smith, A.A. Jr. (2010), “Inferring Labor Income Risk From Economic Choices: An Indirect Inference Approach,” Unpublished Manuscript, available at http://dx.doi.org/10.2139/ssrn.1649092.

[17] Haliassos, M. Bertaut, C. C. (1995) ”Why Do So Few Hold Stocks?,” Economic Journal, Royal Economic Society, Royal Economic Society, vol. 105(432), pages 1110-29, September.

[18] Haliassos, M., and A. Michaelides (2003): choice and liquidity constraints,” International Economic Review, 44, 143-177.

[19] Heaton, J., Lucas, D. (1997) “Market Frictions, Savings Behavior, And Portfolio Choice,” Macroeconomic Dynamics, Cambridge University Press, vol. 1(01), pages 76-101, January.

[20] Knüpfer, S., Rantapuska, E. H., Sarvimäki, M. (2014), Labor Market Experiences and Portfolio Choice: Evidence from the Finnish Great Depression, unpublished paper.

[21] Lillard, L.A., Weiss Y. (1979), “Components of Variation in Panel Earnings Data: American Scientists 1960-1970”, Econometrica, 47, 437-454.

[22] MaCurdy, T.E. (1982), The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis, Journal of Econometrics, 18, 83-114.

[23] Meghir, C., and Pistaferri, L. (2004), “Income Variance Dynamics and Heterogeneity,” Econometrica, 72 (1), 1–32.

[24] Merton, R. C. (1971) “Optimum consumption and portfolio rules in a continuous-time model,” Journal of Economic Theory, Elsevier, vol. 3(4), pages 373-413, December.

[25] Mussida, C., Picchio, M. (2014) “The trend over time of the gender wage gap in Italy,” Empirical Economics, Springer, vol. 46(3), pages 1081-1110, May.

[26] Perraudin, W. R. M. Sorensen, B. E. (2000), “The demand for risky assets: Sample selection and household portfolios,” Journal of Econometrics, Elsevier, vol. 97(1), pages 117-144, July.

[27] Pelizzon, L. , Weber, G. (2009) ”Efficient portfolios when housing needs change over the life cycle,” Journal of Banking Finance, Elsevier, vol. 33(11), pages 2110-2121, November.

[28] Sanromar, G. (2014) “Cost and preference heterogeneity in risky financial markets” Journal of Applied Econometrics, forthcoming.

[29] Santos Silva, J.M.C. Tenreyro, S., Wei, K., (2014), “Estimating the extensive margin of trade,” Journal of International Economics, Elsevier, vol. 93(1), pages 67-75.

[30] Vissing-Jorgensen, A. (2002), “Limited Asset Market Participation and the Elasticity of Intertemporal Substitution,” Journal of Political Economy, University of Chicago Press, vol. 110(4), pages 825-853, August.