

# Hedging Labor Income Risk over the Life-Cycle

Fabio C. Bagliano<sup>1,3</sup>, Raffaele Corvino<sup>2</sup>, Carolina Fugazza<sup>1</sup>, and Giovanna Nicodano<sup>1,3</sup>

<sup>1</sup>University of Torino, <sup>2</sup>Cass Business School, <sup>3</sup>Collegio Carlo Alberto

**This draft:** February 2019

## Abstract

We show that the decision to participate in the stock market depends on the ability of equities to hedge the individual permanent earnings shocks, consistent with implications of life-cycle models. Those households who refrain from stock investing display positive correlation between their own permanent income innovations and market returns. These results owe to a two-step empirical strategy. First, a minimum distance estimation disentangles the aggregate from the idiosyncratic permanent component of labor income risks. The second step reconstructs the individual life-cycle dynamics of persistent shocks through a Kalman filter applied to the estimated labor income process. We are thus able to obtain the full cross-sectional distribution of individual correlations between permanent shocks and market returns.

*Keywords:* Stock market participation, Labor income-risk return correlation, Permanent income shocks, Kalman filter

*JEL classification:* G10,G11, D14, C15

Address: Dipartimento ESOMAS, Università di Torino, Corso Unione Sovietica 218bis, 10134, Torino (Italy).

E-mails: fabio.bagliano@unito.it; raffaele.corvino@cass.city.ac.uk; carolina.fugazza@unito.it; giovanna.nicodano@unito.it

# 1 Introduction

Consumption smoothing is a central tenet in economics and the implications of the underlying theory are clear-cut. Individual savings and investments should help hedge individual permanent labor income shocks (Haliassos and Michaelides, 2003), that display consequences also in the future and are otherwise uninsurable. Moreover, savers should reduce their investment in equities (Merton, 1969), and possibly stay out of the stock market (Bagliano, Fugazza, and Nicodano, 2014), if their permanent income shocks display positive correlation with stock returns. Indeed, positive correlation would imply that equities contribute to the amplification of permanent earnings shocks. The evidence is less clear-cut, though. First, disentangling the permanent component of earnings shocks from the transitory one is challenging because they are unobservable. Thus, the study of households' hedging motives relies on the correlation between labor income growth and stock returns (see e.g. Heaton and Lucas, 2000, Angerer and Lam, 2009, Bonaparte, Korniotis, and Kumar, 2014). Second, available estimates of such correlation on U.S. data are usually not statistically different from zero, being sample correlation between just a few annual market returns and labor income residuals from a time-series regression based on the Panel Study of Income Dynamics (PSID). Moreover, there is a huge variation in available estimates of the correlation between labor income shocks and stock returns.<sup>1</sup>

In this paper, we devise a novel econometric strategy to estimate individual labor income risk over the life cycle, that allows for an exploration of heterogeneous hedging motives. Once we obtain the dynamics of the unobservable component of labor income at the individual level, we show that both the variability of permanent shocks to income and their correlation with stock returns are significant determinants of the propensity to participate. On the contrary, the correlation with total income shocks is not. Our results confirm the theoretical prediction that what is relevant for households' risk taking decision is the ability of risky financial assets to hedge permanent, instead of pure transitory, labor income risk.

The dominant approach in both asset pricing and household finance rests on the calibration of the individual income process. Given the evidence that agents are subject to substantial and highly persistent shocks to earnings (see e.g. Abowd and Card, 1989), the individual's log labor income is modeled as the sum of a Mincerian function of demographic and personal characteristics (e.g. education) and a stochastic trend hit by permanent and transitory shocks. The former have permanent effects on the level of individual labor income, whereas the latter change only the current level of the income, without effects on the future earnings. Moreover, the permanent shock contains

---

<sup>1</sup> Cocco, Gomes, and Maenhout (2005) report estimated values not significantly different from zero. Campbell, Cocco, Gomes, and Maenhout (2001) and Campbell and Viceira (2002) find values, ranging from 0.33 for households with no high-school education to 0.52 for college graduates. Cocco et al. (2005) provide estimates between 0.01 and 0.02, while Heaton and Lucas (1996) between 0.07 and 0.14. Munk and Sorensen (2010) report a correlation of 0.17.

an idiosyncratic and an aggregate component which may be correlated with market returns. While sticking to such mainstream approach, we adopt an estimation strategy based on the method of moments to make efficient use of information in PSID. This method exploits both the cross-sectional variation in labor income as well as its individual time variation available from the panel structure of the data. Thus, we are able to disentangle the idiosyncratic from the aggregate permanent income risks. We uncover that more than 90% of the variance of the permanent shock is idiosyncratic while the systematic component counts only to 7.5% of the variance of the overall permanent shock. This evidence may explain why the absolute value of the correlation between the systematic component of permanent shock and stock return innovation is relatively small in the aggregate. It also suggests to explore sample heterogeneity of hedging needs.

In the second step, we thus proceed with the reconstruction of labor income risk over the life cycle for each sample household. To this aim, we retrieve the dynamics over time of the unobservable component of labor income at the individual level with a Kalman filter. This exploits the assumed relationship between the observed and the unobserved variables in order to infer the dynamics of the latter. Recovering the individuals' dynamics of the permanent component allows the measurement of the empirical cross-sectional distribution of (otherwise) unobservable individual correlations between such permanent shocks and market returns. Our results confirm a substantial heterogeneity in the comovement of persistent earning shocks with stock returns.

Finally, we perform a probit analysis of stock market participation. The predicted level of aggregate participation, that is the average probability of participation across individuals, aligns with the actual level of stock market participation, defined as the percentage of individuals reporting positive equity holdings. Both are close to 40% over all the sample years. We also run the experiment of assigning to all individuals a correlation equal to either  $-0.2$  or  $0.2$ , respectively, instead of the observed ones. The predicted level of participation shifts from around 20% to around 60%, holding other parameters fixed.

A closely related study of hedging motives (Angerer and Lam, 2009) indicates that it is the variance of the permanent component of labor income shocks that mainly affects the share of risky assets (including both stocks and bonds) in household portfolios, without addressing stock market participation and correlations. Bonaparte et al. (2014) pin down the role of the correlation between labor income shocks and stock returns in explaining stock market participation. However, the estimate of correlation, based on their method, does not explain individual participation in PSID while our method does. Moreover, our method provides an estimate of the correlation that takes into account the theoretical restrictions on the different component of labor income shocks. In particular, we contribute to this literature an efficient estimate of individual correlation, showing that it is the permanent component of labor income shocks that affects hedging in financial markets, while the

correlation with total income shocks does not. These results support the theoretical implication that a sufficiently high and positive value of such correlation is able to explain the non-participation to the stock market in addition to the observed low equity share in participants' portfolios (Bagliano et al., 2014). We therefore do not have to rely on competing explanation of non participation, such as unawareness, participation costs, crash risk among others.

Recent studies document that individuals exhibit substantial heterogeneity in terms of labor income , and that such inequality has increased over time - see Guvenen (2009), among others. We add the observation that individuals are heterogeneous also in terms of correlation between income growth and stock market return. Moreover, we show a powerful implication of such heterogeneity. We point out that the estimate of the correlation coefficient over the whole sample yields a statistical zero value since individual correlations are widely scattered across negative and positive values. This argument supports and extends similar observations in Eiling (2013) and Fugazza, Giofrè, and Nicodano (2011).

Our study may contribute to improvements in the personalized design of target date funds. Often the time-series dimension of the data is limited, because of the short tenure of workers. The method we suggest exploits the cross-sectional dimension of the data to circumvent this problem.

The rest of the paper is organized as follows. Section 2 presents the labor income process and the estimation methodology, assuming that parameters of the income process are common to either sub-groups or to the entire sample. In section 3, we describe the data and the estimation of the labor income process. In section 4, we reconstruct the individual dynamics of the permanent stochastic trend, relating it to the stock market participation decision. Section 5 concludes.

## 2 Model specification and estimation method

### 2.1 Modeling labor income and participation

In this study, we model the log-labor income of the investor  $i$ , at time  $t$ , following Cocco et al. (2005):

$$\log(Y_{i,t}) = f(t, Z_{i,t}) + v_{i,t} + \epsilon_{i,t} \tag{1}$$

where  $f(t, Z_t)$  is a deterministic function of individual characteristics, such as age, race, marital status, and family size. The residuals are assumed to be the sum of two stochastic processes,  $v_{i,t}$ , and  $\epsilon_{i,t}$ , where

$$v_{i,t} = v_{i,t-1} + u_{i,t} \tag{2}$$

follows a random walk.  $u_{i,t} \sim \mathcal{N}(0, \sigma_u^2)$  and  $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2)$  are mutually uncorrelated. The term  $v_{i,t}$  represents the permanent shock hitting the labor income of individual  $i$  at time  $t$ , with innovation  $u_{i,t}$ . On the contrary, the shock to labor income,  $\epsilon_{i,t}$ , is transitory and uncorrelated across investors. The permanent shock is the sum of an aggregate and an idiosyncratic (i.e. individual specific) component:

$$u_{i,t} = \xi_t + \omega_{i,t} \tag{3}$$

where  $\xi_t \sim \mathcal{N}(0, \sigma_\xi^2)$ , and  $\omega_{i,t} \sim \mathcal{N}(0, \sigma_\omega^2)$ . The aggregate component  $\xi_t$  may be correlated with the excess market return. We denote their correlation coefficient with  $\rho_{\xi, r_m}$ , while  $\omega_{i,t}$  is orthogonal to  $r_{m,t}$ . Therefore, the variance of the permanent shock  $u_{i,t}$  is equal to:

$$\sigma_u^2 = \sigma_\xi^2 + \sigma_\omega^2 \tag{4}$$

and the correlation between permanent shock and excess stock return is computed as

$$\rho_{u, r_m} = \rho_{\xi, r_m} \frac{\sigma_\xi}{\sigma_u} \tag{5}$$

since the idiosyncratic component  $\omega_{i,t}$  is uncorrelated with the excess stock return.

As for participation, we adopt the reduced form proposed by Bonaparte et al. (2014). Participation is a binary variable, that depends on *individual* hedging needs. Aside from socio-economic characteristics, these depend on the level of *individual* log-labor income, on its correlation with the excess stock return and on the standard deviation of its innovations. Importantly, it is the permanent shock that should matter according to the theory (see Haliassos and Michaelides, 2003).

The following section explains how to exploit both the time-series and the cross-sectional dimension of the data in order to separately identify the moments of the idiosyncratic and aggregate components of the permanent income shock. We postpone to section 5 the details concerning the reconstruction of the permanent income shocks at the individual level.

### 2.1.1 Theoretical (Co-)moment of the labor income process

We derive the total labor income shock (TLIS) for individual  $i$  at time  $t$ ,  $e_{i,t}$ , from equation (1):

$$e_{i,t} = \log(Y_{i,t}) - f(t, Z_{i,t}) = v_{i,t} + \epsilon_{i,t} \tag{6}$$

Let us now compute the time variation in TLIS for each individual over a time interval of length  $d$ ,

DTLIS:

$$\Delta_d e_{i,t} = e_{i,t+d} - e_{i,t} \quad (7)$$

where  $d = \{1, 2, \dots, D\}$ , and  $D$  is the maximum number of lags.

A useful property of  $\Delta_d e_{i,t}$  is that it only contains the innovations of the permanent income shock and the transitory shock. Using equations (1)-(3), we obtain:

$$e_{i,t+d} - e_{i,t} = \sum_{s=t+1}^{t+d} u_{i,s} + \epsilon_{i,t+d} - \epsilon_{i,t}, \quad (8)$$

since the deterministic part of the random walk,  $v_{i,t}$ , cancels out. For instance, when  $d = 1$ ,  $\Delta_1 e_{i,t}$  is the first difference of the TLIS, and is equal to:

$$\Delta_1 e_{i,t} = u_{i,t+1} + \epsilon_{i,t+1} - \epsilon_{i,t} \quad (9)$$

Let  $N$  be the number of individuals and  $T$  the length of their working spell. Then, we construct two sets of variance-covariance matrices of DTLIS. The first set includes the  $[N \times N]$  matrix containing the variance of each individual's DTLIS time series on the main diagonal, and the covariance between individuals' DTLIS time series off the main diagonal.

For the second set, we construct  $D$  matrices, one for each lag, with dimension  $[(T-d) \times (T-d)]$ . Each matrix has on the main diagonal the cross-sectional variance, for each point in time. This is the variance of the  $N$ -dimensional vector containing the DTLIS of the  $N$  individuals at each time  $t$ . Off the main diagonal, there are the covariances between time periods, that are the covariances between the  $N$ -dimensional vectors containing the DTLIS of the  $N$  individuals at different time periods. Then, each  $[(T-d) \times (T-d)]$  matrix has the following symmetric form:

$$\begin{bmatrix} C_d(1,1) & C_d(2,1) & C_d(3,1) & \dots & C_d(T-d,1) \\ C_d(2,1) & C_d(2,2) & \cdot & \cdot & \cdot \\ C_d(3,1) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ C_d(T-d,1) & \cdot & \cdot & \cdot & C_d(T-d,T-d) \end{bmatrix} \quad (10)$$

The generic element of the matrix is denoted by  $C_d(t, t+l)$ , and is equal to

$$C_d(t, t+l) = \text{cov}(\Delta_d e_t, \Delta_d e_{t+l}), \quad (11)$$

where  $e_t$  is the  $N$ -dimensional vector containing the TLIS of the  $N$  individuals at each time  $t$ . Hence,

when  $l = 0$ ,  $C_d(t, t + l)$  is the cross-sectional variance, at time  $t$ , of the DTLIS corresponding to lag  $d$ :

$$C_d(t, t) = \text{var}(\Delta_d e_t) = d\sigma_\omega^2 + 2\sigma_\epsilon^2 \quad (12)$$

since the aggregate component is constant across individuals for each point in time. When  $l > 0$ ,  $C_d(t, t + l)$  identifies the covariance terms between time periods, that are equal to

$$\begin{bmatrix} C_d(t, t + l) = (d - l)\sigma_\omega^2 & d > l \\ C_d(t, t + l) = -\sigma_\epsilon^2 & d = l \\ C_d(t, t + l) = 0 & d < l \end{bmatrix} \quad (13)$$

Observe that we have isolated the variance of the individual-specific transitory income shock from the variance of the individual-specific permanent income shock, exploiting the temporal variation of DTLIS for each individual.

We now turn to the cross-sectional variation for each time  $t$ , that is to the characterization of the theoretical moment conditions for the  $[N - N]$  matrix. The  $[N - N]$  matrix has the following symmetric form:

$$\begin{bmatrix} C(1, 1) & C(2, 1) & C(3, 1) & \cdot & \cdot & C(N, 1) \\ C(2, 1) & C(2, 2) & \cdot & \cdot & \cdot & \cdot \\ C(3, 1) & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ C(N, 1) & \cdot & \cdot & \cdot & \cdot & C(N, N) \end{bmatrix} \quad (14)$$

The generic element of the matrix is denoted by  $C(i, j)$ . It is the covariance between the one-lag DTLIS time series of individuals  $\{i, j\}$  when  $i \neq j$ , and is the variance of the one-lag DTLIS time series of an individual when  $i = j$ :

$$\begin{bmatrix} C(i, j) = \sigma_\xi^2 & i \neq j \\ C(i, j) = \sigma_u^2 + 2\sigma_\epsilon^2 & i = j \end{bmatrix} \quad (15)$$

since the covariance across individuals is only due to the aggregate component. Observe that we have isolated the variance of the aggregate permanent shock, exploiting the variation of DTLIS across individuals, at each  $t$ .

Finally, we derive the moment condition for the covariance between each individual's one-lag DTLIS time series and the excess stock returns. This is equal to:

$$\text{cov}(\Delta_1 e_i, r_m) = \rho_{\xi, r_m} \sigma_\xi \sigma_{r_m} \quad (16)$$

where  $\sigma_{r_m}$  is the standard deviation of the excess stock return.

### 2.1.2 Empirical counterparts of labor income shocks

We must now identify the empirical counterparts of the total income shocks,  $e_{i,t}$  in equation (6). We estimate a panel regression of log-labor income on an age polynomial up to third order, and a set of observable socio-economic characteristics, such as race, marital status, and family size. Then, we obtain a time-series of log-labor income residuals for each individual by removing the deterministic component  $f(t, Z_{i,t})$ , where  $f(t, Z_{i,t})$  is assumed additively separable in  $t$  and  $Z_{i,t}$  includes age and individual fixed effects.

Let us now denote with  $N$  the number of individuals appearing in the sample, and with  $T$  the length of the sample time-series. We call *dres* the differentials of the regression residuals, that are the empirical counterparts of  $\Delta_d e_{i,t}$ . Then, we construct two sets of variance-covariance matrices of the *dres*.

Specifically, we construct the sample counterparts for the matrix  $[N - N]$ , by using the  $\Delta_1 e_{i,t}$ . For the set of matrices  $[(T - d) - (T - d)]$ , we use all the  $\Delta_d e_{i,t}$  up to  $d = D$ .

For each lag  $d$ , the elements  $C_d(t, t)$  are on the main diagonal of the  $[(T - d) - (T - d)]$  matrix, and the elements  $C_d(t, t + l)$  are on the  $l$ -th diagonal below the main one. We finally expand the matrix  $[N - N]$  with the empirical covariances between individual residuals and excess stock return, and with the variance of the excess stock return, thus forming a matrix  $[(N + 1) - (N + 1)]$ .

In total, from the  $[(T - d) - (T - d)]$  matrices, we obtain a number of sample conditions equal to

$$M_1 = \sum_{d=1}^D \sum_{l=0}^{\min(l, T-d)} (T - d - l) \quad (17)$$

and a number of sample conditions equal to  $M_2 = \frac{(N+1)(N+2)}{2}$  from the  $[(N + 1) - (N + 1)]$  matrix, as empirical counterparts for the non-zero theoretical moment conditions. Then, we stack the sample conditions, that we denote  $g_M$ , in one  $M$ -vector, with  $M = M_1 + M_2$ , and we perform numerically the following optimization to estimate the model parameters

$$\min_{\theta} (g_M - G_M(\theta))' I_M (g_M - G_M(\theta)) \quad (18)$$

where  $G_M(\theta)$  is the  $M$ -vector containing the non-zero theoretical moment conditions, and  $I_M$  is an identity matrix of size  $M$ .



### 2.1.3 Standard Errors

We obtain  $t$ -statistics through Monte Carlo simulations. We simulate the log-labor income residuals according to the equations (2)-(4) of the model, and using our estimates for the model parameters. In practice, we generate random transitory and permanent idiosyncratic shocks, for each individual  $i$ , and each point in time  $t$ , sampling from two univariate and independent normal distributions with variance equal to the estimates of  $\sigma_\epsilon^2$ , and  $\sigma_\omega^2$ , respectively. Then, we generate random stock returns and permanent aggregate shocks for each point in time  $t$ , sampling from a bivariate normal distribution, with zero mean vector, and covariance matrix equal to

$$\begin{bmatrix} \sigma_\xi^2 & \sigma_{r_m} \sigma_\xi \rho_{\xi, r_m} \\ \sigma_{r_m} \sigma_\xi \rho_{\xi, r_m} & \sigma_{r_m}^2 \end{bmatrix} \quad (19)$$

By using equation (3), we then reconstruct the permanent shock, and finally the log-labor income residual, for each individual  $i$ , and each point in time  $t$ . Each simulation produces a sample of log-labor income residuals, that forms the input for the estimation algorithm, thus yielding a set of parameter estimates. Repeating this procedure for a number  $S$  of simulations, we obtain a distribution of  $S$  estimates for each parameter, with corresponding mean and standard deviation.

## 3 The labor income process

### 3.1 Data

Information on individual labor income are collected from the Panel Study Income Dynamics (PSID), that provides survey data on personal, demographic, and income characteristics of US households. PSID data are available at annual frequency between 1968 and 1996, and every two years from 1997. We use the 26 waves covering the period from 1971 to 1997 in the first part of the paper, when we implement our estimation methodology of the labor income process, to allow comparability with previous studies (see e.g., Cocco et al., 2005, Guvenen, 2009).

We extend the time dimension of our sample in the second part of the paper, up to 2011, when we focus our attention on the stock market participation. Data on stock market participation, in fact, are available on PSID from 1999. The data convey information on both direct and indirect stock holdings. They refer to the household that owns any shares of stock in traded corporations including mutual funds, investment trusts and/or IRAs. As for investment opportunities in risky assets, finally, we use the US stock market excess return (from Kenneth French's website). Note that since PSID reports earnings data for the previous year, we use labor income data and stock market return for

the period 1970-1996.

Throughout the analysis, we apply the same selection criteria as Guvenen (2009), who also uses the PSID data. We consider an individual only if, for 20 out of 26 years, she (i) reports positive labor income and working hours, (ii) worked between 520 and 5110 hours in a year, (iii) reports hourly earnings included in a given interval, (iv) does not belong to the poverty SEO sub-sample in 1968, (v) is a male head of the household, and (vi) is between 20 and 64 years old. The selection procedure leaves a sample of 1107 individuals. The definition of labor income is the same as in Guvenen (2009), including wage income and additional earnings, such as bonuses, commissions, farm income and business income.

### 3.2 The Minimum Distance estimates of the labor income process

This section presents empirical results regarding the parameter vector  $\theta = \{\sigma_\epsilon, \sigma_\xi, \sigma_\omega, \rho_{\xi, r_m}\}$ . We first disentangle the aggregate and the idiosyncratic component of the permanent shock, a specificity of our Minimum Distance (MD) strategy. We then compare the MD estimates of the variance of the permanent income shock,  $u$ , and of its correlation with stock returns,  $corr(u, r)$ , with previous studies by Guvenen (2009) and Cocco et al. (2005). To ensure comparability, we perform the estimation on the same time interval, by using data up to 1996, that is the last year for which PSID reports annual data.

In Table 1, the first row lists the parameters. The second row reports the estimates based on the whole sample of 1107 individuals. The volatility of the aggregate component of labor income shocks,  $\sigma_\xi$  in column 1, is far smaller than the variance of the idiosyncratic component,  $\sigma_\omega$  in column 2. While it is known that aggregate income is less variable than individual income (see e.g. Pischke, 1995), here this result directly emerges out of individual data and is consistent across sub-samples in the other rows.

In column 6, the estimated correlation between the aggregate permanent shocks to labor income and stock returns is quantitatively small. This is not surprising, since the ratio of aggregate to total volatility of the permanent shock never exceeds 0.17. The estimated parameters are all statistically different from zero, but for the correlation coefficient between the aggregate permanent component of labor income shocks and stock returns,  $\rho_{\xi, r_m}$ . This implies that also the correlation between the overall permanent component and stock returns,  $corr(u, r)$ , is not significantly different from zero, a pattern that is common to benchmark estimates in other rows, as well.<sup>2</sup>

Such correlation, in column (7) is low and negative ( $-0.06$ ) at the aggregate level. It is positive for

---

<sup>2</sup>In Appendix A, we consider all the subsequent waves and show that the parameter estimates are robust to the extension of the time dimension of our sample

**Table 1 Labor Income Process: Estimated Parameters and Benchmarks**

The table reports the Minimum Distance estimates of the parameters  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , based on the PSID waves 1971-1996. We compute  $\sigma_u^2$ , and  $corr(u, r)$  using the equations in section 2. The market return,  $r_m$ , is the contemporaneous US excess stock return from Kenneth French’s website. Row (1) lists the parameters. Row (2)-(3) refer to the sample of 1107 individuals, (4)-(13) to groups based on educational attainment, (College/No-College), occupation (White collar/Blue collar), and sector of employment (Manufacturing/Non Manufacturing). Benchmark estimates are from Cocco et al. (2005) and Guvenen (2009).

	$\sigma_{\xi}^2$	$\sigma_{\omega}^2$	$\sigma_u^2$	$\sigma_{\epsilon}^2$	$\rho$	corr(u,r)	$\frac{\sigma_{\xi}}{\sigma_u}$	N
<b>All</b>	<b>0.001</b>	<b>0.015</b>	<b>0.017</b>	<b>0.121</b>	-0.225	-0.062	<b>0.075</b>	1107
<i>Guevenen (2007)</i>			<b>0.015</b>	<b>0.061</b>				
<b>College</b>	<b>0.003</b>	<b>0.020</b>	<b>0.023</b>	<b>0.128</b>	-0.270	-0.101	<b>0.140</b>	298
<i>Cocco et al. (2005)</i>			<b>0.017</b>	<b>0.006</b>		-0.018		
<i>Guevenen (2007)</i>			0.009	0.047				
<b>High School</b>	<b>0.002</b>	<b>0.014</b>	<b>0.015</b>	<b>0.120</b>	-0.130	-0.043	<b>0.108</b>	809
<i>Cocco et al. (2005)</i>			<b>0.011</b>	<b>0.074</b>		0.006		
<i>Guevenen (2007)</i>			<b>0.011</b>	<b>0.052</b>				
<b>White Collars</b>	<b>0.003</b>	<b>0.018</b>	<b>0.021</b>	<b>0.122</b>	-0.264	-0.099	<b>0.140</b>	351
<b>Blue Collars</b>	<b>0.003</b>	<b>0.013</b>	<b>0.015</b>	<b>0.121</b>	-0.066	-0.027	<b>0.174</b>	379
<b>Manufacturing</b>	<b>0.002</b>	<b>0.012</b>	<b>0.013</b>	<b>0.091</b>	-0.169	-0.060	<b>0.126</b>	218
<b>Non Manufacturing</b>	<b>0.001</b>	<b>0.017</b>	<b>0.018</b>	<b>0.113</b>	-0.354	-0.092	<b>0.068</b>	430

high school educated workers and negative for those with a college degree, indicating that the value at the aggregate level is mainly tracked by college graduates’ earnings.<sup>3</sup> Campbell and Viceira (2002) evaluate the correlation between permanent labor income shocks and lagged stock returns. They find high positive values in the range of 0.32-0.52. This result is common to Guevenen, Schulhofer-Wohl, Song, and Yogo (2017) who use administrative data to evaluate the regression coefficient between the growth rate in real earnings and lagged stock returns<sup>4</sup>.

Let us turn to the variance of the total permanent shock,  $\sigma_u^2$  and the variance of the transitory shock,  $\sigma_{\epsilon}^2$ . Both our estimates (0.017 and 0.121, respectively) and those in (Guvenen, 2009) in the third row (0.015 and 0.061) are based on the same PSID data and the same sample selection criteria. They reveal that the permanent shock is less volatile than the transitory one. This pattern finds confirmation in the estimates performed on subgroups, in the remaining rows. This observation applies both to our estimates, and to the benchmark ones by both Guvenen (2009) and Cocco et al. (2005). This analysis indicates common patterns across our estimates and previous ones, suggesting that we can rely upon our new estimates of the two components of permanent income shocks.

<sup>3</sup>At group level, Davis and Willen (2000) find a negative value for college educated and for high school dropouts, on US data, while Fagereng, Gottlieb, and Guiso (2017) find no significant correlation, on Norwegian data.

<sup>4</sup>Indeed, we do find a higher and positive value of 0.14 for the correlation when considering the relation with lagged stock returns.

At both aggregate and group level, estimates in Table 1 uncover considerable volatility in the idiosyncratic component of labor income. This suggests that we should observe a large dispersion in its co-movement with stock returns across individuals. In Appendix B.1, we report the distribution of the correlation between the first difference of labor income residuals and excess stock returns. Our results show considerable heterogeneity in such correlation across individuals. Thus, we repeat the estimates on clusters of individuals formed on the basis of their empirical correlation. For all the investigated sub-samples, the estimated correlation between stock returns and permanent income shocks is high in absolute value - conflicting with evidence in Table 1. In particular, we find that the estimated correlation between the permanent component and stock returns is significant and ranges from 0.53 to -0.48.<sup>5</sup> Such heterogeneity motivates us to reconstruct the income paths of our sample individuals. This will allow to obtain the individual correlations between the permanent component and stock returns.

## 4 Individual hedging motives and portfolio choice

So far, the estimation does not capture the heterogeneity in the correlation between permanent labor income shocks and stock returns across individuals. Hence, in this section, we perform an additional step that allows to retrieve the *individual's* income shocks as well as the correlation with stock market returns. We then show that the correlation between permanent labor income shocks and stock returns explains participation in the stock market, as predicted by the theory.

### 4.1 Individual dynamics of labor income shocks

We estimate the dynamics over time of the unobservable stochastic labor income components at the individual level through a linear Kalman filter. The Kalman filter enables to retrieve the dynamics of the latent variable, the permanent component of labor income shocks (equation (2)), by using the regression residuals as observable variable and the ex-ante known relationship between these two variables (equation (4)). To implement the Kalman filter, we use the estimates of the previous step as parameters of the model<sup>6</sup>.

As result, for each individual, we obtain the dynamics of the permanent shocks to labor income for each individual over the entire sample time series.

---

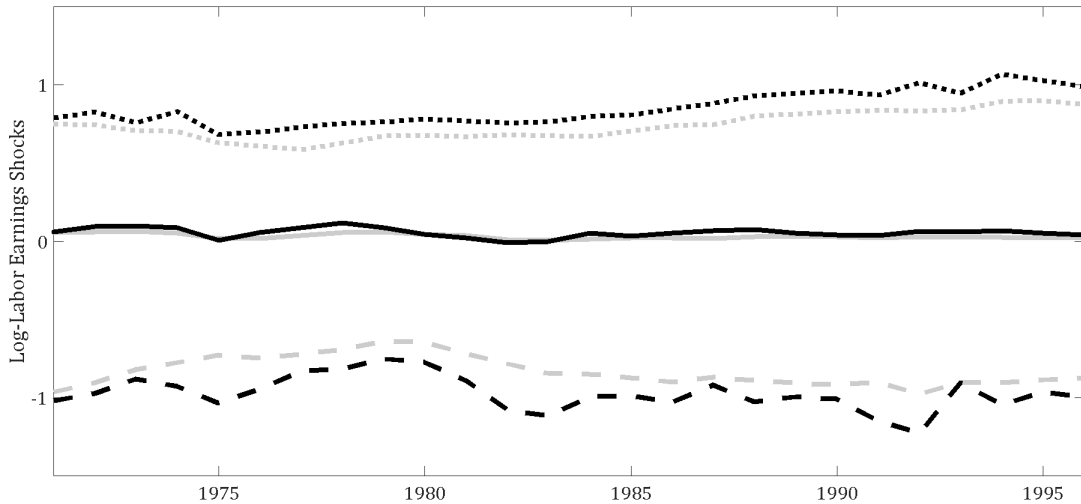
<sup>5</sup>In addition, in Appendix B.2 we obtain the parameter estimates for individual grouped according to their decision to participate and clustered according to the value of the empirical correlation with stock returns. Results reported in Appendix B.2 confirm the wide heterogeneity of the estimated correlation between the permanent component and stock returns.

<sup>6</sup>In Appendix C, we detail the implementation of the Kalman filter.

Figure 1 shows the dynamics of the individual permanent component of labor income shocks (grey lines) for the 5-th (dashed), the 50-th (solid), and the 95-th percentile (dotted), and compares them with the corresponding log-labor income residuals (black lines). Unsurprisingly, given that the permanent component ( $v_{i,t}$ ) is the sum of mean-zero shocks, the median value of the reconstructed  $v_{i,t}$  is zero.

**Figure 1. Individual Dynamics of Permanent Income Shocks**

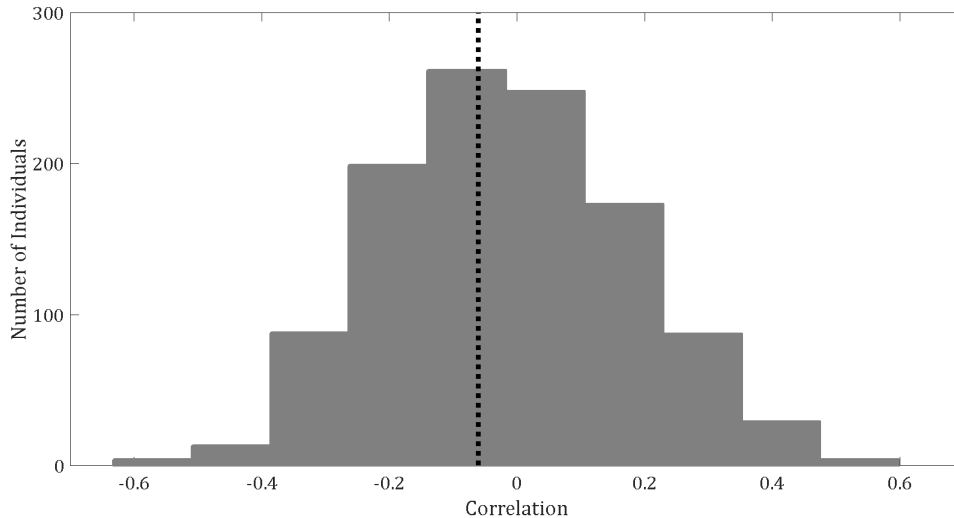
The figure shows the dynamics of the permanent component of the log-labor income residuals (grey lines) for the 5-th (dashed), the 50-th (solid), and the 95-th percentile (dotted), and the corresponding log-labor income residuals (black lines), for the time period going from 1971 to 1996. We estimate the individual’s permanent component of the log-labor income residuals by applying the Kalman filter methodology to the residuals obtained from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data, and the final sample of 1107 individuals.



By pinning down the variation over time of the permanent stochastic component of each individual, we are able to compute the correlation between each individual’s labor income innovation and stock market returns. In Figure 2, we report the distribution of such correlation coefficients evaluated across individuals. The distribution turns to be centered at -0.053 (emphasized with black dotted line), that is a value very similar to the estimated parameter for the aggregate correlation (see section 4.1). The standard deviation is 0.21.

## Figure 2. Individual Correlations

The figure shows the distribution of the correlation between the estimated individual's permanent innovations to the log-labor income and stock return, for the time period going from 1971 to 1996. We estimate the individual's permanent component of the log-labor income residuals by applying the Kalman filter methodology to the residuals obtained from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data, and the final sample of 1107 individuals.



The range of individual correlations is indeed very wide, between -0.6 and 0.6, with one third of individuals displaying a correlation not significantly different from zero. Thus, it is not surprising to obtain point estimates near zero or not significant. However, our results highlight that this low point estimate is due to high heterogeneity in individual labor earnings rather than to the absence of comovement with market returns.

## 4.2 Explaining stock market participation

In this section, we relate the *individual* decision to participate in the stock market directly to *individual* income hedging motives, along with observable personal characteristics. As described in section 3.1, we use waves between 1999 and 2011 to collect data on stock market participation.

Moreover, we use waves between 1988 and 2011 to estimate the individuals' labor income dynamics, and their correlation with stock returns. We adopt this time dimension to use a long enough time series of data for evaluating the individual's sample correlation, and at the same time for preserving a large enough cross-section of individuals to estimate the process parameters. In fact, further than applying the filters described in section 3.1, we also rule out individuals with less than 12 valid observations on labor income, labor income standard deviation larger than 5, and missing data

on demographic characteristics (age, family size, marital status and education). In this additional sample selection, we follow Bonaparte et al. (2014). The final sample, then, includes 944 individuals.

We use this sample to estimate a probit model for the effects of a set of explanatory variables on the decision to participate to the stock market. The dependent variable is the stock market participation, that is a dummy variable  $H$  that takes value 1 in case of participation, and 0 otherwise. The main independent variables of interest are the standard deviation and the correlation with stock market return of different specifications of shocks to (log)-labor income. As explained in Section 2, the total (log)-labor income  $\log(Y_{i,t})$  is the sum of a deterministic part  $f(t, Z_{i,t})$  and a stochastic part  $e_{i,t}$ , where  $f(t, Z_{i,t})$  is a function of individual's observable characteristics, and  $e_{i,t}$  is the stochastic component. This latter is the sum of a permanent component  $v_{i,t}$  and a transitory component  $\epsilon_{i,t}$  (equation (6)).

The probit regression is performed *quasi* out-of-sample: the estimation of the deterministic function  $f(t, Z_{i,t})$ , and of the income process parameters, are performed using all the waves between 1988 and 2011, while both standard deviation and correlation with stock market return are computed up to 1997, that is one wave before the first wave with stock market participation data. The idea is that individuals decide whether to invest in stock from time  $t$  onwards, given the information on standard deviation and correlation up to  $t$ .

In Table 2, we include the shocks to the total log-income  $\log(Y_{i,t})$ , and the permanent component of the stochastic shocks to the labor income computed with the Kalman filter, following the estimation methodology described in the previous section. In all the regressions, we control for time effects and individual's demographic characteristics, such as age, family size, marital status, education, and (log)-labor income level. We use standard maximum likelihood method to estimate the probit regression and we report results in Table 2.

Column (1) just considers observable characteristics, and all the characteristics explain the propensity to participate. In Column (2), we include only the empirical labor income characteristics: log-labor income level, the standard deviation of total income shocks, and their correlation with stock returns. According to our results, the level and the correlation are significant determinants of the participation decision, as in Bonaparte et al. (2014). However, once the individual characteristics are included, in Column (3), then the effect of income-return correlation is no longer significant (as in Angerer and Lam, 2009).

In Column (4), we consider the distribution of the individual permanent income shocks,  $v_{i,t}$ , obtained from the Kalman filter: the standard deviation of the permanent income shocks dynamics, and the correlation between the permanent income shocks dynamics and the stock market return. Both variables are statistically significant, and they are still significant when we include demographic

characteristics (Column (5)). Finally, results in Column (6) show that the individual’s permanent income shocks characteristics have significant explanatory power even when we add the total income shocks standard deviation and correlation with stock return.

**Table 2 Probit Regression Results**

The table reports the probit regression results considering as independent variables the participation decision. The dependent variable is equal to 1, if the individual reports non-zero wealth allocation to stock market on a given wave, and equal to 0 otherwise. The independent variables are individual’s demographic characteristics, and the standard deviation and the correlation with stock market return of the shocks to different specifications of (log)-labor income: total (log)-labor income, and permanent component of the stochastic part computed with the Kalman filter (KF) following the estimation methodology described in the previous section. The coefficients are estimated with standard maximum likelihood, and we report in bold those significant at 5% level. We include in all regression the time effects. The total number of observations is  $N(944) \times T(7) = 6608$

	(1)	(2)	(3)	(4)	(5)	(6)
Age	<b>0.009</b>		<b>0.009</b>		<b>0.008</b>	<b>0.008</b>
Family Size	<b>-0.042</b>		<b>-0.041</b>		<b>-0.043</b>	<b>-0.042</b>
Marital Status	<b>-0.140</b>		<b>-0.139</b>		<b>-0.139</b>	<b>-0.140</b>
Education	<b>0.165</b>		<b>0.165</b>		<b>0.166</b>	<b>0.166</b>
Labor Income	<b>0.191</b>	<b>0.309</b>	<b>0.191</b>	<b>0.308</b>	<b>0.191</b>	<b>0.191</b>
Specifications of (Log)-Labor Income Shocks						
<i>Total</i>						
Standard Deviation		-0.039	-0.018			0.049
Correlation		<b>-0.100</b>	-0.089			0.065
<i>Permanent (KF)</i>						
Standard Deviation				<b>-0.564</b>	<b>-0.348</b>	<b>-0.486</b>
Correlation				<b>-0.118</b>	<b>-0.152</b>	<b>-0.205</b>
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations = 6586						
Pseudo $R^2$	0.091	0.035	0.092	0.036	0.093	0.093

These results confirm the theoretical prediction that what is relevant for households’ risk taking decision is the ability of risky assets to hedge the permanent labor income risk instead of the pure transitory labor income risk<sup>7</sup>.

Previous results leave open the possibility that our Kalman Filter estimates do not outperform other empirical measures of permanent income shocks. We next compare the explanatory power of our method with the one in Bonaparte et al. (2014).

To this aim, we decompose the shocks to the total log-income, reported in Table 2, into its deter-

<sup>7</sup>For completeness, in Appendix D, we calibrate the standard life cycle model of consumption and portfolio choice (see Cocco et al., 2005) according to our estimates, and show that, ceteris paribus, what is relevant in determining the stock market participation is the correlation between the permanent component of labor income shock.



ministic part  $f(t, Z_{i,t})$  and stochastic part  $e_{i,t}$ . In Table 3, Columns (1) to (3) report the coefficients of the two parts included separately, and jointly, respectively.

**Table 3 Probit Regression - Kalman filtered permanent component vs moving average permanent component**

The table reports the probit regression results considering as independent variables the participation decision. The dependent variable is equal to 1, if the individual reports non-zero wealth allocation to stock market on a given wave, and equal to 0 otherwise. The independent variables are individual's demographic characteristics, and the standard deviation of shocks to (log)-labor income as well as their correlation with stock returns. The shocks to (log)-labor income are measured according to alternative specifications: the total (log)-labor income, the deterministic part, the stochastic part, the permanent and the transitory components of the stochastic part as computed by Bonaparte et al. (2014) (BKK), and the permanent and the transitory components of the stochastic part computed with the Kalman filter (KF) following the estimation methodology described in section 4.1 and in Appendix D. The coefficients are estimated with standard maximum likelihood, and we report in bold those significant at 5% level. We include in all regression the time effects. The total number of observations is  $N(944) \times T(7) = 6608$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Age	<b>0.009</b>	<b>0.008</b>	<b>0.01</b>	<b>0.007</b>	<b>0.009</b>	<b>0.008</b>	<b>0.008</b>	<b>0.009</b>	<b>0.009</b>	<b>0.007</b>	<b>0.009</b>	<b>0.009</b>	<b>0.011</b>
Family Size	<b>-0.042</b>	<b>-0.041</b>	<b>-0.041</b>	<b>-0.042</b>	<b>-0.041</b>	<b>-0.043</b>	<b>-0.043</b>	<b>-0.039</b>	<b>-0.039</b>	<b>-0.044</b>	<b>-0.044</b>	<b>-0.044</b>	<b>-0.042</b>
Marital Status	<b>-0.140</b>	<b>-0.139</b>	<b>-0.140</b>	<b>-0.138</b>	<b>-0.138</b>	<b>-0.137</b>	<b>-0.138</b>	<b>-0.136</b>	<b>-0.136</b>	<b>-0.139</b>	<b>-0.135</b>	<b>-0.135</b>	<b>-0.136</b>
Education	<b>0.165</b>	<b>0.165</b>	<b>0.165</b>	<b>0.166</b>	<b>0.165</b>	<b>0.168</b>	<b>0.166</b>	<b>0.166</b>	<b>0.166</b>	<b>0.165</b>	<b>0.165</b>	<b>0.167</b>	<b>0.165</b>
Labor Income	<b>0.191</b>	<b>0.191</b>	<b>0.189</b>	<b>0.190</b>	<b>0.188</b>	<b>0.177</b>	<b>0.184</b>	<b>0.189</b>	<b>0.189</b>	<b>0.190</b>	<b>0.189</b>	<b>0.170</b>	<b>0.166</b>
Specifications of (Log)-Labor Income Shocks													
<i>Total</i>													
Standard Deviation		-0.018											0.045
Correlation		-0.089											0.211
<i>Deterministic</i>													
Standard Deviation			0.043		0.098								0.070
Correlation			0.131		0.076								-0.069
<i>Stochastic</i>													
Standard Deviation				<b>-0.177</b>	<b>-0.239</b>								-1.974
Correlation				<b>-0.187</b>	<b>-0.175</b>								0.821
<i>Permanent (BKK)</i>													
Standard Deviation						<b>-0.329</b>		0.256				0.260	0.231
Correlation						<b>0.213</b>		<b>0.130</b>				<b>0.142</b>	0.124
<i>Transitory (BKK)</i>													
Standard Deviation							<b>-0.172</b>	<b>-0.236</b>				0.006	-0.014
Correlation							<b>-0.165</b>	<b>-0.116</b>				-0.234	<b>-0.293</b>
<i>Permanent (KF)</i>													
Standard Deviation									-0.094		<b>0.474</b>	0.446	1.946
Correlation									<b>-0.320</b>		<b>-0.418</b>	<b>-0.405</b>	<b>-0.912</b>
<i>Transitory (KF)</i>													
Standard Deviation											<b>-0.303</b>	<b>-0.523</b>	-0.596
Correlation											<b>-0.139</b>	0.071	0.366
Time Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations = 6586													
Pseudo $R^2$	0.092	0.092	0.092	0.094	0.094	0.094	0.095	0.096	0.094	0.094	0.096	0.098	0.100

Further, we include separately the permanent and transitory shocks in the set of explanatory variables. We consider both our disentangling strategy based on the Kalman filter, and the disentangling methodology of Bonaparte et al. (2014). In Bonaparte et al. (2014), the permanent component of the stochastic shocks to the labor income at time  $t$  is the equally weighted average of the stochastic shocks to the labor income at time  $t - 1$ ,  $t$ , and  $t + 1$ , and the transitory component of the stochastic shocks to the labor income at time  $t$  is the residual part.

Column (2) confirms that, once the individual characteristics are included, the effect of the standard deviation of *total* income and its correlation with stock return is not significant. Columns (3) to (5), however, suggest that only the deterministic part of the total income has no explanatory power, while both the standard deviation of the entire stochastic component and its correlation with stock return have a negative and significant effect on the propensity to invest in stocks.

In columns (6) to (8), we benchmark results obtained by Bonaparte et al. (2014) on two different datasets. In PSID, the correlation of the permanent component, computed according to Bonaparte et al. (2014), predicts a higher propensity to participate while it is the standard deviation and the correlation of the transitory component of the stochastic part that predict a lower propensity to participate. Finally, the correlation between the permanent component, obtained on the Kalman filter methodology, and stock return has always the expected impact on stock market participation. The coefficient of this correlation is always negative and significant, robust to the inclusion of all the characteristics of other shock specifications. We conclude that our method is able to explain households' participation behavior, despite the short time dimension of the data.

**Table 4 Probit Regression - Kalman filtered permanent component vs moving average permanent component with interactions**

The table reports the probit regression results considering as independent variables the participation decision. The dependent variable is equal to 1, if the individual reports non-zero wealth allocation to stock market on a given wave, and equal to 0 otherwise. The independent variables are individual’s demographic characteristics, and the interaction between the standard deviation of shocks to (log)-labor income and their correlation with stock returns. The shocks to (log)-labor income are measured according to alternative specifications: the total (log)-labor income, the deterministic part, the stochastic part, the permanent and the transitory components of the stochastic part as computed by Bonaparte et al. (2014) (BKK), and the permanent and the transitory components of the stochastic part computed with the Kalman filter (KF) following the estimation methodology described in section 4.1 and in Appendix D. The coefficients are estimated with standard maximum likelihood, and we report in bold the significant variables, at 5 % significance level. We include in all regression time effects. The total number of observations is  $N(944) \times T(7) = 6608$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	<b>0.009</b>	<b>0.009</b>	<b>0.009</b>	<b>0.009</b>	<b>0.009</b>	<b>0.009</b>	<b>0.009</b>
Family Size	<b>-0.040</b>	<b>-0.040</b>	<b>-0.037</b>	<b>-0.037</b>	<b>-0.036</b>	<b>-0.038</b>	<b>-0.037</b>
Marital Status	<b>-0.131</b>	<b>-0.132</b>	<b>-0.128</b>	<b>-0.128</b>	<b>-0.127</b>	<b>-0.128</b>	<b>-0.128</b>
Education	<b>0.165</b>	<b>0.164</b>	<b>0.164</b>	<b>0.166</b>	<b>0.165</b>	<b>0.165</b>	<b>0.164</b>
Labor Income	<b>0.189</b>	<b>0.186</b>	<b>0.186</b>	<b>0.176</b>	<b>0.182</b>	<b>0.187</b>	<b>0.186</b>
Specifications of (Log)-Labor Income Shocks							
<i>Total</i>	-0.014						
<i>Deterministic</i>		<b>0.027</b>					
<i>Stochastic</i>			<b>-0.062</b>				
<i>Permanent(BKK)</i>				<b>0.20</b>			
<i>Transitory(BKK)</i>					<b>-0.052</b>		
<i>Permanent (KF)</i>						<b>-0.186</b>	
<i>Transitory (KF)</i>							<b>-0.083</b>
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations = 6604							
Pseudo $R^2$	0.091	0.091	0.092	0.092	0.093	0.093	0.092

### 4.3 Predicting stock market participation

In this section, we use the probit regression estimates to predict the level of stock market participation, for each year. We adopt the standard normal distribution results at the base of the probit model. In particular, given the dependent dummy variable  $H$ , the probability that  $H$  is equal to 1 is computed as follows:

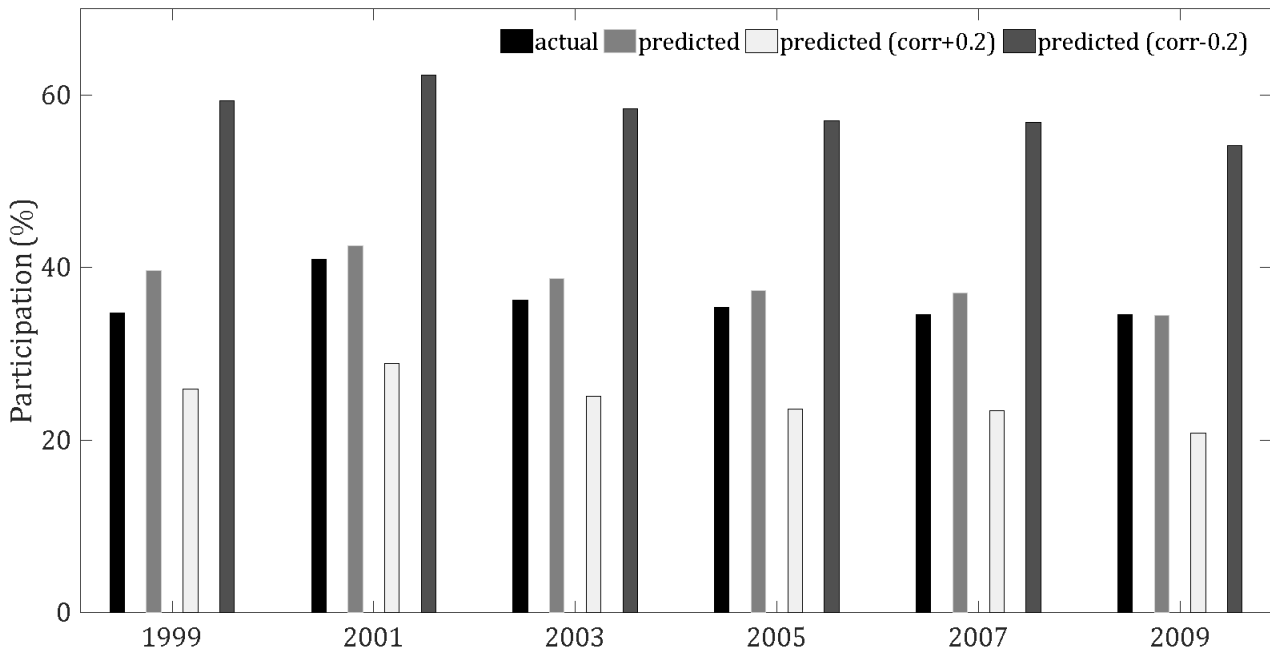
$$P(H = 1) = \Phi(\beta' * X) \tag{20}$$

where  $\Phi$  stands for the cumulative distribution function of a standard normal variable,  $\beta$  is the  $M \times 1$  vector of coefficients estimated with the probit regression, and  $X$  is the  $N \times M$  matrix of  $M$  explanatory variables, where  $N$  is the number of individuals in the sample. We compute  $P(H = 1)$  for each year. As a result, for each individual in the sample, we obtain the predicted probability that he participates to the stock market in each year covered by our data. Then, for each year, we

compute the predicted level of aggregate participation as the average probability across individuals, and we compare it with the actual level of stock market participation, defined as the percentage of individuals in the final sample reporting value of dummy variable  $H$  equal to 1. Figure 3 shows that probit regression estimates allow to match quite well actual data on stock market participation. Both predicted and actual level of participation lie around 40% over all the years in the sample.

**Figure 3. Actual and Predicted Stock Market Participation**

The figure shows the actual stock market participation, and the stock market participation predicted by using the estimates of the Probit regression of Table 3, for each year. Stock market participation is defined as the percentage of individuals in the final sample with a dummy variable  $H$  equal to 1. We also show the predicted stock market participation imputing a correlation between the estimated individual’s permanent component of the log-labor income shocks and stock return equal to -0.2 and 0.2, respectively, to all individuals in the sample. The predicted level of stock market participation is computed by using standard normal distribution results:  $P(H = 1) = \Phi(\beta' * X)$ , where  $\Phi$  stands for the cumulative distribution function of a standard normal variable,  $\beta$  is the  $M \times 1$  vector of estimated coefficients, and  $X$  is the  $N \times M$  matrix of  $M$  explanatory variables.



Furthermore, to stress the impact of the correlation between the stock market returns and permanent income shocks retrieved using the Kalman filter, we obtain the participation probability by assigning to all individuals in our sample correlation equal to either -0.2 or 0.2 respectively, and we calculate again equation (20). Results reported in figure 3 show that the individual correlation of the permanent component substantially matters in determining the stock market participation: varying this correlation from -0.2 to 0.2, the level of participation goes from around 20% to around 60%.

## 5 Summary and Conclusions

Understanding idiosyncratic risks faced by investors is a central issue in the economics of incomplete markets. This source of *ex-post* heterogeneity may have strong implications on households' portfolio choice and hence on the behavior of asset prices. To the extent that uninsurable labor income risks are highly correlated with risky assets, more prudent investors should reduce their exposure to financial risks (Merton, 1969). Conversely, if labor income exhibits a low correlation with financial markets, then investing in risky assets can serve as a hedge against labor income risk.

This study contributes to the measurement of the joint distribution of individual labor income shocks and stock returns over the life cycle. We first derive efficient estimates of the parameters of individual labor income processes using public panel data. In particular, the applied method offers more flexibility than the maximum likelihood methodology, allowing to disentangle the idiosyncratic and the aggregate permanent income risk as well as the correlation between the latter and market returns. In addition, we recover the individual dynamics of labor income profiles and obtain the overall distribution of correlations between permanent income shocks and stocks returns. The dispersion of this correlation across individuals leads us to conclude that the lack of correlation estimated at the aggregate level is due to the wide range of significant (positive and negative) correlations observed at individual level rather than to the absence of comovement of earnings shocks with market returns.

These achievements allow to relate the decision to participate in the stock market to income hedging motives, alongside personal characteristics. Our findings show that both the variability of shocks to income that have permanent effects and their correlation with stock returns are significant in determining the propensity to participate while the correlation with total income shocks is not. Our results confirm the theoretical prediction that what is relevant for households' risk taking decision is the ability of risky assets to hedge permanent labor income risk rather than pure transitory shocks.

## References

- Abowd, J., and D. Card, 1989, On the covariance structure of earnings and hours changes, *Econometrica* 57, 411–445.
- Angerer, X., and P.S. Lam, 2009, Income risk and portfolio choice: an empirical study, *Journal of Finance* 64, 1037–1055.
- Bagliano, F.C., C. Fugazza, and G. Nicodano, 2014, Optimal life-cycle portfolios for heterogeneous workers, *Review of Finance* 18, 2283–2323.
- Bonaparte, J., G. Korniotis, and A. Kumar, 2014, Income hedging and portfolio decisions, *Journal of Financial Economics* 113, 300–324.
- Campbell, J. Y., J. Cocco, F. Gomes, and P. Maenhout, 2001, Investing retirement wealth: a life-cycle model, *Risk Aspects of Investment-Based Social Security Reform* University of Chicago Press, 439–483.
- Campbell, J.Y., and L. Viceira, 2002, Strategic asset allocation: Portfolio choice for long-term investors, *Oxford University Press, Oxford, UK* .
- Cocco, J., F. Gomes, and P. Maenhout, 2005, Consumption and portfolio choice over the life cycle, *Review of Financial Studies* 18, 491–533.
- Davis, S. J., and P. Willen, 2000, Occupation-level income shocks and asset returns: Their covariance and implications for portfolio choice, *Unpublished working paper, GSB University of Chicago* .
- Eiling, E., 2013, Industry-specific human capital, idiosyncratic risk and the cross-section of expected stock returns, *The Journal of Finance* 68(1), 43–84.
- Fagereng, A., C. Gottlieb, and L. Guiso, 2017, Asset market participation and portfolio choice over the life cycle, *Journal of Finance* 72(2), 705–750.
- Fugazza, C., M. Giorfrè, and G. Nicodano, 2011, International diversification and industry-related labor income risk, *International Review of Economics and Finance* 20(4), 764–783.

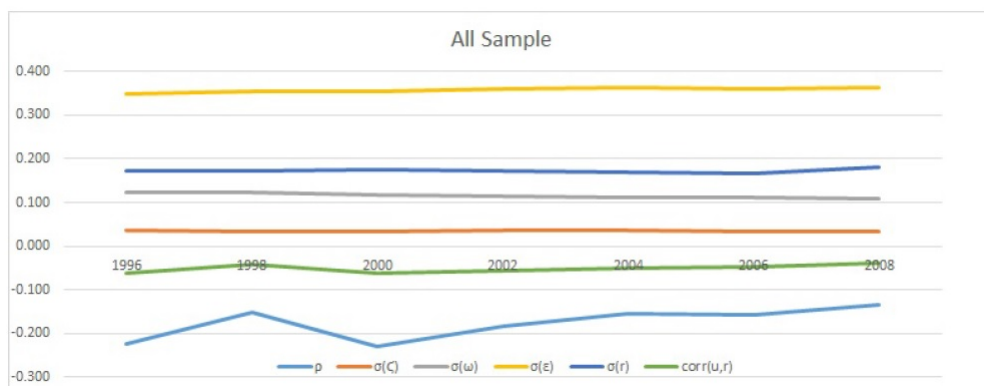
- Guevenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo, 2017, Worker betas: five facts about systematic earnings risk, *American Economic Review* 107, 398–403.
- Guvenen, F., 2009, Empirical investigation of labor income process, *Review of Economic Dynamics* 12.
- Haliassos, M., and A. Michaelides, 2003, Portfolio choice and liquidity constraints, *International Economic Review* 44, 144–177.
- Heaton, J., and D. Lucas, 1996, Evaluating the effects of incomplete markets on risk sharing and asset pricing, *Journal of Political Economy* 104, 443–487.
- Heaton, J., and D. Lucas, 2000, Portfolio choice and asset prices: The importance of entrepreneurial risk, *Journal of Finance* 55, 1163–1198.
- Merton, R., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, *Review of Economics and Statistics* 51, 247–257.
- Munk, C., and C. Sorensen, 2010, Dynamic asset allocation with stochastic income and interest rates, *Journal of Financial Economics* 96, 433–462.
- Pischke, J. S., 1995, Individual income, incomplete information, and aggregate consumption, *Econometrica* 63, 805–840.

## Appendix A Parameter estimates over subsequent waves

In this appendix, we analyze the stability of the parameter estimates reported in Table 1 (section 3.2) over time. In Figure A.1, we plot the value of parameters estimated considering all the individuals and expanding the sample by adding sequentially the subsequent waves (from 1998 to 2008). The figure shows that our estimates are robust across waves.

### Figure A.1. Stability of parameter estimates

The figure shows Minimum Distance (MD) estimates of the parameters, when the final year of the panel data varies from 1996 to 2008. The  $y$ -axis indicates the point estimate of the corresponding parameter, and the  $x$ -axis indicates the final year of the time dimension of the data-sample used for the estimation. The number of individuals included in the sample changes according to the different estimation periods.



## Appendix B Empirical Distributions

In Appendix B.1, we measure the distribution of the empirical correlation between the first difference of labor income residuals and excess stock returns and re-estimate all the parameters of Table 1 on clusters of individuals formed on the basis of this correlation. Our results show considerable heterogeneity in such correlation across individuals. In Appendix B.2, we repeat the point estimates of Table 1 conditioning on the decision to (not) participate in the stock market.

For all the investigated sub-samples, the estimated correlation between stock returns and permanent income shocks is high in absolute value - conflicting with evidence in Table 1. In particular, we find that the estimated correlation between the permanent component and stock returns is significant and ranges from 0.53 to -0.48. Such heterogeneity motivates us to reconstruct the income paths of our sample individuals. This will allow to obtain the individual correlations between the permanent component and stock returns.

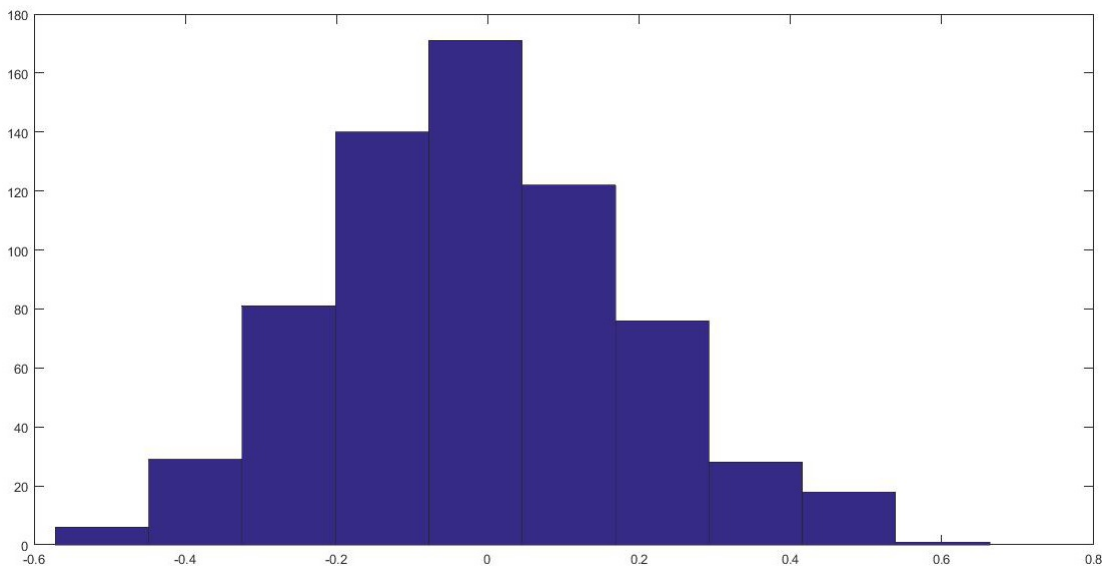


## B.1 Estimates by clusters of individuals

Figure B.1 reports the distribution of the empirical correlation between the first difference of labor income residuals and stock returns.

### Figure B.1. Empirical correlation between the first difference of labor income residuals and stock returns

The figure shows the empirical distribution of the observed correlation between individual's log-labor income residuals and stock return, for the time period going from 1971 to 1996. The individual's time-series correlation is computed using annual log-labor income residuals, and annual stock returns. We obtain residuals from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data, using the final sample of 1107 individuals.



The figure shows that the individual correlations are widely dispersed across negative and positive values. Thus it is not surprising that the estimate of the single correlation coefficient for all agents yields a statistically close-to-zero value (see Table 1 in the main text).

Consequently, we proceed by re-estimating all the parameters by clusters of individuals. To this aim, we take all the samples considered in section 3.2 of the main text, and split them into sub-samples according to the 30th and the 60th percentiles of distribution of the empirical correlation (plotted in Figure B.1). In particular, for each sample we consider, on the one hand, the subsample of those individuals for whom the correlation is lower than  $-0.1$  and, on the other hand, the sub-sample of those individuals with a correlation higher than  $0.1$ . In particular, we select the clusters on all sample of 1107 individuals and on the subgroups of workers formed on the basis of education and occupational characteristics. For each cluster, we estimate the labor income process parameters for all the pairs of subgroups by implementing the MD estimation procedure described in section 2 of the

main text. We obtain residuals from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data. Then, we estimate equation (5) by using the log-income residuals as observable variable. We obtain point estimates for  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , and then we compute  $\sigma_u^2$ , and  $corr(u, r)$  according to equations described in section 2. In Table B.1, we report results.

**Table B.1 Clusters and Groups**

The table reports the estimates for the model parameters by implementing the estimation methodology described in section 2 on clusters of individuals of our final sample, selected on the base of the observed empirical correlation between log-labor income residuals and stock return. The cluster of individuals are formed by grouping those with correlation lower than -0.1, and those with correlation higher than 0.1. We select the clusters on all sample of 1107 individuals, and on the subgroups considered in section 3. The market variable is the contemporaneous US excess stock return as reported by the Kenneth French's website.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Non college		College		Blue		White		Second		Third		All	
	>0.1	<-0.1	>0.1	<-0.1	>0.1	<-0.1	>0.1	<-0.1	>0.1	<-0.1	>0.1	<-0.1	>0.1	<-0.1
$\rho$	<b>0.97</b>	<b>-0.92</b>	<b>0.92</b>	<b>-0.93</b>	<b>0.96</b>	<b>-0.93</b>	<b>0.94</b>	<b>-0.93</b>	<b>0.93</b>	<b>-0.94</b>	<b>0.92</b>	<b>-0.87</b>	<b>0.95</b>	<b>-0.93</b>
$\sigma_{\xi}$	<b>0.09</b>	<b>0.08</b>	<b>0.11</b>	<b>0.09</b>	<b>0.08</b>	<b>0.07</b>	<b>0.09</b>	<b>0.08</b>	<b>0.07</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>	<b>0.10</b>	<b>0.09</b>
$\sigma_{\omega}$	<b>0.13</b>	<b>0.13</b>	<b>0.16</b>	<b>0.17</b>	<b>0.12</b>	<b>0.12</b>	<b>0.16</b>	<b>0.16</b>	<b>0.15</b>	<b>0.15</b>	<b>0.14</b>	<b>0.14</b>	<b>0.14</b>	<b>0.15</b>
$\sigma_{\epsilon}$	<b>0.26</b>	<b>0.27</b>	<b>0.27</b>	<b>0.25</b>	<b>0.22</b>	<b>0.21</b>	<b>0.24</b>	<b>0.23</b>	<b>0.21</b>	<b>0.22</b>	<b>0.25</b>	<b>0.23</b>	<b>0.28</b>	<b>0.26</b>
$\sigma_r$	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>	<b>0.17</b>
N	249	264	97	114	89	102	122	129	79	77	130	131	370	397
$\sigma_{\xi}^2$	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.00</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>
$\sigma_{\omega}^2$	<b>0.02</b>	<b>0.02</b>	<b>0.03</b>	<b>0.03</b>	<b>0.01</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>
$\sigma_u^2$	<b>0.02</b>	<b>0.02</b>	<b>0.04</b>	<b>0.04</b>	<b>0.02</b>	<b>0.02</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>
$\sigma_{\epsilon}^2$	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	<b>0.06</b>	<b>0.05</b>	<b>0.05</b>	<b>0.06</b>	<b>0.05</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.05</b>	<b>0.08</b>	<b>0.07</b>
$corr(u, r)$	<b>0.55</b>	<b>-0.51</b>	<b>0.50</b>	<b>-0.43</b>	<b>0.54</b>	<b>-0.45</b>	<b>0.45</b>	<b>-0.43</b>	<b>0.41</b>	<b>-0.41</b>	<b>0.48</b>	<b>-0.46</b>	<b>0.53</b>	<b>-0.48</b>
$stdevcorr(u, r)$	0.0594	0.0637	0.0713	0.0680	0.0708	0.0679	0.0644	0.0616	0.0674	0.0664	0.0642	0.0711	0.0604	0.0614

According to our results, for all sub-samples the estimated correlation between stock returns and permanent income shocks is high in absolute value. In particular, when focusing on the overall population (column 7 of Table B.1), for those individuals who exhibit an empirical correlation higher than 0.1 the estimated correlation between stock returns and permanent shocks is significant and about 0.53, while for those whose empirical correlation is lower than -0.1 the estimated correlation is significant and about -0.48. Similar values are obtained when considering sub-samples over individuals grouped by education, occupation and industry. Consequently, our findings imply that two otherwise equal individuals may display opposite stock market exposure to the stock market risk. For example, two workers in the financial industry have opposite exposures if one works in private equity and the other in restructuring; two workers in oil have different exposures if one imports from Saudi and the other is in fracking. Our results confirm the high level of heterogeneity across individuals in terms of correlation between permanent labor income shocks and stock returns.

## B.2 Individual correlation and stock market participation

In this section, we estimate the correlation between permanent income shocks and stock returns for participants and non participants to the stock market. We create a dummy variable for participation ( $H = 1$ ) and we divide the individuals in two groups based on  $H$ . Moreover, since the decision to participate may be affected by other observable characteristics, we match individuals on age, occupation, education, industry categories. In practice, each individual with  $H = 1$  is paired with an individual who has  $H = 0$ , and the same observable characteristics. We perform the estimation on the two similar groups, in the sense that they are equal unless that the first group participates, and the second group does not. For each group we perform the method of moment estimation of all parameters considering the time span from 1971 to 2008 and to 2010, alternatively. We report results in Table B.2. In the table,  $T_h$  stands for the number of waves for which the individual participates<sup>8</sup>. Our results show that the correlation between permanent income shocks and stocks returns is negative and not significant for those who participate to the stock market while it is positive and significant for those who do not participate.

---

<sup>8</sup>In particular,  $T_h > 1$  indicates that workers declared to hold stocks (direct and/or indirect) at least once across all waves.  $T_h > 2$  indicates that workers declared to hold stocks (direct and/or indirect) at least twice all waves.  $T_h > 3$  indicates that workers declared to hold stocks (direct and/or indirect) more than three times across all waves.

**Table B.2 Estimates on Matched Groups - Only Equal Pairs**

The table reports the estimated parameters resulting from the minimization in equation (5), using the log-income residuals as observable variable. We obtain residuals from the panel regression of log-labor income on an age polynomial and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data. We first estimate  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ . Then we compute  $\sigma_u^2$ , and  $corr(u, r)$  by using the model equations described in section 2. % corr >0 is the percentage of workers with positive correlation.  $H = 1$  identifies the subgroup of individuals that report non-zero allocation of wealth to stock investments for more than a given number of waves, and  $H = 0$  identifies the subgroup of individuals that report non-zero allocation of wealth to stock investments for less than, or equal to, a given number of waves. We apply our estimation methodology on subgroups of individuals of our final sample based on the number of surveys-waves for which an individual reports non-zero allocation of wealth to stock investments. For each threshold number of waves, we match two subgroups based on observable characteristics by using propensity score estimator, selecting only pairs of individuals with bounded differences in annual wage. Information on investments allocation are available in PSID from the wave 1999. The market return is the contemporaneous US excess stock return from Kenneth French's website.

	$T_h > 0$		$T_h > 1$		$T_h > 2$		$T_h > 3$	
	H=1	H=0	H=1	H=0	H=1	H=0	H=1	H=0
	2008							
$\rho$	<b>-0.361</b>	<b>0.298</b>	<b>-0.466</b>	<b>0.362</b>	<b>-0.138</b>	0.330	0.968	<b>0.408</b>
$\sigma_{\xi}$	<b>0.028</b>	<b>0.084</b>	<b>0.022</b>	<b>0.086</b>	<b>0.022</b>	<b>0.099</b>	<b>0.002</b>	<b>0.077</b>
$\sigma_{\omega}$	<b>0.124</b>	<b>0.098</b>	<b>0.130</b>	<b>0.100</b>	<b>0.139</b>	<b>0.093</b>	<b>0.147</b>	<b>0.097</b>
$\sigma_{\epsilon}$	<b>0.373</b>	<b>0.305</b>	<b>0.388</b>	<b>0.312</b>	<b>0.395</b>	<b>0.312</b>	<b>0.407</b>	<b>0.309</b>
$\sigma_r$	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.182</b>	<b>0.181</b>
$corr(u, r)$	<b>-0.080</b>	<b>0.194</b>	<b>-0.078</b>	<b>0.236</b>	<b>-0.022</b>	<b>0.241</b>	<b>0.013</b>	<b>0.254</b>
$st.dev(corr(u, r))$	0.382	0.200	0.544	0.193	0.545	0.194	7.719	0.223
$\frac{\sigma_{\xi}^2}{\sigma_u^2}$	<b>0.049</b>	<b>0.424</b>	<b>0.028</b>	<b>0.425</b>	<b>0.024</b>	<b>0.531</b>	<b>0.000</b>	<b>0.387</b>
% corr >0	39%	52%	39%	55%	43%	54%	46%	52%
N	132	132	105	105	78	78	64	64
	2010							
$\rho$	<b>-0.293</b>	<b>0.258</b>	<b>-0.377</b>	<b>0.335</b>	<b>-0.147</b>	<b>0.385</b>	<b>-0.197</b>	<b>0.385</b>
$\sigma_{\xi}$	<b>0.039</b>	<b>0.081</b>	<b>0.035</b>	<b>0.085</b>	<b>0.037</b>	<b>0.099</b>	<b>0.044</b>	<b>0.107</b>
$\sigma_{\omega}$	<b>0.121</b>	<b>0.096</b>	<b>0.126</b>	<b>0.097</b>	<b>0.130</b>	<b>0.094</b>	<b>0.138</b>	<b>0.095</b>
$\sigma_{\epsilon}$	<b>0.386</b>	<b>0.337</b>	<b>0.393</b>	<b>0.344</b>	<b>0.405</b>	<b>0.353</b>	<b>0.407</b>	<b>0.363</b>
$\sigma_r$	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>
$corr(u, r)$	<b>-0.090</b>	<b>0.166</b>	<b>-0.101</b>	<b>0.221</b>	<b>-0.040</b>	<b>0.279</b>	<b>-0.060</b>	<b>0.288</b>
$st.dev(corr(u, r))$	0.100	0.132	0.097	0.131	0.104	0.140	0.113	0.145
$\frac{\sigma_{\xi}^2}{\sigma_u^2}$	<b>0.094</b>	<b>0.416</b>	<b>0.072</b>	<b>0.434</b>	<b>0.075</b>	<b>0.526</b>	<b>0.092</b>	<b>0.559</b>
% corr >0	38%	44%	38%	47%	41%	49%	42%	50%
N	136	136	112	112	89	89	71	71

Finally, since the labor income is a strong determinant of investment decisions, we include (log)-Labor income within the matching variable. In particular, for each threshold, in terms of number of waves, we match two subgroups on the base of observable characteristics by using propensity score estimator, and ignoring differences in terms of annual wage. Results reported in Table B.3 confirm that the correlation between permanent income shocks and stocks returns is negative and

not significant for those who participate to the stock market while it is positive and significant for those who do not participate. Results are reported in Table B.4.

**Table B.3 Estimates on Matched Groups - All Pairs**

The table reports the results on the model parameters estimation by implementing the estimation methodology described in section 4. We estimate the equation (5) by using the log-income residuals as observable variable. We obtain residuals from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data. We estimate  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , and then we compute  $\sigma_u^2$ , and  $corr(u, r)$  by using the model equations described in section 2. We apply our estimation methodology and then report results on subgroups of individuals of our final sample, based on the number of surveys-waves for which an individual reports non-zero allocation of wealth to investments in stock market.  $H = 1$  identifies the subgroup of individuals that report non-zero allocation of wealth to investments in stock market for more than a given number of waves, and  $H = 0$  identifies the subgroup of individuals that report non-zero allocation of wealth to investments in stock market for less than, or equal to, a given number of waves. Information on investments allocation are available in PSID from the wave 1999. The market variable is the contemporaneous US excess stock return as reported by the Fama-French website.

	$T_h > 0$		$T_h > 1$		$T_h > 2$		$T_h > 3$	
	H=1	H=0	H=1	H=0	H=1	H=0	H=1	H=0
	2008							
$\rho$	-0.226	<b>0.399</b>	-0.262	<b>0.479</b>	-0.240	<b>0.442</b>	-0.113	<b>0.333</b>
$\sigma_{\xi}$	<b>0.050</b>	<b>0.049</b>	<b>0.055</b>	<b>0.073</b>	<b>0.048</b>	<b>0.091</b>	<b>0.051</b>	<b>0.084</b>
$\sigma_{\omega}$	<b>0.111</b>	<b>0.093</b>	<b>0.117</b>	<b>0.096</b>	<b>0.121</b>	<b>0.096</b>	<b>0.126</b>	<b>0.087</b>
$\sigma_{\epsilon}$	<b>0.340</b>	<b>0.333</b>	<b>0.359</b>	<b>0.299</b>	<b>0.358</b>	<b>0.320</b>	<b>0.359</b>	<b>0.299</b>
$\sigma_r$	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>	<b>0.181</b>
$corr(u, r)$	-0.093	<b>0.186</b>	-0.111	<b>0.290</b>	-0.088	<b>0.304</b>	-0.042	<b>0.231</b>
$\frac{\sigma_{\xi}^2}{\sigma_u^2}$	0.169	0.217	0.181	0.366	0.136	0.473	0.141	0.482
$st.dev(corr(u,r))$	0.092	0.101	0.103	0.118	0.099	0.127	0.096	0.137
% corr > 0	38%	47%	37%	63%	42%	53%	45%	45%
N	189	189	154	154	118	118	91	91
	2010							
$\rho$	-0.182	<b>0.465</b>	-0.214	<b>0.425</b>	-0.137	<b>0.388</b>	-0.109	-0.007
$\sigma_{\xi}$	<b>0.058</b>	<b>0.082</b>	<b>0.059</b>	<b>0.059</b>	<b>0.061</b>	<b>0.050</b>	<b>0.064</b>	<b>0.056</b>
$\sigma_{\omega}$	<b>0.115</b>	<b>0.097</b>	<b>0.120</b>	<b>0.095</b>	<b>0.119</b>	<b>0.101</b>	<b>0.126</b>	<b>0.094</b>
$\sigma_{\epsilon}$	<b>0.362</b>	<b>0.349</b>	<b>0.371</b>	<b>0.330</b>	<b>0.365</b>	<b>0.317</b>	<b>0.371</b>	<b>0.326</b>
$\sigma_r$	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>	<b>0.180</b>
$corr(u, r)$	-0.082	<b>0.300</b>	-0.094	<b>0.224</b>	-0.062	0.172	-0.049	-0.004
$\frac{\sigma_{\xi}^2}{\sigma_u^2}$	<b>0.203</b>	<b>0.417</b>	<b>0.195</b>	<b>0.278</b>	<b>0.208</b>	<b>0.197</b>	<b>0.205</b>	<b>0.262</b>
$st.dev(corr(u,r))$	0.094	0.114	0.096	0.111	0.110	0.100	0.116	0.104
% corr > 0	38%	66%	37%	47%	41%	44%	44%	50%
N	225	225	185	185	145	145	112	112

## Appendix C Kalman filter

In this appendix, we detail how to implement the Kalman filter to retrieve the dynamics of the unobserved components of labor income shocks. We consider the state-space model described by the equations (2)-(4), and we reconstruct the dynamics of the unobservable random walk  $v_{i,t}$  for

each individual, by implementing a linear Kalman filter, using the regression residuals as observable variable, and the parameters estimated in the optimization (5). In few words, the Kalman filter exploits the assumed relationship between the observed and the unobserved variables in order to infer the dynamics of the latter. This relationship forms the *measurement equation*, and here is given by the equation (6) in the text. The equation (2) in the text, instead, forms the *transition equation*, which describes the evolution over time of the latent variable. We initialize the filter with two arbitrary conditions on the initial value of the latent variable, and its variance:

$$v_{i,0} \quad P_{i,0} \tag{A.1}$$

Then, we use the *prediction equations* to estimate the one-step ahead value of the latent variable, and the variance:

$$\begin{aligned} E_0[v_{i,1}] &= v_{i,0} \\ E_0[P_{i,1}] &= P_{i,0} + Q \end{aligned} \tag{A.2}$$

where the first prediction comes from the transition equation, and  $Q$  is the variance of the transition equation, here given by  $\sigma_u^2$ . Then, we use the measurement equation for making a forecast about the observed variable, and we compare our forecast with the actual observation, thus obtaining a measurement error<sup>9</sup>.

$$h_{i,1} = e_{i,1} - E_0[e_1] = e_{i,1} - v_{i,0} \tag{A.3}$$

We take into account the measurement error for updating the estimate of  $v_{i,1}$ , and  $P_{i,1}$ :

$$\begin{aligned} cc\hat{v}_{i,1} &= v_{i,0} + K * h_1 \\ \hat{P}_{i,1} &= (1 - K) * E_0[P_{i,1}] \end{aligned} \tag{A.4}$$

where  $K$  is the key *Kalman gain*, which weighs the measurement error in the estimate update, and it is equal to

$$K = \frac{E_0[P_{i,1}]}{E_0[P_{i,1}] + \sigma_\epsilon^2} \tag{A.5}$$

In fact,  $\sigma_\epsilon^2$  plays here the role of the measurement error variance, that is the reliability of the new available observation in improving the estimate of the latent variable. These steps are repeated recursively over the entire time series, thus obtaining an estimate of the latent variable dynamics. As final result, we reconstruct the dynamics of the permanent stochastic component of the log-income process, for each individual, over the sample time series.

---

<sup>9</sup>Based on information at  $t = 0$ , the expected value of  $u_{i,t}$  and  $\epsilon_{i,t}$  is zero.

## Appendix D Optimal life-cycle stock market participation and asset allocation

In this appendix, we consider the standard life-cycle model of consumption and portfolio choice (see e.g., Cocco et al., 2005) to show that what is relevant in shaping optimal stock market participation and portfolio allocation is the correlation between permanent income shocks and stock returns. First, we outline the theoretical details of the model. Next, we solve it under the standard benchmark calibration for all parameters but those governing the labor income process. For these latter, we consider the estimates on the PSID obtained according to the MD methodology of section 3.

### D.1 The life-cycle model

In particular, we model an investor who maximizes the expected discounted utility of consumption over her entire life and wishes to leave a bequest as well. The effective length of her life, which lasts at most  $T$  periods, is governed by age-dependent life expectancy. At each date  $t$ , the survival probability of being alive at date  $t + 1$  is  $p_t$ , the conditional survival probability at  $t$ . The investor starts working at age  $t_0$  and retires with certainty at age  $t_0 + K$ . Investor's  $i$  preferences at date  $t$  are described by a time-separable power utility function:

$$\frac{C_{it_0}^{1-\gamma}}{1-\gamma} + E_{t_0} \left[ \sum_{j=1}^T \beta^j \left( \prod_{k=0}^{j-1} p_{t_0+k} \right) \left( p_{t_0+j} \frac{C_{it_0+j}^{1-\gamma}}{1-\gamma} + (1-p_{t_0+j}) b \frac{(X_{it_0+j}/b)^{1-\gamma}}{1-\gamma} \right) \right] \quad (\text{A.6})$$

where  $C_{it}$  is the level of consumption at time  $t$ ,  $X_{it}$  is the amount of wealth the investor leaves as a bequest to her heirs in case of death,  $b \geq 0$  is a parameter capturing the strength of the bequest motive,  $\beta < 1$  is a utility discount factor, and  $\gamma$  is the constant relative risk aversion parameter.

#### D.1.1 Labor and retirement income

Available resources to finance consumption over the agent's life cycle derive from accumulated financial wealth and from the stream of labor income. At each date  $t$  during *working life*, the exogenous labor income  $Y_{it}$  is assumed to be governed by a deterministic age-dependent growth process  $f(t, \mathbf{Z}_{it})$ , and is hit by both a permanent shock  $u_{it}$  and a transitory disturbance  $\epsilon_{it}$ , according to process outlined in section 1.

During *retirement*, income is certain and equal to a fixed proportion  $\lambda$  of the permanent component

of income in the last working year:

$$\log Y_{it} = \log \lambda + f(t_{0+K}, \mathbf{Z}_{it_{0+K}}) + u_{it_{0+K}} \quad t_0 + K < t \leq T \quad (\text{A.7})$$

where the level of the replacement rate  $\lambda$  is meant to capture at least some of the features of Social Security systems.

### D.1.2 Investment opportunities

We allow savings to be invested in a short-term riskless asset, yielding each period a constant gross real return  $R^f$ , and in stocks, characterized as “*stocks*”. The risky asset yields stochastic gross real returns denoted as  $R_t^s$ . We maintain that the investment opportunities in the risky assets do not vary over time and model excess returns of stocks over the riskless asset as

$$R_t^s - R^f = \mu^s + \nu_t^s \quad (\text{A.8})$$

where  $\mu^s$  is the expected stock premium, and  $\nu_t^s$  is the normally distributed innovation, with mean zero and variance  $\sigma_s^2$ .

At the beginning of each period, financial resources available for consumption and saving are given by the sum of accumulated financial wealth  $W_{it}$  plus current labor income  $Y_{it}$ , that we call *cash on hand*  $X_{it} = W_{it} + Y_{it}$ . Given the chosen level of current consumption,  $C_{it}$ , next period cash on hand is given by:

$$X_{it+1} = (X_{it} - C_{it})R_{it}^P + Y_{it+1} \quad (\text{A.9})$$

where  $R_{it}^P$  is the portfolio return

$$R_{it}^P = \alpha_{it}^s R_t^s + (1 - \alpha_{it}^s) R^f \quad (\text{A.10})$$

with  $\alpha_{it}^s$ , and  $(1 - \alpha_{it}^s)$  denoting the shares of the investor’s portfolio invested in stocks and in the riskless asset respectively.

### D.1.3 Solving the life-cycle problem

In this standard intertemporal optimization framework, the investor maximizes the expected discounted utility over life time, by choosing the consumption and the portfolio rules given uncertain



labor income and asset returns. Formally, the optimization problem is written as:

$$\max_{\{C_{it}\}_{t_0}^{T-1}, \{\alpha_{it}^s, \alpha_{it}^b\}_{t_0}^{T-1}} \left( \frac{C_{it_0}^{1-\gamma}}{1-\gamma} + E_{t_0} \left[ \sum_{j=1}^T \beta^j \left( \prod_{k=0}^{j-2} p_{t_0+k} \right) \left( p_{t_0+j} \frac{C_{it_0+j}^{1-\gamma}}{1-\gamma} + (1-p_{t_0+j}) b \frac{(X_{it_0+j}/b)^{1-\gamma}}{1-\gamma} \right) \right] \right) \quad (\text{A.11})$$

$$s.t. \quad X_{it+1} = (X_{it} - C_{it}) \left( \alpha_{it}^s R_t^s + (1 - \alpha_{it}^s) R^f \right) + Y_{it+1}$$

with the labor income and retirement processes specified above and short sales and borrowing constraints imposed.

Given its intertemporal nature, the problem is restated in a recursive form, rewriting the value of the optimization problem at the beginning of period  $t$  as a function of the maximized current utility and of the value of the problem at  $t + 1$  (Bellman equation):

$$V_{it}(X_{it}, u_{it}) = \max_{\{C_{it}\}_{t_0}^{T-1}, \{\alpha_{it}^s, \alpha_{it}^b\}_{t_0}^{T-1}} \left( \frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta E_t [p_t V_{it+1}(X_{it+1}, u_{it+1}) + (1-p_t) b \frac{(X_{it+1}/b)^{1-\gamma}}{1-\gamma}] \right) \quad (\text{A.12})$$

At each time  $t$  the value function  $V_{it}$  describes the maximized value of the problem as a function of the two state variables, the level of cash on hand at the beginning of time  $t$ ,  $X_{it}$ , and the level of the stochastic permanent component of income at beginning of  $t$ ,  $u_{it}$ . In order to reduce the dimensionality of the original problem to one state variable we exploit the homogeneity of degree  $(1-\gamma)$  of the utility function, and normalize the entire problem by the permanent component of income  $u_{it}$ . Thus, we can rewrite (A.12) as

$$V_{it}(X_{it}) = \max_{\{C_{it}\}_{t_0}^T, \{\alpha_{it}^s\}_{t_0}^T} \left( \frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta E_t [p_t V_{it+1}(X_{it+1}) + (1-p_t) b \frac{(X_{it+1}/b)^{1-\gamma}}{1-\gamma}] \right) \quad (\text{A.13})$$

This problem has no closed form solution: hence the optimal values for consumption and portfolio shares at each point in time are obtained by means of numerical techniques. To this aim, we apply a backward induction procedure and obtain optimal consumption and portfolio rules in terms of the state variable starting from the last possible period of life  $T$ . In particular, in the presence of

bequest, the terminal condition is:

$$V_{iT+1}(X_{iT+1}) = b \frac{(X_{iT+1}/b)^{1-\gamma}}{1-\gamma}$$

and the Bellman equation (A.12) at  $T$  becomes

$$V_{iT}(X_{iT}) = \frac{C_{iT}^{1-\gamma}}{1-\gamma} + \beta E_t \left( b \frac{(X_{iT+1}/b)^{1-\gamma}}{1-\gamma} \right) \quad (\text{A.14})$$

from which the optimal consumption and portfolio share policy rules are obtained for each possible value of the state variable (the initial level of cash on hand at  $T$ ) using the standard grid search method.<sup>10</sup> Going backwards, for every period  $t = T - 1, T - 2, \dots, t_0$ , the Bellman equation (A.12) is used to obtain the optimal rules for consumption and the portfolio shares. For each level of the state variable  $X_{it}$ , the value function at the beginning of time  $t$ ,  $V_{it}(X_{it})$ , is obtained by picking the levels of consumption and portfolio shares that maximize the sum of the utility from current consumption  $U(C_{it})$  and the discounted expected value from continuation  $\beta E_t[\cdot]$ , computed using  $V_{it+1}(X_{it+1})$  obtained from the previous iteration. In particular, given  $V_{it+1}(X_{it+1})$ , the expectation term is evaluated in two steps. We use numerical integration performed by means of the standard Gaussian Hermite quadrature method to approximate the distribution of shocks to labor income and asset returns. Then, cubic spline interpolation is employed to evaluate the value function at points that do not lie on the state space grid.

## D.2 Optimal asset allocation decisions on calibrated parameters

In this section, we report the results from solving the standard life cycle model for consumption and portfolio decisions considering the values of the relevant parameters reported in Table D.1 (Cocco et al., 2005).

---

<sup>10</sup>According to this method, the problem is solved over a grid of values covering the space of the state variables and the controls, to ensure that the solution found is a global optimum.

**Table D.1 Calibration parameters**

Description	Parameter	Value
Working life (max)	$T$	20 -65
Retirement (max)	$t_0 + K$	65 -100
Discount factor	$\beta$	0.96
Risk aversion	$\gamma$	5
Replacement ratio	$\lambda$	0.68
Riskless rate	$r$	0.02
Excess returns on stocks	$\mu^s$	0.04

In addition, we calibrate the relevant parameters of the labor income process against the estimates obtained by taking alternatively 2 and 4 clusters of individuals formed on the basis of the empirical correlation of the total income residuals and stock returns displayed in Appendix B.1. In particular, we obtain residuals from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data. We rank individuals on the basis of the empirical correlation between labor income residuals and stock returns and group them into two clusters that cover workers with negative correlation and those with positive correlation, respectively. Then we estimate the model equations (18) to obtain  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , and then we compute  $\sigma_u^2$ , and  $corr(u, r)$ . We report results in Table D.2. In the second case, workers are grouped into 4 clusters on the basis of the four quartiles of the distribution of the empirical correlation between labor income residuals and stocks returns. On each group of workers, we estimate the model equations (18) of section 2 to obtain  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , and then we compute  $\sigma_u^2$ , and  $corr(u, r)$ . Results are reported in Table D.3.

**Table D.2 Estimates by 2 clusters**

The table reports the results on the model parameters estimation by implementing the estimation methodology described in section 3. We estimate the equation (18) of the main text by using the log-income residuals as observable variable. We obtain residuals from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data. We obtain the point estimates for  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , and then we compute  $\sigma_u^2$ , and  $corr(u, r)$  by using the model equations described in section 2. We apply our estimation methodology and then report results on clusters of individuals selected on the base of the observed empirical correlation between log-labor income residuals and stock return. In particular, we report in the table results for two clusters workers, i.e. those with negative correlation and those with positive correlation, respectively. We select the clusters on all sample of 1107 individuals. The market variable is the contemporaneous US excess stock return as reported by the Kenneth French's website.

Clusters	-0.71-0	0-0.66
Parameters		
$\sigma_u$	<b>0.150</b>	<b>0.163</b>
$\sigma_{\epsilon}$	<b>0.296</b>	<b>0.317</b>
$\sigma_r$	<b>0.411</b>	<b>0.411</b>
$corr(u, r)$	<b>-0.195</b>	<b>0.214</b>

**Table D.3 Estimates by 4 clusters**

The table reports the results on the model parameters estimation by implementing the estimation methodology described in section 3. We estimate the equation (18) in the main text by using the log-income residuals as observable variable. We obtain residuals from the panel regression of log-labor income on an age polynomial, and demographic characteristics, by using the waves 1971-1996 of the PSID as source of data. We estimate  $\rho_{\xi, r_m}, \sigma_{\xi}, \sigma_{\omega}, \sigma_{\epsilon}, \sigma_{r_m}$ , and then we compute  $\sigma_u^2$ , and  $corr(u, r)$  by using the model equations described in section 2. We apply our estimation methodology and then report results on clusters of individuals, selected on the base of the observed empirical correlation between log-labor income residuals and stock return. In particular, we report in the table results for four clusters, corresponding to the four quartiles of the empirical distribution of the individual correlations. We select the clusters on all sample of 1107 individuals. The market variable is the contemporaneous US excess stock return as reported by the Kenneth French's website.

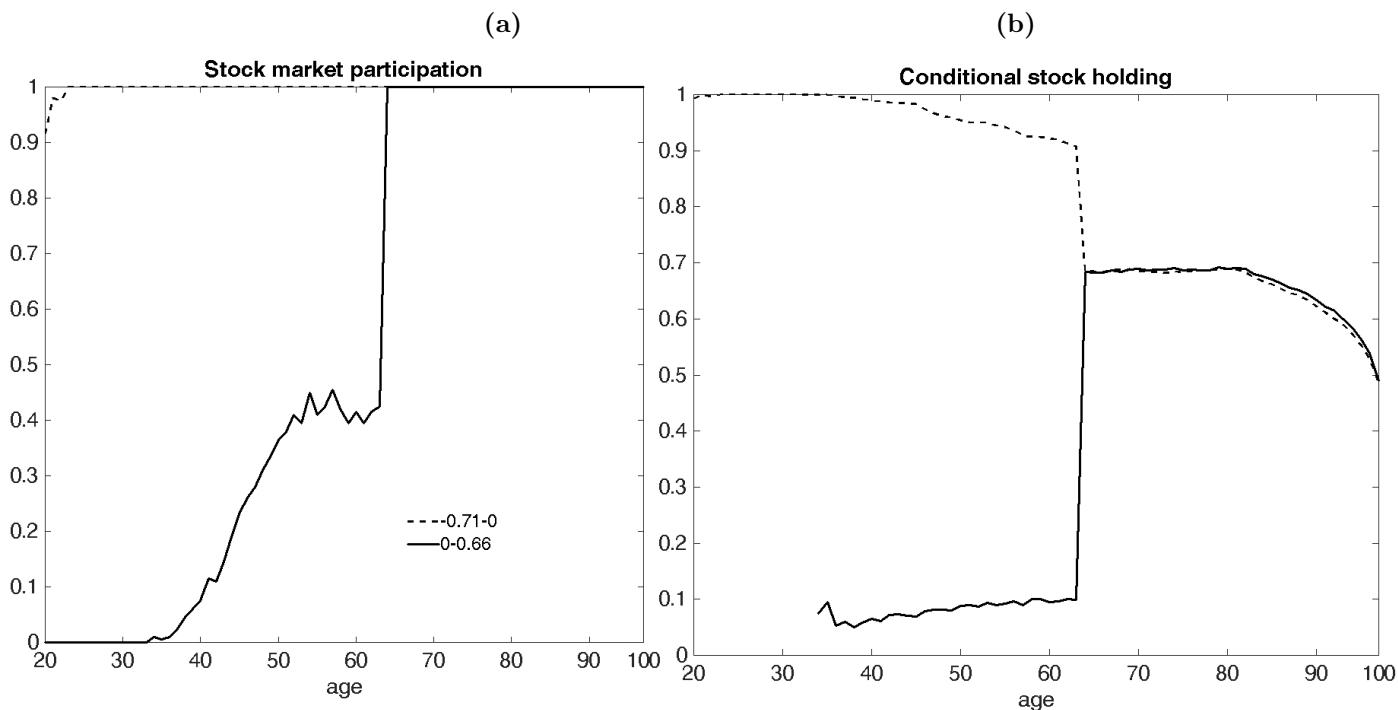
Clusters	-0.71-0.17	-0.17-0	0-0.18	0.18-0.66
Parameters				
$\sigma_u$	<b>0.172</b>	<b>0.145</b>	<b>0.141</b>	<b>0.199</b>
$\sigma_{\epsilon}$	<b>0.260</b>	<b>0.325</b>	<b>0.324</b>	<b>0.308</b>
$\sigma_r$	<b>0.411</b>	<b>0.411</b>	<b>0.411</b>	<b>0.411</b>
$corr(u, r)$	<b>-0.267</b>	<b>-0.087</b>	<b>0.104</b>	<b>0.278</b>

Our results evidence that the estimated correlation between the permanent component of labor income shocks and stock returns,  $corr(u, r)$ , is significant for all the clusters of workers. In particular,  $corr(u, r) = -0.195$ , for workers whose labor income shocks are negatively correlated with stock returns,  $corr(u, r) = 0.214$ , for workers whose labor income shocks are positively correlated with stock returns ( $corr(u, r) = -0.267$  for the bottom quartile and  $corr(u, r) = 0.214$  for the top quartile). We solve the life-cycle model for each set of parameters and report results of the optimal life cycle profiles

of stock market participation and conditional stock holdings in figures D.1 (for the case of 2 clusters) and D.2 (for the case of 4 clusters). Our results show that, workers displaying a negative correlation between the permanent component of labor income and stock returns should always participate in the stock market (dashed line in panel (a) of figure D.1, and dashed and the dotted lines in panel (a) of figure D.2) investing almost all their financial wealth in stocks during their working life (dashed line in panel (b) of figure D.1, and dashed and the dotted lines in panel (b) of figure D.2). For these workers, labor income provides a hedge against shocks to stock returns, thus it is optimal to tilt their portfolios to stocks. On the contrary, for those workers displaying a positive and sufficiently high correlation between the permanent component of labor income and stock returns it is optimal to not participate in the stock market when young (solid line in panel (a) of figure D.1, and the solid and the dashed-dotted lines in panel (a) of figure D.2) and invest a relatively low fraction of their financial wealth in the stock market later in life (solid line in panel (b) of figure D.1, and the solid and the dashed-dotted lines in panel (b) of figure D.2). With positive correlation between permanent labor income shocks and stock returns, it is optimal to offset such high background risk by holding a considerably lower fraction of financial portfolio in stocks and not to participate at all to the stock market at young ages, when human capital represent the bulk of total wealth.

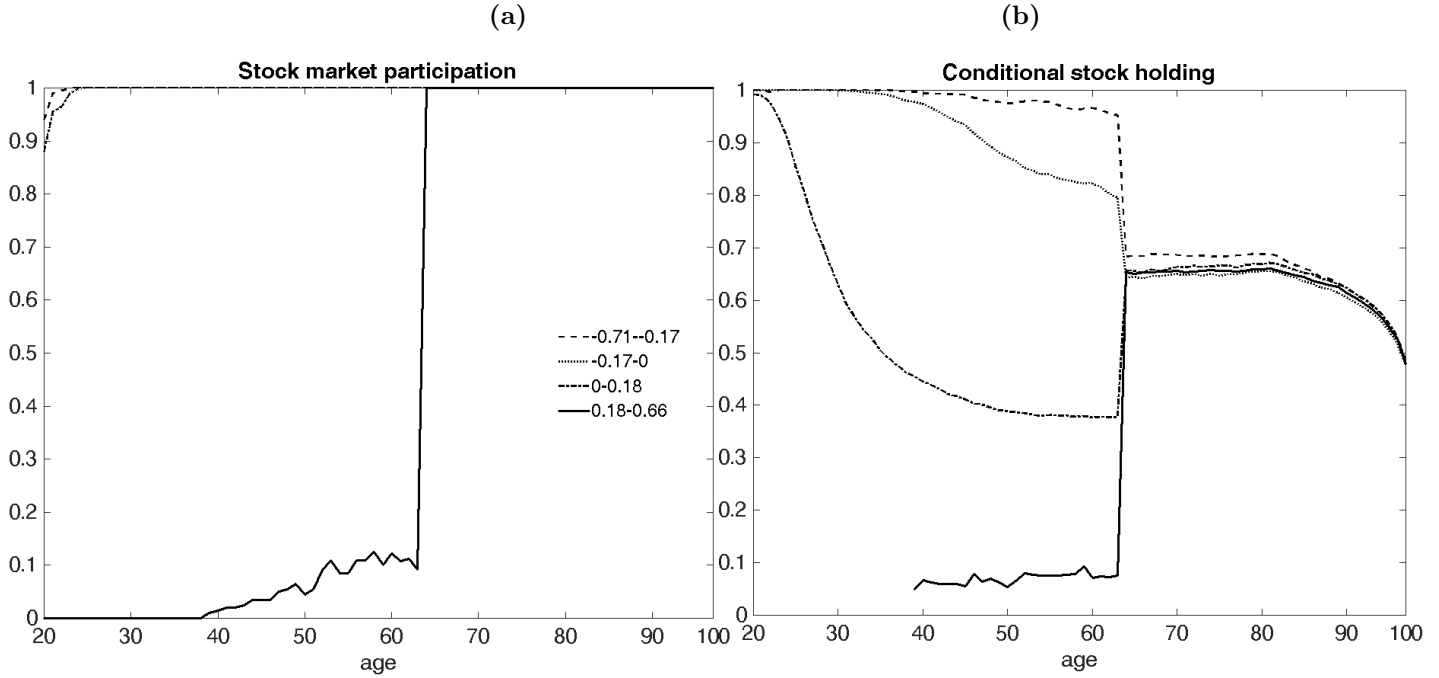
**Figure D.1. Optimal conditional stock holding and stock market participation - two clusters**

The figures displays the profiles of stock market participation and conditional stock holding for two clusters of individuals. Labor income parameters take the values estimated in Table D.2. Preferences and financial asset returns are calibrated according to values in Table D.1.



### Figure D.2. Optimal conditional stock holding - four clusters

The figures displays the profiles of stock market participation and conditional stock holding for four clusters of individuals. Labor income parameters take the values estimated in Table D.3. Preferences and financial asset returns are calibrated according to values in Table D.1.



### D.3 Optimal asset allocation decisions: transitory vs permanent income shock correlation

In this section, we consider a counterfactual experiment to proof that what is relevant in determining the stock market participation is the positive correlation between the permanent component of labor income shocks and stock returns (see Haliassos and Michaelides, 2003 and Bagliano et al., 2014). To this aim, we solve the standard life cycle according to the calibrated parameters of Table D.1 and assuming, counterfactually, that labor income is hit by transitory shocks only and that they are correlated to stock returns. In particular, we set the variance of the transitory shock such that the total labor income risk is equal to our estimate of section 3. In addition, we consider two counterfactual extreme values for its correlation with stock returns (see Table D.4).

**Table D.4 Counterfactual correlated transitory shock**

Correlation transitory shock and stock returns	Var. Permanent shock	Variance Transitory shock	Dev. Stand. stock returns
-0.214	-	0.110	0.1688
0.237	-	0.127	0.1688

Our results show that, the value of the correlation between transitory shock and stocks do not affect substantially the optimal stock market participation and stock holdings (see figure D.4).

**Figure D.3. Optimal conditional stock holding and stock market participation - correlated transitory shock**

