Hedging Permanent Income Shocks.*

Fabio C. Bagliano[1], Raffaele Corvino[1], Carolina Fugazza[1] and Giovanna Nicodano[3]

[1]University of Torino and CeRP (Collegio Carlo Alberto)

[3]University of Torino, CEPR, CeRP (Collegio Carlo Alberto) and Netspar

This draft: February 12th, 2022

Abstract

Do hedging motives predict portfolio choice and stock market participation? While prior evidence is mixed, this paper shows that they do both over time and in an out-of-sample analysis. Conceptually, we highlight that non-zero correlations between individual permanent income shocks and stock market returns imply co-movements across individuals' income growth. Empirically, we exploit the longitudinal dimension of the data to deliver precise estimates of such correlations. Risk taking is highly sensitive to individual correlations that capture heterogeneous exposure to aggregate risk.

Keywords: Permanent income shocks, Portfolio Choice, Stock market participation, Incomplete markets, Life-cycle portfolio design

JEL classification: G10,G11, D14, C15

Address: Collegio Carlo Alberto, Piazza Arbarello 8, 10122, Torino (Italy).

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

*We are grateful to Alex Michaelides, Marie Brière, Bernard Dumas, Davide Fiaschi, Christian Gollier, George Korniotis, Roberto Marfé, Ignacio Monzon, Henriette Prast, Theo Nijman, Alberto Quaini and an anonymous referee, as well as participants in the 2021 CeRP Workshop, the 2020 Consob ESMA Bocconi webinar, the 2021 Inquire Europe webinar, the 2019 Meetings of the European Economic Association, the International Association of Applied Econometrics, the Italian Labor Economics Association, the 2019 Paris Finance Meeting (Eurofidai), the OEE Conference 2020, and seminars at Collegio Carlo Alberto, Imperial College London and the University of Pisa for helpful suggestions. We are also grateful to the Institute for Quantitative Investment Research Europe, the Observatoire de l'Epargne Européenne, and the Italian MIUR ("Department of Excellence" 2018–2022) for funding.

1 Introduction

According to life-cycle theory, portfolio investments help hedge individual permanent labor income (PI) shocks which are otherwise uninsured. The theory encourages investors to reduce (Merton, 1969; Viceira, 2001) and possibly avoid (Benzoni, Collin-Dufresne, and Goldstein, 2007; Bagliano, Fugazza, and Nicodano, 2014) equity holdings when stock market returns display positive correlation with their PI shocks, implying that equities amplify earnings risk. Moreover, hedging choices change over time because earnings and return realizations affect them through both consequent budget constraints and learning about earnings ability (as in Chang, Hong, and Karabarbounis, 2018). However, the evidence on income hedging motives [1] as drivers of observed portfolios is less clear-cut. It is indeed well known that both stock market participation and portfolio choice are usually moderately sensitive to the estimated correlations between PI shocks and stock market returns. Moreover, the statistical significance of estimated correlations between realized labor income growth and realized stock market returns may suffer from the data's relatively short time-series dimension. Finally, the average correlation coefficients are often estimated to be close to zero, so that hedging motives are unlikely to account for individuals' moderate risk taking in equity markets.

This paper uses a novel approach to estimate the individual correlation between PI shocks and stock market returns. We demonstrate that these new estimates imply large hedging motives and can explain portfolio choice. Moreover, they keep stronger explanatory power than competing methods. We also update these correlation estimates based on new information. Consistent with the learning hypothesis, we find that these updates explain individual equity portfolio shares and participation over time. These results confirm the theoretical prediction that portfolio holdings of equities respond to their ability to hedge permanent labor income risk.

Our approach posits that individuals display heterogeneous exposure to stock market returns, as discussed in Campbell and Viceira (2002) and Guvenen et al. (2017). Accordingly, we allow for correlation between stock market returns and individual PI shocks in a labor income process that is commonly used (see Carroll and Samwick, 1997 and Cocco et al., 2005) in the literature. We then take the resulting implied moment conditions to the data. Importantly, these conditions enable exploitation of the large cross-sectional dimension of the data to estimate the individual correlation coefficients. This is because the labor income shocks of any two individuals co-move, due to their

[1] See e.g. Heaton and Lucas (2000), Campbell, Cocco, Gomes, and Maenhout (2001), Campbell and Viceira (2002), Cocco, Gomes, and Maenhout (2005), Angerer and Lam (2009), Bonaparte, Korniotis, and Kumar (2014).

common exposure to the stock market return, in proportion to their correlation coefficients.

More precisely, individual 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 permanently affect the level of individual labor income, whereas the latter only impact the current earnings. Moreover, PI shocks feature both idiosyncratic and systematic components, with the latter that co-moves with the stock market return in our setting. The modeling of this income process embeds two usual restrictions that we exploit to identify individual-specific PI shocks. A first restriction is that of a zero intertemporal covariance of the idiosyncratic transitory shock. This implies that we can retrieve the variances of all the unobservable components of income shocks through the observable intertemporal covariances of the total income shocks at different lags. The second restriction is that of a zero covariance of the idiosyncratic component across different agents. This implies that the observable co-movements between any two individuals' shocks to total labor income are due to each shock's dependence on the stock market return through its permanent component. We use both sets of restrictions, while previous methods only exploit the former.

We proceed to estimate the relevant parameters characterizing the joint distribution of incomes and stock market returns with data from the Dutch National Bank Household Survey (DNB), a large annual panel covering from 1993 to 2019 period. Having multiple years of data for each individual allows us to obtain more precise estimates of the correlation between income growth and stock market returns. The rich information about individuals is the reason why a closely related paper (Bonaparte et al., 2014) prefers it to commonly used US data sets such as the Survey of Consumer Finances (SCF) or the Panel Study of Income Dynamics (PSID). We show how to exploit such rich information on participants in the DNB survey to perform the MD estimation on 80 clusters (MD80), grouping individuals according to time-invariant observable characteristics, including education, urbanization of the household residence, and risk aversion. Finally, we study whether such MD80 correlation estimates pin down the variation in equity investments.

We also test the strength of our method, relative to existing ones that exclusively rely on the time dimension of the data, as we vary the number of waves. We therefore perform our analysis based on MD80 estimates not only on the full sample but also on the 1993-2011 waves. This allows to compare results with (Bonaparte et al., 2014), that finds support for income hedging motives using a moving-average method. The use of a subset of waves also enables to perform an out-of-sample experiment on the remaining waves. Finally, we apply our Minimum Distance (MD) method also on PSID data on stock market participation, obtaining similar results.

We find that the propensity to participate in the stock market is highly sensitive to the MD80 correlation between PI shocks and stock market returns. Moreover, the explanatory power of the MD80 correlation estimates for portfolio holdings is robust, also relative to competing methods, not only in long samples ($T = 26, 18$) but also also in shorter ones ($T = 14$). These results imply that income hedging motives are more relevant to individual risk-taking than previously thought and that they can be precisely measured using a limited time-series data dimension. This second feature allows us to implement an out-of-sample analysis, which confirms that our MD80 estimates are able to explain participation decisions.[2]

This paper also departs from the extant literature by studying changes in individuals' decisions over time based on the realized sequence of PI shocks. Despite a common cluster-based correlation, different realizations of individual labor income will imply heterogeneous updated correlations. We estimate such revised correlations after reconstructing the dynamics of individual PI shocks using a Kalman filter. This exploits the assumed relationship between labor incomes and stock market returns, their realizations and the MD80 parameter estimates. We relate these revised correlations, estimated on an expanding window, to a counting variable equal to the number of waves in which the individual has invested in stocks. The regression results show that the decision to remain in the market is informed by both this revised correlation between PI shocks and stock market returns and the MD80 correlation. These revised correlations always display explanatory power, adding to the impact of hedging needs on risk-taking compared to previous approaches not exploiting the model-implied moment conditions.

This analysis, aside from its positive implications for stock market participation, is relevant to delegated portfolio managers and (robo-) advisors. For example, a new participant can initially be assigned to one of the 80 clusters based on their observable characteristics. Then, the manager or advisor revises asset allocation over time after updating their cluster's correlation. An online appendix calibrates a life-cycle portfolio choice model using our estimates. This delivers the optimal age-glide path for heterogeneous groups of individuals displaying different correlations between PI shocks and stock market returns. When stock market returns and PI shocks are uncorrelated, the optimal share invested in stocks until the age of 30 is 100% (100%); then, it gradually decreases,

[2]This supports the importance, stressed in Guiso and Terlizzese (1996), of *ex-ante* measures of future income risk in explaining portfolio choice.

reaching 40% (20%) at retirement for a relative risk aversion equal to 5 (8). With correlation equal to 0.5, the optimal share invested in stocks drops to 47% (0%) at the beginning of a person's working life, before reaching 25% (11%) at retirement. These findings indicate the possibility of personalized implementation of target-date funds, a common vehicle for pension investing (Mitchell and Utkus, 2020).

Our approach builds on Carroll and Samwick (1997), which breaks down individual earnings shocks into a permanent shock and a transitory shock. It is well known that identifying PI shocks requires a long labor income time series data for estimation (e.g., Carroll, Hall, and Zeldes, 1992; Carroll, 1997), especially for structural methods (Meghir and Pistaferri, 2004). Guvenen (2009) adopts a parsimonious structural representation to provide MD estimates of the persistence of income shocks. Using a similar parsimonious methodology, we estimate the individual correlations between PI shocks and stock market returns. This is made possible by explicitly allowing for such correlations in the stochastic processes characterizing the joint distribution of earnings shocks. Exploiting the implied co-movement of earnings growth across agents, that was so far ignored, our approach precisely pins down the heterogeneous correlations between PIH and stock returns.

Several papers estimate the co-movement between wage risk and stock market returns, based on the earnings decomposition pioneered by Carroll and Samwick (1997), with inconclusive results. Campbell and Viceira (2002) find that the contemporaneous correlation between labor income shocks and stock returns is low (0.06-0.1), while the correlation with lagged stock returns is higher (0.32-0.5). In Cocco, Gomes, and Maenhout (2005), there is a negligible correlation between stock market returns and labor income shocks. Angerer and Lam (2009) find that the effects of covariance measures on risky asset share are insignificant both statistically and economically. According to Guvenen et al. (2017), such traditional approach underestimates systematic risk by ignoring the differential exposure across workers to aggregate risk factors. It therefore misinterprets the residual from the wage regression as purely idiosyncratic (that is, unrelated to aggregate outcomes) when in fact it contains systematic risk. Guvenen et al. (2017) estimate the "wage betas" by clustering individuals and regressing earnings on aggregate risk factors to make the residual closer to the theoretical concept of idiosyncratic risk. Our approach accounts for both clusters and one aggregate risk factor. The mean correlation between PIH and stock market returns is as high as $0.2 - 0.3$ when we cluster individuals, while it is as low as 0.05 when we do not. This happens both on Dutch and on US data. Imposing the same correlation for each agent within a cluster reveals the common exposure to aggregate risk that each individual may occasionally be able to shield through a new job (as in Low, Meghir, and Pistaferri (2010)) or informal insurance (Guvenen and Smith (2014)). In order to assess the added empirical contribution of our approach, we consider alternative correlation metrics in our regression analysis following Campbell and Viceira (2002) and Guvenen et al. (2017). [3]

Our paper thus contributes to the understanding of heterogeneous hedging motives in financial risk-taking. An early study of PI shocks (Angerer and Lam, 2009) indicates that the variance of the permanent component of labor income shocks affects the share of risky assets in household portfolios. Later, although Betermier, Jansson, Parlour, and Walden (2012) find strong relationships—over time for the same individuals—between changes in the volatility of human capital and changes in portfolio holdings, the cross-sectional relationship is weaker. We also find a relatively weak response to income volatility, as in Betermier et al. (2012), especially for the direct component of equity investments. However, our conclusion is that individual agents appear to hedge labor income risk consistently, both in the cross section and in the time series, provided that the correlation between PI shocks and stock market returns is controlled for. Recently, Fagereng, Guiso, and Pistaferri (2018) find substantial sensitivity of portfolio decisions to uninsured wage risk in a long panel data of firms and their workers, through an ingenious identification strategy, also demonstrating that a large part of firm-level permanent shocks is passed on to wages.

While this evidence pins down the role of earnings volatility in risk taking, the one of correlation between earnings shocks and stock returns remains inconclusive also in more recent papers. [4] Massa and Simonov (2006) focus on explaining individual portfolio tilts away from the market portfolio, finding that familiarity is significant in explaining them while hedging motives are not. A subsequent study, Arrondel, Pardo, and Oliver (2010), uses a survey-based proxy for both correlation and earnings uncertainty. They find that earnings risk affects the decision to hold risky assets for French households whose earnings are non-negatively correlated with financial returns, only. In Calvet and Sodini (2014), the beta of income shocks on a household's portfolio return does not comove with that household's risky share. Against this background, Bonaparte et al. (2014) identifies the

[3] Differently from Guvenen et al. (2017), we overlook differential exposure to both employer- and industry-level risk when using DNB data because of availability.

[4] This leaves equity market participation largely disconnected from hedging motives. Alternative mechanisms suggested to explain participation include a fixed participation cost (Haliassos and Michaelides, 2003), the degree of trust in the stock market (Guiso, Sapienza, and Zingales, 2008), ambiguity aversion (Dimmock, Kouwenberg, Mitchell, and Peijnenburg, 2016), a stochastic interest rate possibly correlated with earnings shocks (Munk and Sorensen, 2010), mean reversion in stock market returns (Michaelides and Zhang, 2017), and cyclical skewness risk (Catherine, Sodini, and Zhang, 2020), among others cited in Gomes, Haliassos, and Ramadorai (2020).

role of the individual sample correlation between labor income shocks and stock market returns in explaining stock market participation. While our broad take-away fully aligns with theirs, our work contributes an original approach to the assessment of hedging motives. Thanks to it, individual hedging motives are able to explain both portfolio composition and equities market participation, also out-of-sample. Moreover, both the MD80 cluster-based and the revised individual correlations we estimate significantly impact the frequency of stock market participation, when we follow the same individuals over time, contrary to the sample correlation used in previous work. These results imply that heterogeneous hedging motives explain both non-participation in the stock market and the observed low equity share in individuals' portfolios, consistent with life-cycle theory.

The rest of the paper is organized as follows. In Section 2, we introduce the model for the individual labor income process and stock market returns and provide details concerning the estimation strategy. In Section 3, we describe the data, the clusters and the distribution of the estimated correlations between PI shocks and stock market returns. In Section 4, we link stock market participation and asset allocation choice to such correlations. We also update correlation estimates on an expanding window, linking them to equities market participation revisions over time. Section 5 confirms the robustness of the results using a shorter sample and reports our out-of-sample analysis. Section 6 presents the paper's concluding remarks.

2 The Model and the Implied Moment Conditions

This section presents the joint stochastic process of individual labor income and stock market returns, capturing the notion that agents are differently exposed to stock market shocks through the permanent component of their labor income shocks. Then, we exploit the restrictions imposed on both the cross-sectional and the time-series distribution of individual income shocks to derive the moment conditions implied by the model. The first restriction is the zero intertemporal covariance of the idiosyncratic transitory shock for each agent. The second restriction is the zero covariance of the idiosyncratic component across different agents. Finally, we present the sample counterparts of the model-implied moment conditions and formalize the MD estimator for the unknown parameters that characterize the joint stochastic process.

To begin, consider an economy with N individuals, indexed by i. At each time t, each individual receives the labor income, $Y_{i,t}$, and the rate of return on the stock market, $r_t = \sigma_r W_t$, is realized. Here, σ_r denotes the standard deviation of the stock market returns and W_t is a standard normal random variable, such that $r_t \sim \mathcal{N}(\mu_r, \sigma_r^2)$.

Each individual works for a period of length T. We describe the log-labor income process, following Cocco et al. (2005), as the sum of a deterministic function of a vector of observable characteristics, $Z_{i,t}$, and a stochastic component, $e_{i,t}$:

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

The stochastic log-labor income is, in turn, the sum of two components:

$$e_{i,t} = v_{i,t} + \epsilon_{i,t}, \tag{2}$$

where $v_{i,t}$ is a random walk with shocks $u_{i,t}$:

$$v_{i,t} = v_{i,t-1} + u_{i,t}, \tag{3}$$

where $u_{i,t} = \sigma_u W_{i,t}^p$, and $\epsilon_{i,t} = \sigma_\epsilon W_{i,t}^q$, where $W_{i,t}^p$ and $W_{i,t}^q$ are two standard normal random variables. We refer to $u_{i,t}$ and $\epsilon_{i,t}$ as the permanent and the transitory shocks, respectively, of the log-labor income.

We can express the PI shock, $u_{i,t}$, as the sum of a systematic component, $\xi_{i,t}$, and an idiosyncratic component, $\omega_{i,t}$:

$$\xi_{i,t} \sim \mathcal{N}(0, \sigma_u^2 \rho_i^2),$$
$$\omega_{i,t} \sim \mathcal{N}\left(0, \sigma_u^2(1-\rho_i^2)\right),$$

and thus we can express the PI shock as a linear combination of two normally distributed random variables,

$$u_{i,t} \sim \mathcal{N}(0, \sigma_u^2 \rho_i^2 + \sigma_u^2(1-\rho_i^2)), \quad \text{i.e.} \quad u_i, t \sim \mathcal{N}(0, \sigma_u^2) \tag{4}$$

The interpretation of (4) is simple: the variance of the PI shocks is the sum of systematic and idiosyncratic variances, where the relative weight of the systematic and the idiosyncratic components is given by the correlation between PI shocks and stock market returns, denoted by ρ_i.

2.1 Model-Implied Moment Conditions

This section derives two moment restrictions. The first retrieves the idiosyncratic variance of the unobservable idiosyncratic PI shock from the observable intertemporal covariance of the total shock at different lags. The second restriction is that the observable covariance between any two individual shocks to total labor income, due to the dependence of each individual PI shock on the stock return, is a linear function of the two individuals' correlation coefficients. This second restriction gives rise to $N(N-1)/2$ conditions that will allow to estimate the N correlation coefficients between PI shocks and stock returns.

The total shocks (TS) to labor income for individual i at time t, $e_{i,t}$, are defined in equation (2). Then, we compute the time variation in TS for each individual over a time interval of length d, and we define it as DTS:

$$\Delta_d e_{i,t} = e_{i,t+d} - e_{i,t} \tag{5}$$

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

A useful property of $\Delta_d e_{i,t}$ is that it contains only permanent and transitory income shocks:

$$e_{i,t+d} - e_{i,t} = \sum_{s=t+1}^{t+d} u_{i,s} + \epsilon_{i,t+d} - \epsilon_{i,t}, \tag{6}$$

because 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 TS and is equal to:

$$\Delta_1 e_{i,t} = u_{i,t+1} + \epsilon_{i,t+1} - \epsilon_{i,t} \tag{7}$$

Then, we construct two sets of variance-covariance DTS matrices. The first set includes the $[N \times N]$ matrix containing the variance of each individual DTS time series on the main diagonal and the covariances between individual DTS 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 of the DTS for each point in time. This is the variance of the N-dimensional vector containing the DTS of N individuals at each time t. Off the main diagonal are the covariances between time periods; that is, the covariances between the N-dimensional vectors containing the DTS of N individuals at different time periods. Then, each $[(T-d) \times (T-d)]$ matrix features the following symmetrical form:

$$\begin{bmatrix} C_d(1,1) & C_d(2,1) & C_d(3,1) & . & . & C_d(T-d,1) \\ C_d(2,1) & C_d(2,2) & . & . & . & . \\ C_d(3,1) & . & . & & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ C_d(T-d,1) & . & . & . & . & C_d(T-d,T-d) \end{bmatrix} \tag{8}$$

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

$$C_d(t, t+l) = cov\left(\Delta_d e_t, \Delta_d e_{t+l}\right), \tag{9}$$

where e_t is the N-dimensional vector containing the TS 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 DTS corresponding to lag d:

$$C_d(t,t) = var\left(\Delta_d e_t\right) = d\sigma_u^2 + 2\sigma_\epsilon^2 \tag{10}$$

When $l > 0$, $C_d(t, t+l)$ identifies the covariance terms between time periods, which are equal to

$$\begin{bmatrix} C_d(t,t+l) = (d-l)\sigma_\omega^2 & d > l \\ C_d(t,t+l) = -\sigma_\epsilon^2 & d = l \\ C_d(t,t+l) = 0 & d < l \end{bmatrix} \tag{11}$$

Observe that we have isolated the variance of the individual-specific transitory income shocks from the variance of the individual-specific PI shocks by exploiting the temporal variation of the DTS for each individual.

We now turn to the $[N \times N]$ covariance matrix, including the variance of each i-th individual time series of shocks on the main diagonal and the covariances between individual shocks off the main diagonal.

The $[N \times N]$ matrix features the following symmetrical form:

$$\begin{bmatrix} C(1,1) & C(2,1) & C(3,1) & . & . & C(N,1) \\ C(2,1) & C(2,2) & . & . & . & . \\ C(3,1) & . & . & & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ C(N,1) & . & . & . & . & C(N,N) \end{bmatrix} \tag{12}$$

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

$$\begin{bmatrix} C(i,j) = \sigma_u \sigma_r \rho_i \rho_j & i \neq j \\ C(i,j) = \sigma_u^2 + 2\sigma_\epsilon^2 & i = j \end{bmatrix} \tag{13}$$

because the covariance between individuals is due to the correlation with the aggregated shocks:

$$cov(\Delta_d e_{i,t}, \Delta_d e_{j,t}) = cov(u_{i,t+1}, u_{j,t+1}) =$$

$$cov(\sigma_u \sigma_r^{-1} \rho_i r_{t+1}, \sigma_u \sigma_r^{-1} \rho_j r_{t+1}) = \sigma_u \sigma_r \rho_i \rho_j,$$

where the first line is due to the orthogonality of the transitory shocks, the second line is due to the orthogonality of the permanent idiosyncratic shocks, and we use $W_t = r_t/\sigma_r$. Then, we expand the matrix using the condition on the covariance between each individual DTS time series and the stock market returns:

$$cov(\Delta_1 e_{i,t}, r_t) = \sigma_u \sigma_r \rho_i, \tag{14}$$

Accordingly, we ultimately obtain a $[(N+1) \times (N+1)]$ matrix in which the last row and the last column are populated with the condition defined in the equation (14).

Note that this matrix's main diagonal elements provide information on the variance parameters,

σ_u^2 and σ_ϵ^2. In contrast, all of the elements off the main diagonal provide information on the correlation parameters, $\{\rho_i\}_{i=1}^N$, because the co-movements from individual shocks to labor income are due to the dependence of the individual shocks on the common aggregate shocks. This feature of the economy modeled here enables exploitation of the data's cross-sectional dimension to make inferences regarding the correlation parameters. In fact, while the number of correlation parameters increases linearly with N, the number of model restrictions depending on those parameters grows exponentially with N and is specifically equal to $(N(N-1))/2$.[5]

2.2 Minimum Distance Estimation

Now, we derive the sample counterparts of the model-implied moment restrictions before turning to the Minimum Distance estimation.

Let us first identify the sample counterparts of the labor income shocks. We estimate a panel regression of the log-labor income on an age polynomial up to the fourth order and a set of observable personal characteristics, including sex, education, and their interactions. The fitted value of this regression is the deterministic component of the log-labor income, $f(t, Z_{i,t})$, with the regression residuals representing the stochastic component.

The empirical counterparts of the individual income shocks, $(\Delta_d e_{i,t})$, are the first differences between regression residuals for each individual and are denoted as $\underline{\text{dres}}$. Then, we construct two sets of variance-covariance matrices of the $\underline{\text{dres}}$. First, we populate the sample counterpart of the matrix $[N \times N]$ as follows: we fill the off-diagonal entries by computing the covariance between each pair of individuals' $\underline{\text{dres}}$ and we fill the on-diagonal entries by computing the variance of each individual's $\underline{\text{dres}}$. Then, we expand this matrix using the empirical covariances between the individuals' $\underline{\text{dres}}$ and the excess stock market returns and the variance of the excess stock market returns, thus forming a matrix $[(N+1) \times (N+1)]$. For the set of matrices $[(T-d) \times (T-d)]$, we use all of 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 $[(T-d) \times (T-d)]$ matrix, and the elements $C_d(t, t+l)$ are on the l-th diagonal below the main one.

Finally, we formalize the MD estimator of the vector θ, which contains the unknown parameters of the labor income process and the stock market returns:

[5] This number is equal to the number of entries of the lower triangular part of a symmetrical matrix of dimension $[N \times N]$.

$$\theta = \left\{ \{\rho_i\}_{i=1}^{N}, \sigma_u, \sigma_\epsilon, \sigma_r \right\}.$$

Using $\{G_m(\theta)\}_{m=1}^{M}$, we denote the set of M moment conditions implied by the model, which depend on the vector of unknown parameters θ, and we stack all of the moment conditions in one M-vector:

$$\mathbf{G}(\theta) = [G_1(\theta), ..., G_M(\theta)].$$

Next, using $\{g_m\}_{m=1}^{M}$, we denote the set of M empirical counterparts, and we stack all of the sample conditions in one M-vector:

$$\mathbf{g} = [g_1, ..., g_M].$$

Then, the MD estimator searches for the value of θ that minimizes the following quadratic form:

$$Q(\theta) = (g_M - G_M(\theta))' I_M (g_M - G_M(\theta)) \tag{15}$$

where I_M is an identity matrix of size M. We choose an identity matrix as a weighting matrix following Guvenen (2009), which shows that an MD estimator that weighs moments with an identity matrix is asymptotically consistent and normal.[6]

3 Data and Clusters

As discussed, our dataset derives from the DNB Household Survey, which has provided information on annual labor income for a representative sample of the Dutch population since 1993.

Three concerns dictate using the DNB survey instead of the US survey data often used in the household finance literature (e.g., the Panel Survey Income Dynamics). The first is the availability of information on financial investments, such as the decision to participate in the stock market, at the individual level over the entire time span. Second, the DNB is the reference dataset in the empirical assessment of hedging motives for stock investing by Bonaparte et al. (2014), which we

[6]We also perform a two-step estimation by replacing—in the second step—the I_M with a diagonal and positive-definite optimal weighting matrix obtained in the first step. The results are identical.

adopt as a natural benchmark to which to compare our results. Finally, the large cross-sectional dimension of these data enables precise measurement of the income-return correlation parameter, even for short time-series data dimensions (e.g., 1993–2011). Moreover, the rich personal characteristics information provided by the DNB—which includes age, education, health, risk aversion, and wealth—allows individuals to be grouped according to such observable features. Meanwhile, we use data from the Dutch stock market index to estimate the correlation between individual labor income growth and stock market returns.

[Table 1 about here.]

We detail the variables used in our analysis in Table 1 and we report descriptive statistics for the sample up to 2019 in table 2. The average age is around 56, half of the individuals have obtained a college degree, slightly more than half are male, one out of ten individuals is unemployed, the average health status is good, and the average level of risk aversion is moderate. One-third of the sample holds stocks either directly or through mutual funds, which aligns with participation rates in other developed countries, such as the US and the UK. [7] There is large cross-sectional heterogeneity in terms of correlation between labor income growth and stock market returns and in terms of variation of labor income over time, as measured by the standard deviation of labor income growth. We also present descriptive statistics for the short sample up to 2011 in table 16, in which we report very similar figures to the longer sample in terms of personal characteristics of the individuals.

[Table 2 about here.]

3.1 Relating Correlations to Observable Characteristics

If income hedging necessities relate to observable personal characteristics, similar correlations will be observed between labor income shocks and stock market returns for individuals with similar traits.

Accordingly, we cluster individuals through a set of observable variables and estimate a correlation coefficient for each cluster. Our estimation approach incorporates this additional restriction. In fact, it is sufficient to prescribe that the correlation matrix defined in equation (4) be cluster-specific rather than individual-specific, an approach that still allows both permanent and transitory

[7] The DNB does not provide information on the individual stocks held by individuals. It does not give information on stocks held through pension funds that are collective, rather than individual, holdings.

income shocks to be individual-specific. Meanwhile, we assume that each individual i belonging to the cluster k has the same correlation, such that equation which describes individual PI shocks, becomes:

$$u_{i,t} = \sigma_u \left(\rho_k W_t + \sqrt{1 - \rho_k^2} W_{i,t}^p \right),$$

where ρ_k denotes the common correlation parameter for the k-th cluster, to which the individual i belongs.

This procedure presents multiple advantages. From the econometric angle, it reduces the number of unknown parameters, increasing the ratio between the number of informative moment conditions and the corresponding parameters requiring estimation. Economically, we become able to link time-invariant traits, which are known at the beginning of the working period, to hedging needs.

To ensure consistency with both the model and the empirical strategy, the clustering variables must fulfill two conditions. First, the observable variables must be almost time-invariant because the cluster-specific correlation does not change over time. Second, the observable variables should explain the individual log-labor income.

Following the first requirement, we select as clustering traits education, sex, level of urbanization of the household's residence, risk aversion, and financial literacy. Education is a discrete variable denoted by five different values corresponding to the highest level of education attained by the individual. Sex is a dummy variable equal to 1 if the individual is male and 0 otherwise. Urbanization is a dummy variable equal to 1 if the individual lives in an urban area and 0 otherwise. Risk aversion is a dummy variable equal to 1 if the individual displays a DNB risk aversion variable value greater than 5 and 0 otherwise (the DNB variable receives a value between 1 and 7, with 7 indicating very high aversion to risk-taking). Financial Literacy is a dummy variable equal to 1 if the individual reports being knowledgeable with respect to financial investing and 0 otherwise. These variables are recorded for each survey wave for each individual. While they are mostly constant over time, we input to an individual the value of the mode for each variable when the variable displays different values over time. In the Appendix A, we corroborate our clustering procedure with a simple regression analysis.

By combining the number of possible outcomes of each clustering variable, we form a grid of 80 clusters to which each individual can belong, and we assign the individuals to the corresponding cluster.[8] Figure **??** shows how the individuals are distributed into the corresponding clusters. While we only need one individual per cluster to estimate the corresponding correlation parameter, very few clusters are either scarcely populated or extremely crowded, with most clusters having a similar number of individuals.

Finally, we estimate 80 correlation coefficients, which are presented in Figure 2. The left-hand hand panel compares the distribution of the MD correlations to the one of sample correlations between stock market returns and TS. The right-hand panel plots the MD correlations against the correlations between stock market returns and PI shocks obtained with the approach of Bonaparte et al. (2014).

Results are striking. The distribution of MD correlation parameters shifts to the right with respect to both alternatives, indicating larger average hedging needs. The positive mean correlation signals that the PI shocks of most agents have the same sign of business cycle movements which are, in turn, anticipated by the stock market return. Thus, it appears that the MD restrictions based on the covariance matrix of contemporaneous income innovations capture these co-movements in individual PI shocks that instead escape methods relying on the time series of individual income shocks only. As suggested by Guvenen et al. (2017), ignoring the differential exposure across clusters of workers to aggregate risk leads to underestimating that exposure. It therefore misinterprets the residual from the wage regression as purely idiosyncratic (that is, unrelated to aggregate outcomes) when in fact it contains systematic risk. [9]

In more detail, 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 $t+1$. Therefore, the estimate for the correlation between PI shocks and stock market returns derived in Bonaparte et al. (2014) is based on the realized correlation between labor income shocks and stock market returns. It transpires that this distribution turns out to be widely dispersed around its mean, like the one between total income shocks and stock returns.

In contrast, MD estimates are based on the correlation between individual labor income shocks and a common risk factor driving the relationships between individuals' labor income through

[8] Given five outcomes for education, two for Sex, two for urbanization, two for risk aversion, and two for financial literacy, we obtain 80 clusters (5x2x2x2x2).

[9] In the Appendix A we report the same distribution for individual, non cluster-based, correlations. Each individual may be able to shield an aggregate shock through a new job (as in Low, Meghir and Pistaferri (2020)) or informal income support (Guvenen and Smith (2014)). The mean correlation drops from 0.257 to 0.057. Imposing the same correlation for each agent within a cluster isolates the common exposure to aggregate risk.

the relationship between individual labor income and the stock market. The distribution of MD correlations is heavily concentrated around positive values, with the proportion of individuals characterized by a negative correlation between PI shocks and stock market returns dropping from 36% to 11% when estimated using the approach of Bonaparte et al. (2014). In Appendix, we report summary statistics of the correlation parameter estimates (Table 3).

[Figure 1 about here.]

[Figure 2 about here.]

[Table 3 about here.]

4 Hedging Heterogeneous Permanent Income Shocks

We now turn to the paper's main question, which is whether the new measure of hedging motives explains financial risk-taking decisions. In the first part of this section, we highlight differences between results based on our parametric approach to the measurement of PI shocks and the moving average one in Bonaparte et al. (2014), using the same specification but for the correlation coefficient between PI shocks and stock returns. We analyze the decision to participate in the stock market in Section 4.1 and asset allocation decisions in Section 4.2, observing the ways our estimates reveal a considerable economic impact of hedging motives on individual risk-taking.

Then, we exploit the panel dimension of our data in order to see whether revisions in individual correlations explain revisions in portfolios choices over time for a given individual. We address revisions to risk-taking choices over time that are associated with realizations of both income and stock market returns. While the previous literature has generally treated each observation over time as a separate agent, we follow each individual over time. In Section 4.3, we reconstruct the dynamics of PI shocks of each individual before estimating updated correlations and linking revisions to participation and asset allocation to these updates. This is a second check on the quality of our approach to estimating PI shocks.

4.1 Stock Market Participation

[Table 4 about here.]

[Table 5 about here.]

We describe the stock market participation decisions using the dummy variable $\mathcal{I}_{i,t}$, which takes a value of 1 if the individual i invests in the stock market at time t and takes 0 otherwise. Our main determinant of interest—for the decision to invest—is the correlation with the stock market returns of shocks to different specifications of labor income: total labor income, the deterministic component (i.e., the fitted value of the panel log-labor income regression), the stochastic component (i.e., the residual of the panel log-labor income regression), the transitory and the permanent components as computed in Bonaparte et al. (2014), and the permanent component estimated using our approach.

[Figure 3 about here.]

First, we present graphical evidence about the relationship between stock market participation and the correlation between income and returns. We rank individuals according to the level of correlation between labor income shocks and stock market returns for different specifications of labor income shocks, and we plot the average participation rate for the bottom (left bar) and the top quartiles (right bar) in Figure 3. The figure shows that individuals displaying lower correlations participate more, on average, compared to those with higher correlation, when considering TS and stochastic income shocks, which aligns with the income hedging motive. Importantly, the figure shows that the average participation rate varies substantially across the bottom and top quartiles when considering the correlation between PI shocks and stock market returns estimated using our approach and upon individuals being clustered in homogeneous groups. In contrast, we do not observe similar significant heterogeneity in terms of average participation rate when considering the correlation between deterministic, transitory, and permanent income shocks and stock market returns estimated using the approach described in Bonaparte et al. (2014).

Then, we estimate the probability of participating in the stock market by performing a probit regression, where the dependent variable is $\mathcal{I}_{i,t}$, and the explanatory variables include our key determinants and a set of personal characteristics that are likely to impact the decision to invest. For all of the regressions, we control for the income and wealth levels, age, education, sex, risk aversion, family size, and health status. We also control for situations in which the individual is retired or unemployed. The choice of control variables is motivated by the correlation that these variables may have with either direct or indirect costs of participation and align with Bonaparte et al. (2014). The results for the stock market participation probit regressions are reported in Table 4 and in Table 5.

In Table 4, columns (1) and (4) confirm the negative association between total income shocks and stock market participation. Columns (2) and (5) split the total income shocks into a deterministic and a stochastic component, confirming the results of Bonaparte et al. (2014) that the latter component plays the most prominent role in reducing participation. While the regression coefficient for the correlation between stochastic income shocks and stock market returns is always negative and significant at the 0.1% level, its counterpart for deterministic shocks is positive and loses statistical significance when considering controls. Columns (3) and (6) show, importantly, the large economic and statistical significance associated with the MD80 correlation between the permanent component of stochastic shocks and stock returns. This result confirms the broad take-away of Bonaparte et al. (2014), that the permanent component of the labor income shocks drives the negative relationship between the income–stock-returns correlation and stock market participation. Our estimates of the marginal effect of the control variables in columns (4)-(6) show that individuals are more willing to participate in the stock market when they are wealthier, more educated, less risk-averse, and have a smaller family. Meanwhile, sex and income level do not play a significant role in column (6). Similarly, it is not significantly important whether the individual is retired or unemployed.

In Table 5, we see that the predictive power of MD80 estimates of correlation for participation holds when we control for the correlation between stock market returns and different components of income shocks computed following the methodology in Bonaparte et al. (2014) (see columns (1)-(2)). [10] In Table 5, we also include the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen, Schulhofer-Wohl, Song, and Yogo (2017) (columns (3-4)) and Campbell and Viceira (2002) (columns 5-6) at the cluster-level. Specifically, we obtain β_{GUV} by estimating pooled OLS regression of the (log)-real earnings growth of individual i in year t over the stock market return in year t separately by cluster, so that individuals belonging to the same cluster share the same β_{GUV}. To estimate β_{CV}, we first compute the cluster-specific average of the (log)-real earnings growth across the individuals belonging to each cluster, in order to obtain cluster-specific permanent income shocks under the restriction that transitory income shocks cancel out across individuals. Then, we regress cluster-specific permanent income shocks in year t over the stock market return in year $t+1$, following the approach of Campbell and Viceira (2002). These cluster-based beta following Campbell and Viceira (2002) display statistical significance (see column 5). However, this disappears when we also allow in the regression the MD80 correlation estimates.

[10] The transitory component is computed by subtracting the permanent component from the stochastic log-labor income.

In Table 5, the regression coefficient for the correlation between PI shocks and stock market returns computed using the approach of Bonaparte et al. (2014) is smaller in magnitude compared with the regression coefficient for the correlation between PI shocks and stock market returns estimated using our MD80 approach.

We then assess the overall model fit of our probit estimates. We compare the model-implied probability of participation in the stock market and the observed stock market participation rate, for each year of our sample. The probability of participation in the stock market, according to the probit model, is given by

$$P(\mathcal{I}_{i,t} = 1) = \Phi\left(\sum_{k=1}^{K} \beta_k X_{k,i,t}\right), \tag{16}$$

where Φ indicates the cumulative distribution function of a standard normal variable, β_k is the coefficient estimated with the probit regression for the variable k, and $X_{k,i,t}$ is the value of the k-th independent variable for individual i at time t.

We compute $P(\mathcal{I}_{i,t} = 1)$ for each year and for each individual, using the probit regression estimates reported in Table 4 (column (6)), and we compute the average probability across individuals for each year. Figure 4 shows that the predicted level of participation based on our MD estimates aligns with the observed one for each year of our sample.

[Figure 4 about here.]

4.1.1 Robustness

We address potential endogeneity issues arising from using individual wealth as control variable in our regressions in two ways. First, we control for non-financial wealth only, rather than total wealth, excluding the financial items of net worth. Non-financial wealth in DHS accounts for the individual housing property, since it includes real-estate items. Second, we construct dummy variables for quartiles of wealth and we include these dummies as control variables in our regressions instead of the original continuous variable, following the approach of Van Rooij, Lusardi, and Alessie (2011). In both cases, we obtain results that are quantitatively equivalent to those reported in the regression tables.

Then, we perform the entire empirical analysis using the individual's total income rather than the individual's labour income. The total income includes additional income components such as transfers from partner and other members of the household. All the results still hold. These robustness check are repeated in all subsequent regression, without noteworthy changes in results.

4.2 Asset Allocation

This section relates the asset allocation decision to the correlation between labor income shocks and stock market returns. The dependent variable is the share of the individual portfolio invested in stocks either directly, through mutual funds, or both. The main explanatory variables of interest are the same as for the earlier analysis concerning the decision to participate.

Table 6 reports results from the Tobit regression for individual asset allocation (see columns (1)-(6). First, we again confirm the broad takeaway from Bonaparte et al. (2014): individuals demonstrating low correlation between labor income shocks and stock market returns invest more in risky assets. In particular, the correlation between PI shocks and stock market returns is important for asset allocation decisions. This correlation always significantly predicts the fraction of wealth invested in risky assets. This result holds whether or not we separately consider, in unreported regression, the total equity share of the individual and either direct investments in stocks or indirect investment in stocks.

In columns (4)-(6), we control for the set of observable characteristics that may impact an individual's portfolio allocation. We find that richer individuals, who are more educated and less risk-averse, generally invest larger fractions of their wealth in stocks, either directly or through mutual funds. Following Bonaparte et al. (2014), we also find that high income-risk individuals prefer to directly allocate wealth to stocks, rather than through mutual funds.

Vissing-Jorgensen (2002), Fagereng, Gottlieb, and Guiso (2017) and Bonaparte et al. (2014) consider simultaneously the market participation and asset allocation decisions, by estimating Heckman (1979) regressions in which the control variables in the selection model (i..e, the participation regression) and the control variables in the asset allocation regression are the same. We follow the same approach and report the results of Heckman (1979) regression estimates in columns (7)-(8). We consider lagged financial wealth and lagged squared financial wealth as additional control variables, as in Bonaparte et al. (2014). We find that the coefficient estimates of the MD80 correlation term are significantly negative, albeit smaller than the ones in the regression combining both participants and non-participant. This result is in line with Bonaparte et al. (2014) and is due to the limited size of the market participants sub-sample. For example, in the specification including the baseline control variables (column (8)), its estimate is -0.018 with 0.1% statistical significance. Also, the statistical significance of lambda confirms that the market participants sub-sample is not random.[11]

Our results are in line with prior empirical results in the literature. The following table reports the benchmarking anlaysis. The coefficients of the correlation between PI shocks and stock market return display the same order of magnitude and the same same level of statistical significance (5%) in columns (1),(3) and (5). When we add the MD80 correlation term in columns (2),(4) and (6), only the one estimated through the moving-average method in Bonaparte et al. (2014) keeps explanatory power.

So far results show that cluster-based MD estimates of correlation, that imply large hedging motives, have stronger explanatory power than estimates based on previous methods relying on survey data. This evidence is based on replicas of existing Probit, Tobit and Heckman regression specifications that focus on the cross-section. In the following section, we are able to perform a time series analysis of each individual over time. This exploits the modelling of the stochastic process for income to reconstruct the dynamics of individual PI shocks together with the MD estimates of the correlation parameter.

[Table 6 about here.]

[Table 7 about here.]

4.3 Learning about Income Hedging Needs

In this section we follow each individual over time to assess the relationship between changes in the participation decision and revisions in the correlation between PI shocks and stock market returns. To this end, we first reconstruct the sequence of unobservable PI shocks at the individual level using a Kalman filter, exploiting both the parameter estimates from previous sections and the sequence of total income shocks and stock market returns. We then construct the sequence of updated correlation coefficients between the PI shocks and stock market returns over an expanding window. These revised correlation coefficients will belong to the set of independent variables explaining the individual's probability to participate in the stock market in each period and the overall frequency

[11] If lambda is not statistically different from zero, the sample of market participants is randomly drawn from the population and the OLS estimator for the asset allocation decision is unbiased. Otherwise, the OLS estimator is biased and the Heckman (1979) correction is needed to obtain consistent estimates of the regression coefficients.

of individual participation.

4.3.1 Reconstructing Permanent Income

We pin down the unobservable PI shocks at the individual level using information on the total income shocks and the stock market returns. We formulate the labor income process described in Section 2 in a state-space model. We obtain the state-space representation using equations (2) and (3). Specifically, equation (2) forms the *measurement* equation that relates the observable total income shocks to the unobservable permanent component:

$$e_{i,t} = v_{i,t} + \epsilon_{i,t}, \tag{17}$$

where $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and equation (3) forms the transition equation that describes the dynamics of the latent permanent component:

$$v_{i,t} = v_{i,t-1} + u_{i,t}, \tag{18}$$

where,

$$u_{i,t} = \sigma_u \left(\rho_i (r_t/\sigma_r)\right) + \omega_{i,t}, \tag{19}$$

and $\omega_{i,t} = \sigma_u \left(\sqrt{1-\rho_i^2} W_{i,t}^p\right) \sim \mathcal{N}(0, \sigma_u^2(1-\rho_i^2))$.

We track the random walk $v_{i,t}$ using a linear Kalman Filter (KF). To implement this filter, we use estimates from Section 3 as parameters of the state-space model. We initialize the filter with an arbitrary value for $v_{i,0}$ and we form a prior for $v_{i,1}$, denoted by $\hat{v}_{i,1}$, by computing the expected value of $v_{i,1}$ conditional on both $v_{i,0}$ and the stock market return r_1. Our approach outlined in Section 2, in fact, allows us to exploit also the information on the stock market return to infer the latent variable. We next form a prediction of the total income shock $e_{i,1}$, $\hat{e}_{i,1}$, by computing the expected value of $e_{i,1}$ conditional on $\hat{v}_{i,1}$. The difference between the actual and the predicted total income shocks is the measurement error that is used to compute the posterior for $v_{i,1}$, which turns to be the prior for the next point in time. We iterate the system up to T and we reconstruct the PI shocks by computing the first differences of the random walk. Consequently, for each individual, we obtain the dynamics of the permanent shocks to the labor income over the entire time series.

The online Appendix C provides details regarding implementation of the KF.

[Figure 5 about here.]

[Figure 6 about here.]

To represent our results, Figure 5 plots the dynamics of the stock market returns and the PI shocks reconstructed using the KF for the individual with minimum and maximum sample correlation between stock market returns and PI shocks, respectively.

Meanwhile, Figure 6 plots the full distribution of the sample correlations between stock market returns and PI shocks reconstructed using the KF and compares it with both the distribution of the sample correlations between stock market returns and TS (left-panel) and with the distribution of individual correlations between stock market returns and PI shocks obtained by MD estimation (right-panel). In the left panel, the distribution of TS is widely dispersed over the entire set of the correlation values, as expected for sample realizations. However, the distribution obtained using the KF is oriented slightly to the right, signaling higher frequency of positive values for the correlation between stock market returns and PI shocks compared to TS. In the second case, the distribution obtained using the KF is more dispersed than the distribution obtained using the MD estimates, which does not embed information on realized individual earnings and stock market returns.

4.3.2 Participation Frequency

This section follows each individual agent over time to investigate whether their decisions to enter and exit the stock market in different waves are explained by their revised income hedging motives. In other words, we allow each individual to learn from their income realization concerning correlations between stock market returns and her PI shocks (Chang et al., 2018).

We track revisions in the correlation between PI shocks and stock market returns by sequentially computing the sample correlation between stock market returns and PI shocks obtained using the KF up to a given wave $t < T$, where T is the total number of available waves in our sample. Then, we estimate a probit model in which the dependent variable is equal to 1 if the individual invests in stocks, either directly or through mutual funds, in wave t, and the main independent variable is the correlation between PI shocks and stock market returns up to t. Accordingly, we relate the decision to enter or exit the stock market between t and T to the revision in correlation between stock market returns and PI shocks.

Table **??** reports results concerning the decision to invest in stocks both directly and through mutual funds, directly only, and through mutual funds only.

The dependent variable is a dummy variable describing the individual decision to participate in the stock market during a given wave, as it is for Table 4. Here, however, the main independent variable is the updated correlation estimated according to an expanding time window, which accounts for successive realizations of PI shocks and stock market returns. This probit regression reveals whether revised hedging needs, due to revised correlation between PI shocks and stock market returns over time, prompt participation revisions.

This table shows that the revised correlations significantly predict the sequence of individual decisions to participate in the stock market in subsequent waves. It demonstrates that the lower the revised correlation between stock market returns and PI shocks, the higher the propensity of the individual to enter (or remain) in the equities market. Both the economic and statistical significance of the MD80 correlation between PI shocks and stock market returns increase when we also control for the revised correlation, since they are based on different types of information. The MD estimates of PI on the one hand exploit the information embedded in the variance-covariance matrices of total income shocks, both the intertemporal one for each agent and the contemporaneous one across agents, in order to clean out the effect of both transitory and idiosyncratic shocks. On the other hand, they also exploit information about clusters. The revised estimates complement the MD estimates of PI shocks relying on the realization of both stock returns and idiosyncratic shocks over time. Thus, the latter information increase the relevance of MD correlation estimates, while both cluster and covariance information increases the economic relevance of the individual KF correlation.

Finally, we study the frequency of participation in the stock market for the whole sample using a Poisson regression, where the dependent variable is a discrete counting variable equal to the number of waves in which the individual invested in stocks. We use the same explanatory variables as in Table 4, including all of the unreported control variables. Table 8 confirms that individuals remain in the market longer if their labor income shocks are negatively correlated with stock market returns.

Importantly, the correlation between stock market returns and PI shocks has a negative and significant impact when this correlation is obtained using the KF, and the impact is much larger than that of the correlation computed using the approach described in Bonaparte et al. (2014).

[Table 8 about here.]

To assess the economic significance of our results, it is worth recalling that the Poisson model assumes that the dependent counting variable $\mathcal{F}_i$ features a Poisson distribution, with an expected value equal to

$$E[\mathcal{F}_i] = e^{\left(\sum_{k=1}^{K} \theta_k X_{k,i}\right)},$$

where $X_{k,i}$ is the k-th individual-specific covariate, and θ_k is an unknown parameter requiring estimation. Therefore, the marginal effect of the k-th variable is given simply by

$$\frac{\partial E[\mathcal{F}_i]}{\partial X_{k,i}} = \theta_k E[\mathcal{F}_i].$$

Consider, for instance, an individual participating in the stock market over 10 years. Increasing the revised correlation between PI shocks and stock market returns from -0.6 to 0.5 reduces participation by 21 months ($(-0.6 - 0.5) \times 0.162 \times 10$) (38 months if we consider 0.239, the sum of the coefficients of both MD80 and KF correlation terms). Instead, if we similarly increase the correlation between PI shocks and stock market returns, computed according to Bonaparte et al. (2014), the individual reduces participation in the equities market by 14 months ($(-0.6 - 0.5) \times 0.11 \times 10$).

5 Small T, Out-of-Sample Analysis and PSID Data: Results

The previous section exploits survey data characterized by a relatively large number of waves, from 1993 until 2019. Often, the temporal dimension of the data that is available to the researcher or the portfolio manager is much smaller. Section 5.1 will therefore show that our results hold when we use a much smaller number of waves. This experiment demonstrates the robustness of the MD estimation method that exploits both the cross-sectional and the time series dimension of the data. Such robustness allows us to set aside some observations in order to perform an out-of-sample analysis of participation, that will be presented in Section 5.2. This will show that the estimated correlation coefficients predict participation also out-of-sample, suggesting their use for improving on portfolio design. Section 5.3 challenges our MD-KF estimates with a different angle. We scrutinize the distribution of both PI and total shocks based across clusters that are similar in all characteristics but cohort, sex and risk aversion. We expect differences in participation to be more closely related to differences in PI shocks than total shocks (TS). Last but not least, Section 5.4 confirms the results on both the size of hedging needs and the determinants of participation on U.S. data from PSID. This confirmation also indicates that results are not an artifact of the clusters' characteristics since we necessarily have to change them based on data availability.

5.1 MD Estimates with small T

The online Appendix B reports summary statistics for the sample up to 2011, which are very similar to those presented in Table 2 regarding the full sample. Similarly, Table 9 and the associated figure displaying the distribution of cluster-based correlations confirm both the high average MD80 correlation between PI shocks and the stock market return (0.3) and the marked shift to the right of the distribution.

[Table 9 about here.]

[Figure 7 about here.]

Table 10 reports the probit regression results for stock market participation using the DNB survey waves from 1993 to 2011 in columns (1-3) and to 2007 in columns (4-6). This table shows the statistical and economic significance of the MD correlation coefficients in predicting participation, similar to the one based on surveys up to and including 2019. There are minor changes such as those in the statistical significance of some control variables, such as sex.

[Table 10 about here.]

Similarly, results of benchmarking in Table 11 confirm the relative strength of the MD method. In the sample including waves up to 2011, the BKK-correlation estimates with PI shocks have a statistical significant coefficient, as reported in Bonaparte et al. (2014). Otherwise, competing methods lose explanatory power. The economic significance of hedging motives, when assessed through the coefficient of MD80 correlation estimates, becomes even larger than in the full sample. Moreover, this coefficient is around three times larger than the coefficient for the correlation between PI shocks and stock market returns estimated according to Bonaparte et al. (2014).

[Table 11 about here.]

This section shows that our results on both the size of hedging motives and the sensitivity of participation to them hold irrespective to the length of the sample. The online Appendix B also repeats the probit analysis on subsamples by education and retirement status, as well as by focusing on a different dependent variables (Only Stocks or Mutual Funds). The robustness of the results for the probit analysis applies also to the unreported analysis for asset allocation, as well as to revisions in the individual decision to participate in the stock market over time.

5.2 Out-of-Sample Analysis of Participation

Results presented so far show that income hedging motives are more relevant to individual risk-taking decisions than previously thought. Moreover, they are able to explain participation decisions both in the cross-section and over time for each individual. Finally, results presented in the previous subsection also indicate that estimates of such hedging motives are precise, even given a limited time-series data dimension.

This section exploits this last feature to implement an out-of-sample analysis. In addition to providing additional evidence for the robustness of our results, this exercise represents the type of analysis that delegated portfolio managers or (robo-)advisors can use to assign investors to portfolios and revise such assignments. This out-of-sample analysis of participation relies on the 80 MD estimates of correlation between PI shocks and stock market returns at the cluster level using data up to 2011.

When considering data up to 2019, we allocate each new survey participant to one of the 80 clusters (defined in Section 3.1) on the basis of their personal characteristics. We then attribute to each individual the correlation parameter of the corresponding cluster estimated in the previous step using data up to 2011.

Then, we relate these correlation estimates to the decision to participate in the stock market in the years after 2011 (i.e., 2012–2019) and rank individuals from highest to lowest according to their correlation parameter, and we compute the participation rate for each quartile of the distribution.

Our results, reported in Figure 8, display a systematic pattern: individuals belonging to clusters with lower estimated correlation parameters participate more than individuals allocated to clusters displaying higher correlation. The difference between the top and bottom quartiles is remarkable given participation rates in 2012 and subsequent years.

As expected, this difference decreases when we step away from the time window used to estimate the correlation parameters. Nonetheless, when moving forward over time, it is possible to extend the estimation window to exploit the additional available information. Thus, the out-of-sample predictive power of the correlation parameter for the 2018 participation rate will be larger when we estimate the clustered correlations using data up to 2017.

A similar out-of-sample response of participation to correlation is obtained with a probit analysis. We regress a dummy variable, indicating participation in the stock market between 2012 and 2019, on the correlation between stock market returns and the PI shocks assigned to each individual agent according to their corresponding cluster and estimated using data up to 2011.

Again, stock market participation occurs through either mutual funds, direct investment in stocks, or both. Table 12 demonstrates that the clustered correlations, estimated using data up to 2011, negatively and significantly predict stock market participation for the period 2012–2019. The economic significance of our estimates is also remarkable. The results in column (1) suggest that an individual allocated to a cluster displaying high correlation (ρ=0.5) between stock market returns and PI shocks is 8% less likely to participate in the stock market than an individual allocated to a low correlation cluster (see equation (16) in Section 4.1).

[Figure 8 about here.]

[Table 12 about here.]

5.3 Income Shocks and Participation by Cohort and by Cluster

[Figure 9 about here.]

In this section, we visually inspect stock market participation and income shocks for different cohorts and different clusters. We expect participation by cohort to be more closely related to PI shocks than total shocks (TS). Moreover, we expect to see patterns that have been uncovered in the existing literature across clusters with different characteristics.

We start with a representation of the pattern of stock market participation choice by age for three cohorts that are relatively more numerous. Figure 9 displays participation patterns that broadly align with those reported by Fagereng et al. (2017). Then, we graphically contrast participation patterns and income shocks. Figure 10 plots average stock market participation by cohort against both the TS and the PI shocks estimated using our MD method. It appears that the pattern of participation responds to the latter while having a limited relationship with the former.

We then check that agents belonging to different clusters respond according to economic intuition. For instance, we compare two clusters with similar characteristics but risk aversion (Figure 11), finding that the age patterns for PI shocks are relatively similar for the two groups (see intermediate panel), while the TS are not (see top panel). In response to these similar PI shocks, the cluster with higher risk aversion participates less in the equities market (see bottom panel).

We perform a similar analysis across two identical clusters (including risk aversion) but for sex (Figure 12), finding that females participate less in the equities market (see bottom panel), a known result. It also appears that the volatility of PI shocks is greater for females than for males (see middle panel).

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

5.4 Correlation and Probit Estimates based on PSID

The PSID database is the workhorse data set for estimating earnings processes for U.S. individuals. For this reason, this section uses our method on PSID data from 1988 to 2011.

In PSID, we do not have information about risk aversion and financial literacy of individuals. Also, the sample is almost entirely composed by men. On the other hand, information about the industry of the household head's job is available and we still have information about the education level. Thus, to estimate the correlation parameters at the cluster-level using the Minimum Distance (MD) methodology as we do with the DHS data, we form 48 clusters based on 4 education groups and 12 industries as in Campbell and Viceira (2002).

We also estimate both β_{GUV} and β_{CV} at the cluster-level. We obtain the cluster-based β_{GUV} by estimating pooled OLS regression of the (log)-real earnings growth of individuals belonging to each cluster in year t over the stock market return in year t. To obtain β_{CV}, we first compute the average of the (log)-real earnings growth across the individuals belonging to each cluster, then we regress the cluster-specific permanent income shocks in year t over the stock market return in year $t+1$.

We then use these correlation parameters in a probit analysis of stock market participation. We report results from the probit regression in Table 14, where the dependent variable refers to ownership of equities or mutual funds. Results are comparable to the ones in Table 10-11, as the sample covers the same years and the main independent variables are the same. [12]. In the probit

[12] Income in PSID refers to the household, differently from DHS.

regression estimates on PSID data, the statistical significance of the coefficients associated with the MD correlation parameters is in line with the one obtained using the DHS data. Moreover, the statistical significance of the coefficients associated with alternative methods is lower, confirming results of the previous benchmarking exercise performed on the DHS. We therefore conclude that our method captures heterogeneous hedging motives also on PSID data.

We summarize the correlation coefficient estimates in Table 13. Importantly, the MD correlation coefficients estimated on PSID are very similar with those reported in Table 10 based on the Dutch Household Survey over the same years, when estimated at both individual and cluster levels. Other parameters estimated on PSID data tend to be larger in size and generally positive compared to the those estimated on the Dutch data. For instance, we estimate on PSID beta coefficients that are consistent with the values obtained by β_{GUV} and β_{CV}, respectively, using US data. [13]

[Table 13 about here.]

[Table 14 about here.]

6 Summary and Conclusions

This paper proposes a new approach to assess individual permanent income shocks and hedging motives. This centers on the measurement of cluster-based correlation between PI shocks and stock returns.

We recognize that individuals' labor income shocks co-move with each other due to their common albeit heterogeneous exposure to aggregate shocks. We therefore exploit the large cross-sectional dimension of the data to infer the distribution of the key parameter that was elusive in prior literature, namely the correlation between the observable stock market returns and the latent PI shocks.

Our estimates of the mean correlation coefficient between stock market returns and PI shocks is positive and in the range 0.2-0.3. This result is not sensitive to the length of the sample and appears both in the Dutch Household Survey and in the US PSID data. This mean correlation is higher than the one estimated in most prior research. It follows that the mean individual hedges labour income risk by reducing exposure to the equity market. This discovery implies that observed

[13] Let us note that when we use only three education clusters to compute β_{CV}, we obtain values that are very in line with those estimated by Campbell and Viceira (2002) on the three education groups.

portfolios are closer than previously thought to the implications of portfolio choice theory.

Indeed, not only the sign but also the size and the precision of the estimated effects confirm the theoretical prediction that risk-taking decisions respond to the ability of risky financial assets to hedge their earnings risk. To the extent that labor income shocks move together with equities market returns, individuals reduce their stock exposure. These results are robust, holding both in-sample and in out-of-sample experiments. They also hold both in the cross section and for each individual over time.

Earnings risks is a central issue in the economics of incomplete markets. These advances may therefore prove useful also beyond the boundaries of household finance.

References

Angerer, X., and P.S. Lam, 2009, Income risk and portfolio choice:an empirical study, Journal of Finance 64, 1037–1055.

Arrondel, Luc, Hector Calvo Pardo, and Xisco Oliver, 2010, Temperance in stock market participation: Evidence from france, Economica 77, 314–333.

Bagliano, F.C., C. Fugazza, and G. Nicodano, 2014, Optimal life-cycle portfolios for heterogeneous workers, Review of Finance 18, 2283–2323.

Benzoni, L., P. Collin-Dufresne, and R.G. Goldstein, 2007, Portfolio choice over the life-cycle when the stock and labour markets are cointegrated, Journal of Finance 62, 2123–2167.

Betermier, Sebastien, Thomas Jansson, Christine Parlour, and Johan Walden, 2012, Hedging labor income risk, Journal of Financial Economics 105, 622–639.

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. .

Carroll, Christopher D, 1997, Buffer-stock saving and the life cycle/permanent income hypothesis, The Quarterly journal of economics 112, 1–55.

Carroll, Christopher D, Robert E Hall, and Stephen P Zeldes, 1992, The buffer-stock theory of saving: Some macroeconomic evidence, Brookings papers on economic activity 1992, 61–156.

Carroll, C.S., and A. Samwick, 1997, The nature of precautionary wealth, Journal of Monetary Economics 40, 41–71.

Catherine, S., P. Sodini, and Y. Zhang, 2020, Countercyclical income risk and portfolio choices: Evidence from sweden, Swedish House of Finance Research Paper 20-20).

Chang, Y., J.H. Hong, and M. Karabarbounis, 2018, Labor market uncertainty and portfolio choice puzzles, American Economic Journal: Macroeconomics 10, 222–262.

Cocco, J., F. Gomes, and P. Maenhout, 2005, Consumption and portfolio choice over the life cycle, Review of Financial Studies 18, 491–533.

Dimmock, S.G., R. Kouwenberg, O.S. Mitchell, and K. Peijnenburg, 2016, Ambiguity aversion and household portfolio choice puzzles: Empirical evidence, Journal of Financial Economics 199, 559–577.

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.

Fagereng, A., L. Guiso, and L. Pistaferri, 2018, Portfolio choices, firm shocks, and uninsurable wage risk, The Review of Economic Studies 85, 437–474.

Gomes, F. J., M. Haliassos, and T. Ramadorai, 2020, Household finance, Journal of Economic Literature forthcoming.

Guiso, Jappelli T., L., and D. Terlizzese, 1996, Income risk, borrowing constraints, and portfolio choice, The American Economic Review 86(1), 158–172.

Guiso, L., P. Sapienza, and L. Zingales, 2008, Trusting the stock market, The Journal of Finance 63, 2557–2600.

Guvenen, F., 2009, Empirical investigation of labor income process, Review of Economic Dynamics 12.

Guvenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo, 2017, Worker betas: Five facts about systematic earnings risk, American Economic Review Papers and Proceedings 107, 398–403.

Haliassos, M., and A. Michaelides, 2003, Portfolio choice and liquidity constraints, International Economic Review 44, 144–177.

Heaton, J., and D. Lucas, 2000, Portfolio choice and asset prices: The importance of entrepreneurial risk, Journal of Finance 55, 1163–1198.

Heckman, James J, 1979, Sample selection bias as a specification error, Econometrica: Journal of the econometric society 153–161.

Low, Hamish, Costas Meghir, and Luigi Pistaferri, 2010, Wage risk and employment risk over the life cycle, American Economic Review 100, 1432–67.

Massa, M., and A. Simonov, 2006, Hedging, familiarity and portfolio choice, The Review of Financial Studies 19, 633–685.

Meghir, Costas, and Luigi Pistaferri, 2004, Income variance dynamics and heterogeneity, Econometrica 72, 1–32.

Merton, R., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, Review of Economics and Statistics 51, 247–257.

Michaelides, A., and Y. Zhang, 2017, Stock market mean reversion and portfolio choice over the life cycle, Journal of Financial and Quantitative Analysis 52(3), 1183–1209.

Mitchell, O.S., and S.P. Utkus, 2020, Target date funds and portfolio choice in 401(k) plans, Working Paper Pension Research Council .

Munk, C., and C. Sorensen, 2010, Dynamic asset allocation with stochastic income and interest rates, Journal of Financial Economics 96, 433–462.

Van Rooij, M., A. Lusardi, and R. Alessie, 2011, Financial literacy and stock market participation, Journal of Financial Economics 101, 449–472.

Viceira, L.M., 2001, Optimal portfolio choice for long-horizon investors with nontradable labor income, Journal of Finance 41, 433–470.

Vissing-Jorgensen, A., 2002, Towards an explanation of household port- folio choice heterogeneity: nonfinancial income and participation cost structures, National Bureau of Economic Research .

Appendix A Individual Correlations and Clustering Variables

This Appendix reports the MD estimates of the correlation parameters at the individual level. It then presents a simple regression of individual correlations onto individual traits.

The left panel of Figure 13 plots the distribution of the estimated individual correlation coefficients between the PI shocks and the stock market returns $\{\hat{\rho}_i\}_{i=1}^N$ against the empirical distribution of the sample's individual correlations between TS and stock market returns. Meanwhile, the right panel compares the empirical distribution of the estimated individual correlation coefficients $\{\hat{\rho}_i\}_{i=1}^N$ against the individual correlations between PI shocks and stock market returns estimated using the methodology described in Bonaparte et al. (2014).

[Figure 13 about here.]

Then, we check whether the clustering variables correlate with the individual correlations between PI shocks and stock market returns. We run a cross-sectional ordinary least squares regression in which the individual correlations obtained using the minimum distance estimation are the dependent variables, and the clustering variables are the independent variables. We report the regression coefficients in Table 15. All of the clustering variables have explanatory power. Moreover, the individual correlation between PI shocks and stock market returns is higher when the individual is male, less risk-averse, more financially educated but with a lower level of general education, and living in an urban area.

[Table 15 about here.]

Appendix B Summary Statistics and Other Analysis with DNB (1993-2011)

This Appendix B reports summary statistics for the sample up to 2011. It then repeats the probit analysis on subsamples by education and retirement status, as well as by focusing on a different dependent variables (Only Stocks or Mutual Funds).

[Table 16 about here.]

[Table 17 about here.]

[Table 18 about here.]

Appendix C Kalman Filter

This Appendix explains how to implement the Kalman filter (KF) to retrieve the dynamics of the unobserved components of labor income shocks.

Considering the state-space model described by the equations (17) and (18), we reconstruct the dynamics of the unobservable random walk $v_{i,t}$ for each individual by implementing a linear KF, using the regression residuals as the observable variable and estimating the parameters using the minimum distance approach. Briefly, the KF exploits the assumed relationship between the observed and unobserved variables to infer the dynamics of the latter. This relationship forms the *measurement equation*, given here by the equation (17) in the text, with the equation (18) forming the *transition equation*, which describes the evolution of the latent variable over time.

We initialize the filter applying two arbitrary conditions to both the initial value of the latent variable and its variance:

$$v_{i,0} \quad P_{i,0}$$

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

$$E_0[v_{i,1}] = v_{i,0},$$

$$E_0[P_{i,1}] = P_{i,0} + Q,$$

where the first prediction derives from the transition equation, and Q is the variance of the transition equation, given here by σ_u^2. Then, we use the measurement equation to make a forecast about the observed variable and compare our forecast with the actual observation to obtain a measurement error[14].

$$h_{i,1} = e_{i,1} - E_0[e_1] = e_{i,1} - v_{i,0}$$

We consider the measurement error to update the estimate of $v_{i,1}$, and $P_{i,1}$:

[14] Based on information at $t = 0$, the expected value of $u_{i,t}$ and $\epsilon_{i,t}$ is zero.

$$\hat{v}_{i,1} = v_{i,0} + K * h_1 \hat{P}_{i,1} = (1 - K) * E_0[P_{i,1}]$$

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

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

Here, σ_ϵ^2 plays the role of measurement error variance, that is the inverse of the reliability of the newly available observation for improving estimation of the latent variable. These steps are repeated recursively over the entire time series, providing an estimate of the latent variable dynamics. As a final step, we reconstruct the dynamics of the permanent stochastic component of the log-income process, for each individual, over the sample time series.

Appendix D Online Appendix: The Life-Cycle Model

This appendix shows the sensitivity of optimal portfolios to correlation estimates, based on the calibration of a life-cycle model. The investor maximizes the expected discounted utility of consumption over her entire life and also wishes to leave a bequest (Cocco et al. (2005)). The effective length of her life, which lasts a maximum period of T, is governed by her age-dependent life expectancy. At each date t, the survival probability of being alive at date is $t+1$ is p_t (i.e., the conditional survival probability at t). The investor starts working at age t_0 and retires with certainty at age t_0+K. The investor's i preferences at date t are described by a time-separable power utility function:

$$U_{t_0} = E_{t_0}\left[\frac{C_{t_0}^{1-\gamma}}{1-\gamma} + \sum_{j=1} \beta^j \left(\prod_{k=-1}^{j-2} p_{t_0+k}\right)\left(p_{t_0+j-1}\frac{C_{t_0+j}^{1-\gamma}}{1-\gamma} + (1-p_{t_0+j-1})\, b\frac{(X_{t_0+j}/b)^{1-\gamma}}{1-\gamma}\right)\right] \quad \text{(A.1)}$$

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 the event of death, $b \geqslant 0$ is a parameter capturing the strength of the bequest motive, $\beta < 1$ is a utility discount factor, and γ is the constant relative risk aversion parameter.

D.1 Labor and Retirement Income

Available resources to finance consumption over the investor's life cycle derive from accumulated financial wealth and from her labor income stream. At each date t during the investor's *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 impeded by both a permanent shock u_{it} and a transitory disturbance ϵ_{it}, according to the process outlined in Section 1 of the main text.

During *retirement*, income is certain and equal to a fixed proportion λ of the permanent component of income during the investor's last working year:

$$\log Y_{it} = \log\lambda + f\left(t_{0+K}, \mathbf{Z}_{it_{0+K}}\right) + u_{it_{0+K}} \qquad t_0+K < t \leqslant T \quad \text{(A.2)}$$

where the level of the replacement rate λ is designed to capture at least some of the features of welfare systems.

D.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 a risky asset characterized as "stocks" yielding a stochastic gross real returns R_t^s. We maintain that investment opportunities do not vary over time, thus modeling excess stock returns over the riskless asset as:

$$R_t^s - R^f = \mu^s + \nu_t^s \tag{A.3}$$

where μ^s is the expected stock premium, and ν_t^s is the normally distributed innovation, with mean zero and variance σ_s^2.

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

$$X_{it+1} = (X_{it} - C_{it})R_{it}^P + Y_{it+1} \tag{A.4}$$

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

$$R_{it}^P = \alpha_{it}^s R_t^s + (1 - \alpha_{it}^s)\, R^f \tag{A.5}$$

where α_{it}^s and $(1 - \alpha_{it}^s)$ respectively denote the proportion of the investor's portfolio invested in stocks and the proportion invested in the riskless asset.

D.3 Solving the Life-Cycle Problem

According to this standard intertemporal optimization framework, the investor maximizes the expected discounted utility over her lifetime by making consumption and portfolio decisions according to her uncertain labor income and asset returns. Formally, the optimization problem is written as:

$$\max_{\{C_{it}\}_{t_0}^{T-1},\{\alpha_{it}^s,\alpha_{it}^b\}_{t_0}^{T-1}} \left(\frac{C_{it_0}^{1-\gamma}}{1-\gamma} + E_{t_0} \left[\sum_{j=1}^{T} \beta^j \left(\prod_{k=0}^{j-2} p_{t_0+k} \right) \left(p_{t_0+j} \frac{C_{it_0+j}^{1-\gamma}}{1-\gamma} + \right.\right.\right.$$
$$\left.\left.\left. + \left(1-p_{t_0+j}\right) b \frac{\left(X_{it_0+j}/b\right)^{1-\gamma}}{1-\gamma} \right) \right] \right) \tag{A.6}$$

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

where the labor income and retirement processes specified are imposed, along with short sales and borrowing constraints.

Given its intertemporal nature, the problem is restated recursively, 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):

$$V_{it}\left(X_{it}, u_{it}\right) = \max_{C_{it},\alpha_{it}^s} \left(\frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta E_t \left[p_t V_{it+1}\left(X_{it_{+1}}, u_{it+1}\right) \right.\right.$$
$$\left.\left. + \left(1-p_t\right) b \frac{\left(X_{it+1}/b\right)^{1-\gamma}}{1-\gamma} \right] \right) \tag{A.7}$$

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 the beginning of time t, u_{it}. 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 using the permanent component of income u_{it}. Accordingly, we can rewrite (A.7) as:

$$V_{it}\left(X_{it}\right) = \max_{C_{it},\alpha^s_{it}} \left(\frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta E_t \left[p_t V_{it+1}\left(X_{it_{+1}}\right) \right.\right.$$
$$\left.\left. + \left(1-p_t\right) b \frac{\left(X_{it+1}/b\right)^{1-\gamma}}{1-\gamma} \right] \right) \tag{A.8}$$

Given the problem features no closed-form solution, the optimal values for consumption and portfolio share at each time point are obtained by means of standard numerical techniques.

D.4 Calibration

This section calibrates the standard life-cycle model for consumption and portfolio decisions. The values of the calibrated parameters are reported in Table 19.

[Table 19 about here.]

Additionally, we calibrate the relevant parameters of the labor income process against estimates derived from the Dutch National bank Household Survey and obtained in the main text, notably the variance of permanent and transitory shocks ($\sigma_u = 0.008$ and $\sigma_\epsilon = 0.095$, respectively).

Figure 14 reports the correlations between the permanent income component and stock market returns as estimated for the 80 clusters of individuals defined in the main text[15].

[Figure 14 about here.]

The estimated individual correlations range from -0.9 to 0.9, with most values concentrated in the range between 0.1 and 0.72 and the median value calculated as equal to 0.31.

D.5 Results

Using standard numerical techniques, we solve the model considering the parameter values presented in Table 19 and the estimated correlations. For the representative investors for each cluster, we obtain the optimal stock share over their working life.

Figure 15 plots the optimal conditional stock share over the investor's working life where risk aversion is 5. The traditional glide path adopted by Target Date Funds is optimal when stock

[15] Clusters are defined on the basis of education, sex, level of urbanization of the household's residence, risk aversion, and financial literacy.

market returns and permanent income shocks are uncorrelated. When the correlation is zero, the optimal share invested in stocks until the age of 30 is 100%; it gradually decreases to reach 40% at retirement. In the event of correlation equal to -0.5, the optimal share invested in stocks until the age of 40 is 100%; it then decreases more slowly, reaching 60% at retirement. In the event of correlation equal to 0.5%, the optimal share invested in stocks is 47% at the beginning of the investor's working life; it decreases to reach 25% at retirement.

(b) Correlation 0.1 **-** 0.9

[Figure 15 about here.]

[Figure 16 about here.]

Figure 16 plots the optimal conditional stock share over an investor's working life when risk aversion is 8. The traditional glide path adopted by Target Date Funds is optimal when stock market return and permanent income shocks are uncorrelated. When the correlation is zero the optimal share invested in stocks until the age of 35 is 100%; it gradually decreases to reach 20% at retirement. In the event of correlation equal to -0.5, the optimal share invested in stocks until the age of 40 is 100%; it then decreases more slowly, reaching 40% at retirement. In the event of correlation equal to 0.5%, the optimal stock share is zero until the age of 28; it increases to reach 11% at retirement.

[Figure 17 about here.]

[Figure 18 about here.]

Our results demonstrate that limited stock market participation can arise endogenously when calibrations are based on estimated parameters of the joint distribution of labor income and asset returns. Moreover, our results imply that a wide range of investment strategies should be offered according to the plan participant's risk aversion and specific labor income characteristics.

Figure 1. Number of individuals in each cluster
The figure displays the number of individuals belonging to each of the 80 clusters formulated according to the personal traits described in the paper. Clusters are based on education (5 groups), Sex (2), urbanization (2), risk aversion (2), and financial literacy (2).

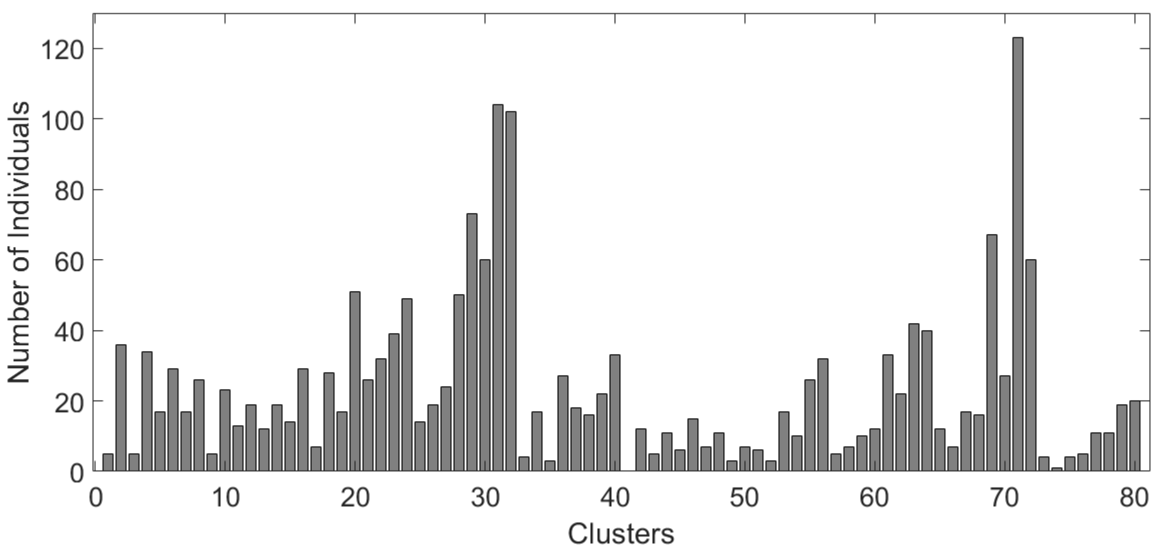


Figure 2. Distribution of cluster-based correlations after clustering: Full sample
The figure's left panel compares—after clustering the individuals in 80 homogeneous groups—the distribution of correlations between stock market returns and permanent labor income shocks estimated through a minimum distance method (MD) with the distribution of sample correlations between stock market returns and total income shocks (Total). The right panel compares the MD distribution with the distribution of correlations between stock market returns and permanent income shocks estimated according to Bonaparte et al. (2014)(BKK). The data are from the DNB Household Survey and cover waves for the period 1993–2019.

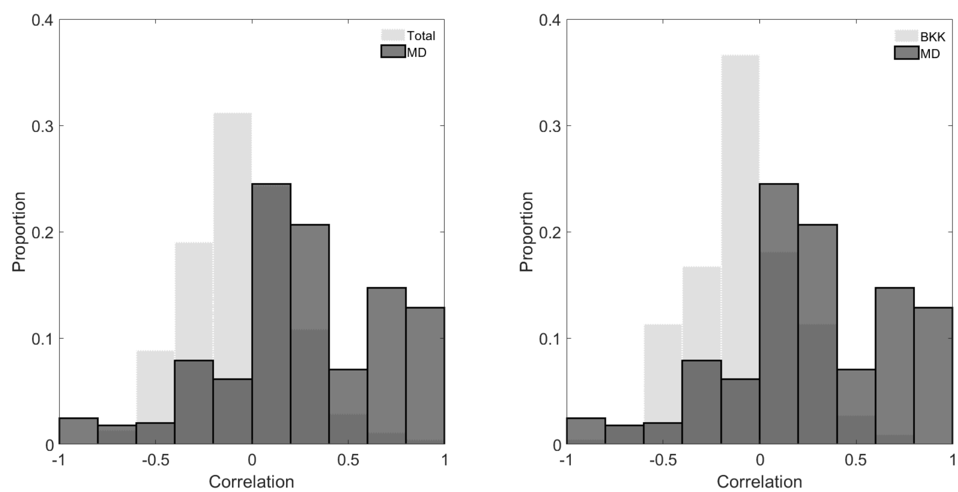


Figure 3. Stock market participation rates for low (grey bar) and high (black bar) correlation subsamples

The figure reports the average stock market participation rates for the low (grey bar) and high (black bar) correlation subsamples. We consider the correlation between stock market returns and different components describing labor income shocks: the total labor income growth rate (Tot), the stochastic component of labor income growth rate (Stoc), the deterministic component of labor income growth rate (Det), the transitory and the permanent shocks estimated according to Bonaparte et al. (2014) (BKK_{Tran} and BKK_{Perm}, respectively), and the permanent component of labor income shocks estimated using the minimum distance method at the cluster-level (MD). Low (high) is defined as the bottom (top) quartile of correlation between income growth and market returns. The data are from the DNB Household Survey and cover all waves for the period 1993–2019.

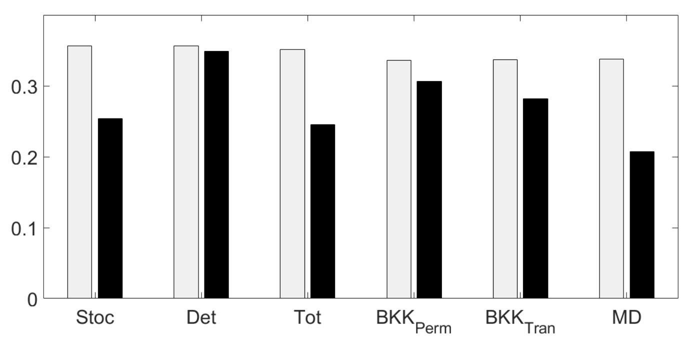


Figure 4. Predicted and actual participation
The figure compares the actual level of participation and the predicted level of participation, for each year, obtained using the probit regression estimates reported in Table 4 (column (6)). To obtain the predicted level of participation for each year of our sample, we compute the individual probability to participate to the stock market using equation (16) and regression estimates reported in Table 4 (column (6)). Then, we compute the average probability across individuals for each year of our sample. The data are from the DNB Household Survey and cover all waves for the period 1993–2019.

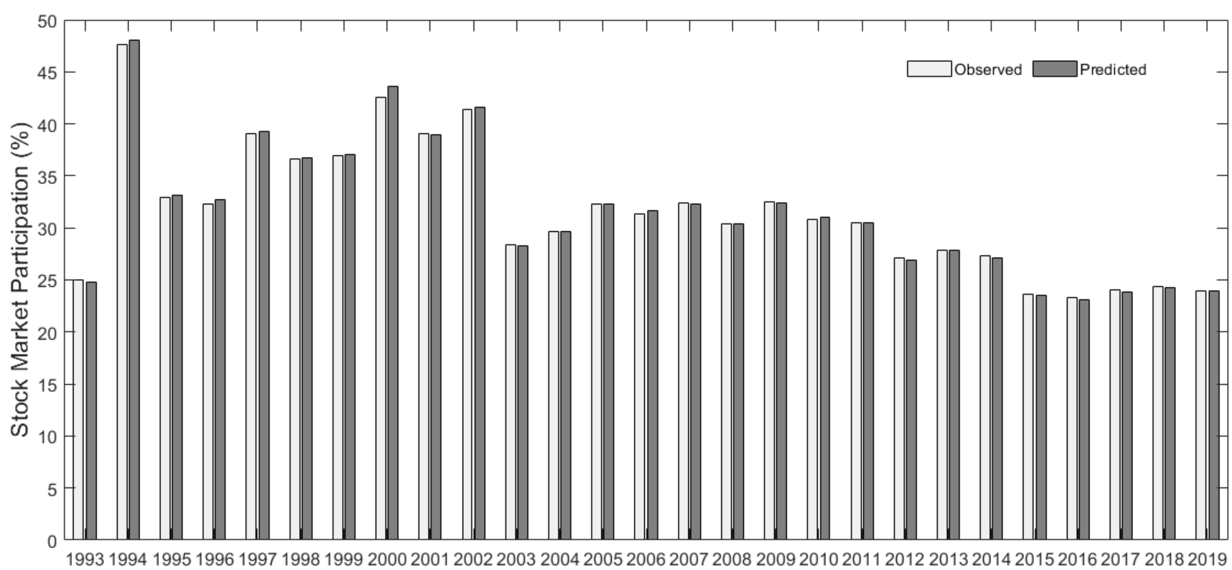


Figure 5. Individual permanent income shocks
The left panel compares the dynamics of stock market returns and permanent labor income (PI) shocks reconstructed using the Kalman filter for the individual with the maximum sample correlation between stock market returns and PI shocks. The right panel compares the dynamics of the stock market returns with the PI shocks reconstructed using the Kalman filter for the individual with minimum sample correlation between stock market returns and PI shocks. The data are from the DNB Household Survey and cover all waves for the period 1993–2019.

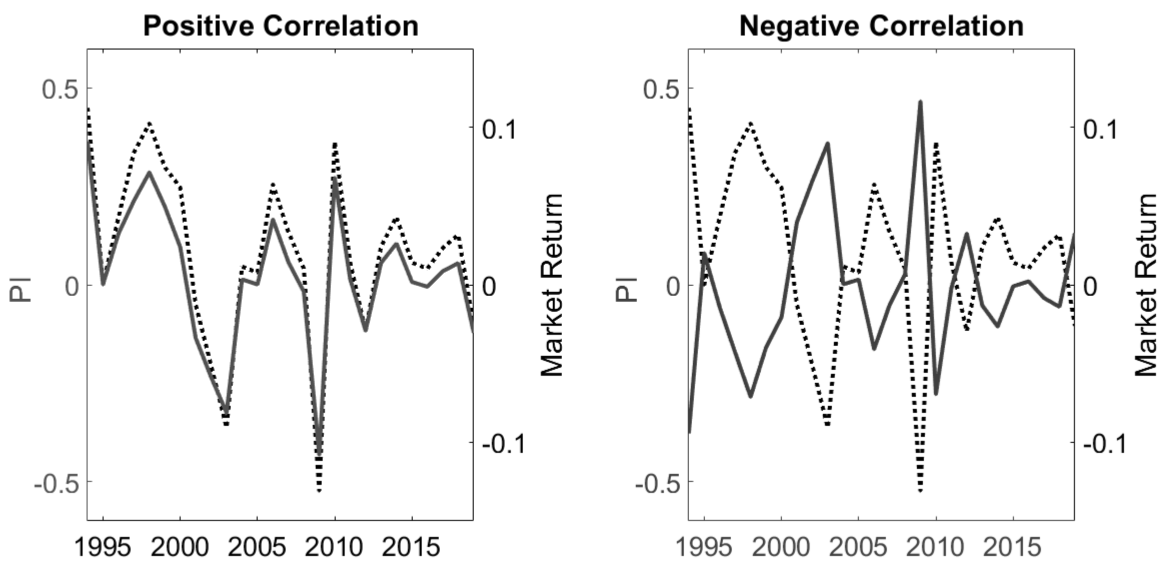


Figure 6. Distribution of individual correlations using Kalman filter
The left panel compares the distribution of sample correlations between stock market returns and the permanent labor income (PI) shocks reconstructed using the Kalman filter (KF) with the distribution of sample correlations between stock market returns and total income shocks (Total). The right panel compares the distribution of sample correlations between stock market returns and PI shocks reconstructed using the KF with the distribution of the correlations between stock market returns and PI shocks estimated using the minimum distance methodology (MD). The data are from the DNB Household Survey and cover all waves for the period 1993–2019.

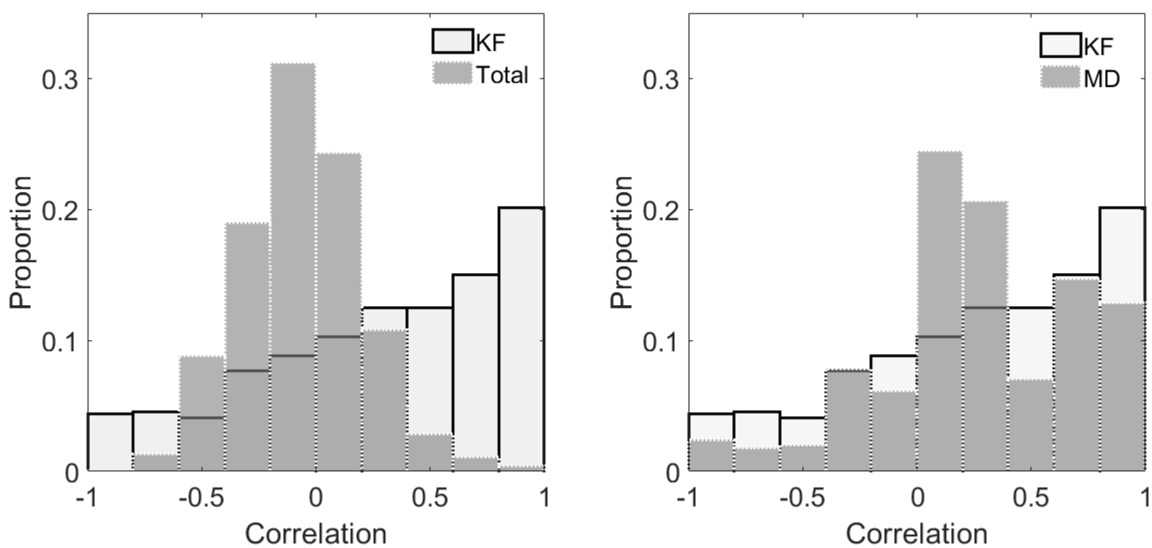


Figure 7. Distribution of cluster-based correlations
The figure's left panel compares—after clustering the individuals in 80 homogeneous groups—the distribution of correlations between stock market returns and permanent labor income shocks estimated through a minimum distance method (MD) with the distribution of sample correlations between stock market returns and total income shocks (Total). The right panel compares the MD distribution with the distribution of correlations between stock market returns and permanent income shocks estimated according to Bonaparte et al. (2014)(BKK). The data are from the DNB Household Survey and cover waves for the period 1993–2011.

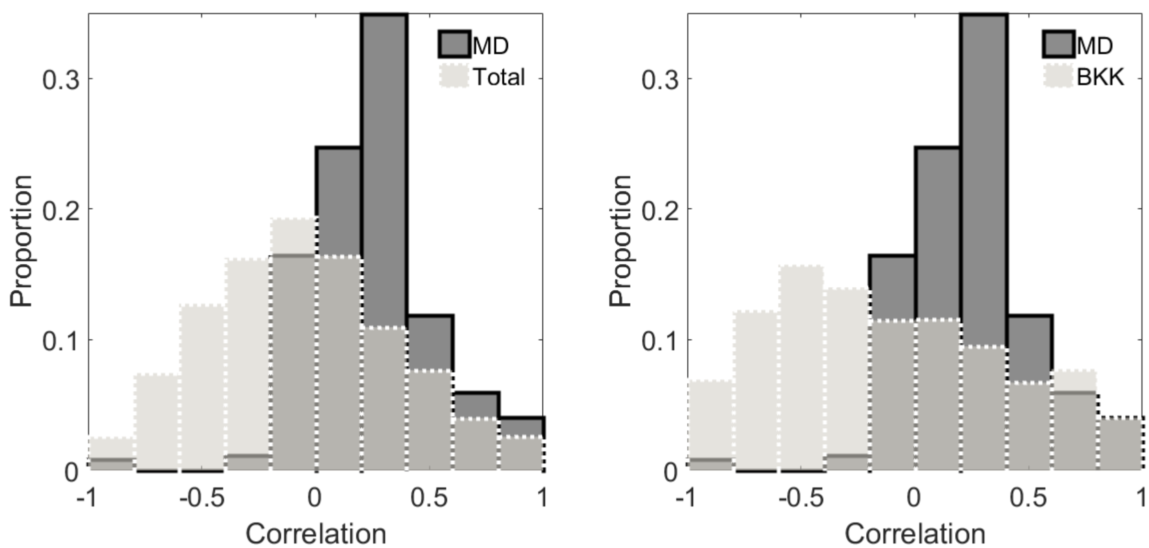


Figure 8. Out-of-Sample prediction of stock market participation
The figure reports the stock market participation rates for the 2012–2019 waves for the low (grey bar) and high (black bar) correlation subsamples. We estimate the correlation between stock market returns and PI shocks using the minimum distance method at the cluster level based on data up to 2011. Then, we allocate individuals to clusters according to their observable characteristics and assign each individual a correlation parameter on the basis of the cluster to which they belong. Low (high) is defined as the bottom (top) quartile of correlation between income growth and stock market returns. The data are from the DNB Household Survey and cover all waves for the period 1993–2019.

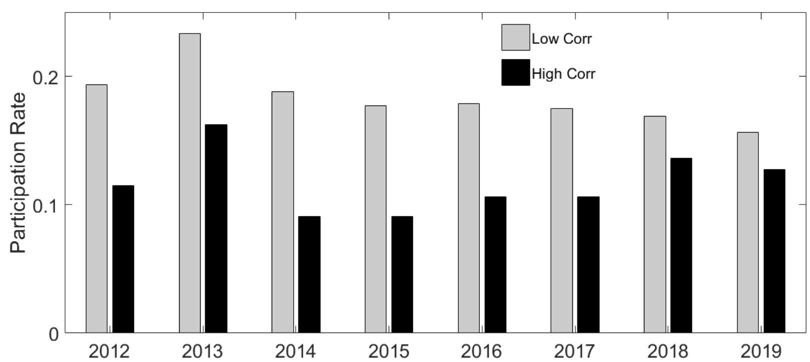


Figure 9. Participation by cohort
The figure shows the stock market participation rate for each cohort over time. Stock market participation is a dummy variable equal to 1 if an individual invests in stocks, either directly or through mutual funds, and zero otherwise. We compute the participation rate for each cohort as the mean of the dummy variable across the individuals belonging to the cohort for each year in the sample. The x-axis reports the variable *Age* obtained as (Year - Cohort). We report results for the three most populated cohorts in our sample (1939-1946-1953).

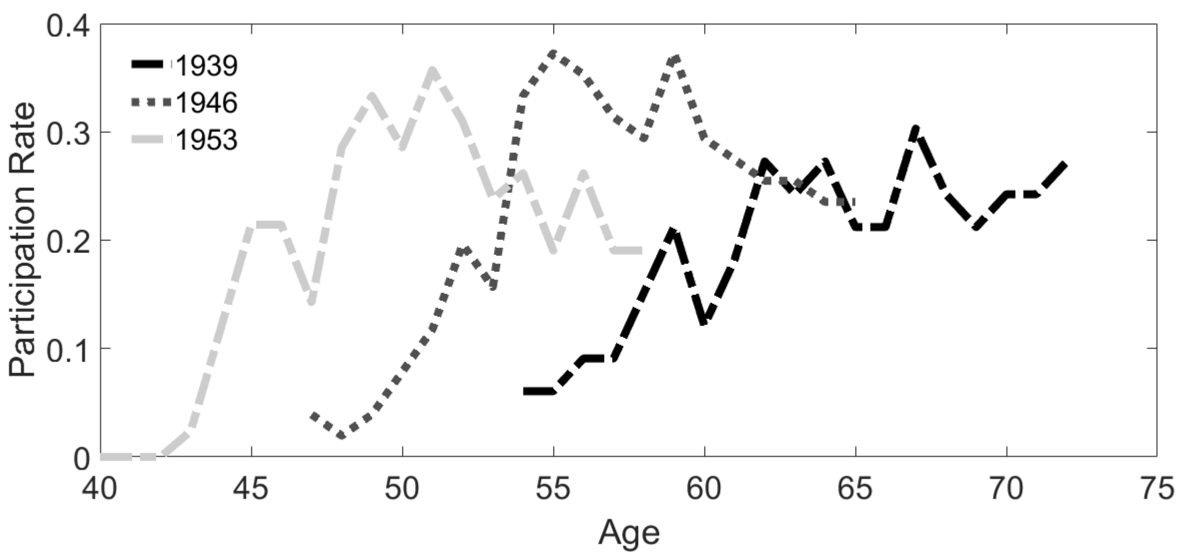


Figure 10. Income shocks and participation by cohort
The figure shows total income shocks (TS, top panel), permanent income shocks (PI, middle panel), and stock market participation rate (Participation, bottom panel) for the three most populated cohorts in our sample: 1939 (black dashed line), 1946 (grey dotted line), and 1953 (grey dashed line). Income shocks are taken as absolute values. For each cohort, we compute the average TS, PI, and Participation rate for individuals belonging to the cohort for each year. Stock market participation is a dummy variable equal to 1 if an individual invests in stocks, either directly or through mutual funds, and zero otherwise. We compute the participation rate for each cohort as the mean of the participation dummy for the individuals belonging to the cohort for each year in the sample.

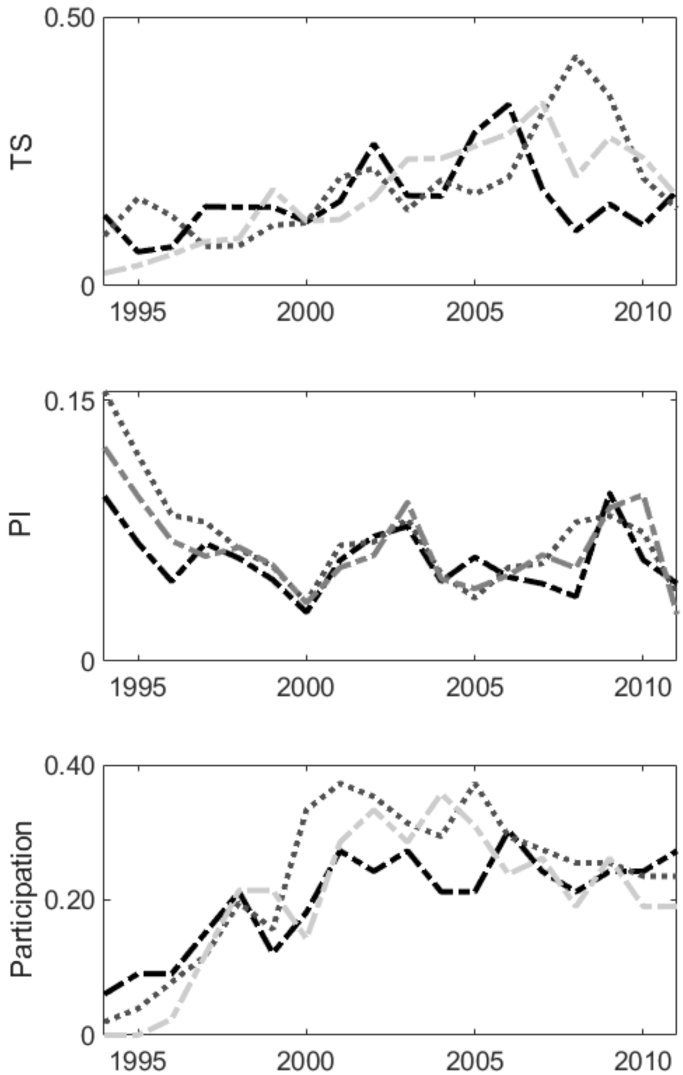


Figure 11. Income shocks and participation by cluster: Risk aversion
The figure shows total income shocks (TS, top panel), permanent income shocks (PI, middle panel), and stock market participation rate (Participation, bottom panel) for the three most populated clusters in our sample. The clusters are similar for all characteristics except risk aversion. The grey dotted line displays patterns for the risk-averse cluster. Income shocks are taken as absolute values. For each cluster, we compute the average TS, PI, and Participation rate for individuals belonging to the cluster for each year. Stock market participation is a dummy variable equal to 1 if an individual invests in stocks, either directly or through mutual funds, and zero otherwise. We compute the participation rate for each cluster as the mean of the participation dummy for the individuals belonging to the cluster for each year in the sample.

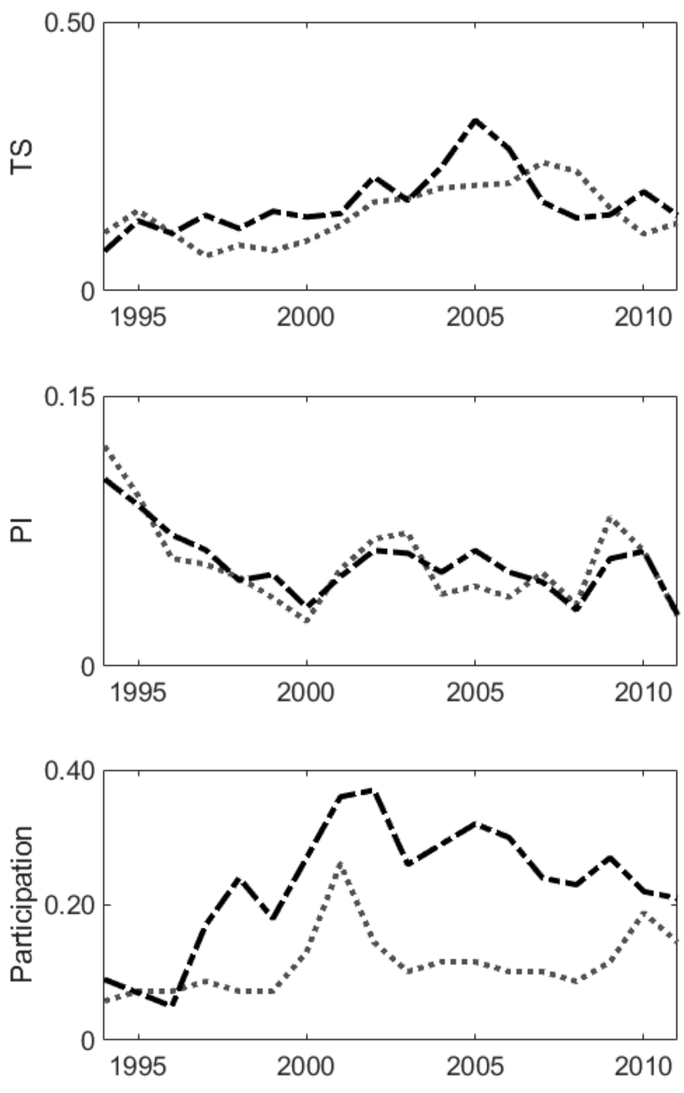


Figure 12. Income shocks and participation by cluster: Sex
The figure shows total income shocks (TS, top panel), permanent income shocks (PI, middle panel), and stock market participation rate (Participation, bottom panel) for the most populated cluster in our sample, including only males (black dashed line), and the cluster with similar characteristics but including only females (grey dotted line). Income shocks are taken as absolute values. For each cohort, we compute the average TS, PI, and Participation rate for individuals belonging to the cluster for each year. Stock market participation is a dummy variable equal to 1 if an individual invests in stocks, either directly or through mutual funds, and zero otherwise. We compute the average participation rate for each cluster as the mean of the participation dummy for the individuals belonging to the cluster for each year in the sample.

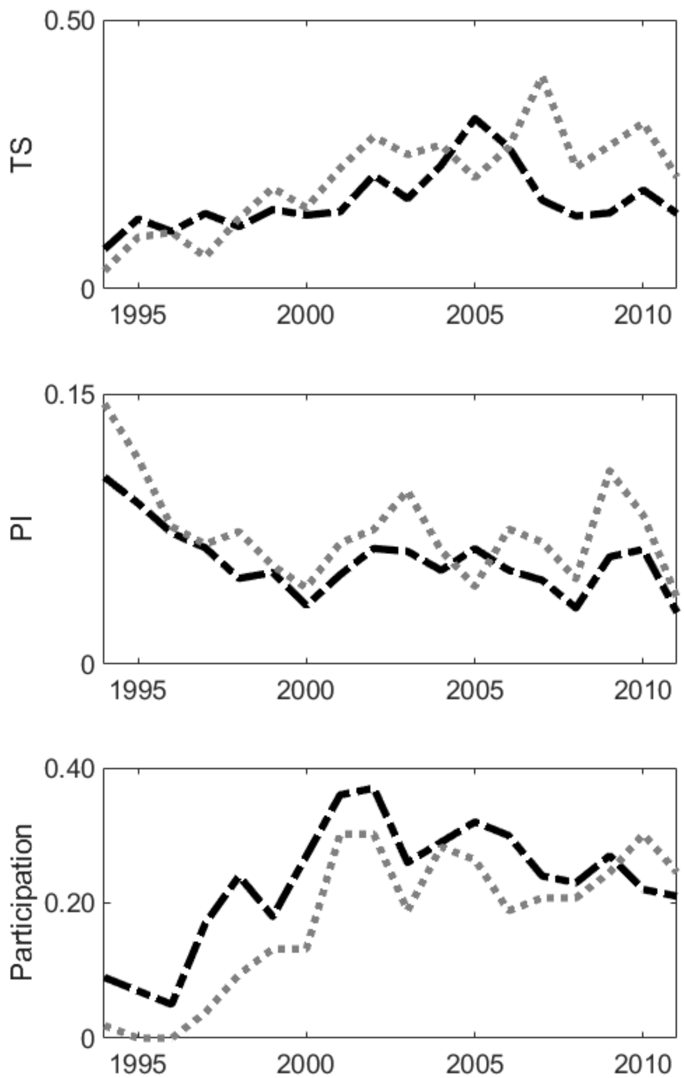


Figure 13. Distribution of individual correlations
The figure's left panel compares the distribution of correlations between stock market returns and permanent labor income shocks estimated through a minimum distance method (MD) at the individual level with the distribution of sample correlations between stock market returns and total income shocks (Total). The right panel compares the MD distribution with the distribution of correlations between stock market returns and permanent income shocks estimated according to Bonaparte et al. (2014)(BKK). The data are from the DNB Household Survey and cover waves for the period 1993–2011.

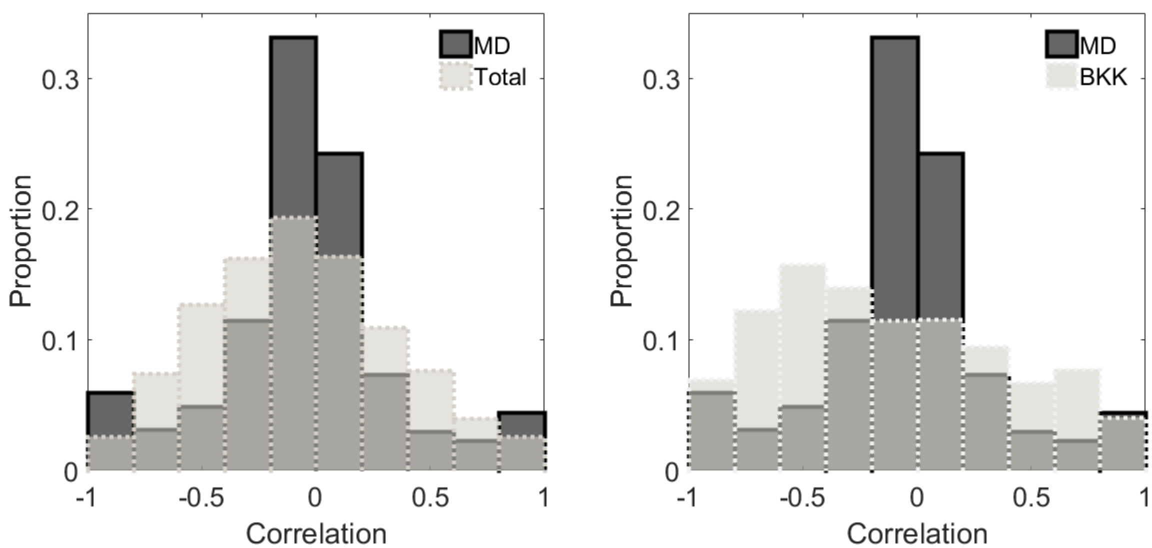


Figure 14. Distribution of individual correlations by clusters
The figure reports the distribution of correlations between stock market returns and permanent labor income shocks as estimated after clustering the individuals in the 80 homogenous groups.

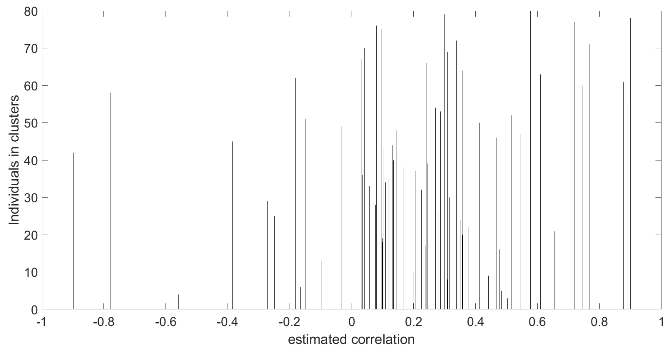
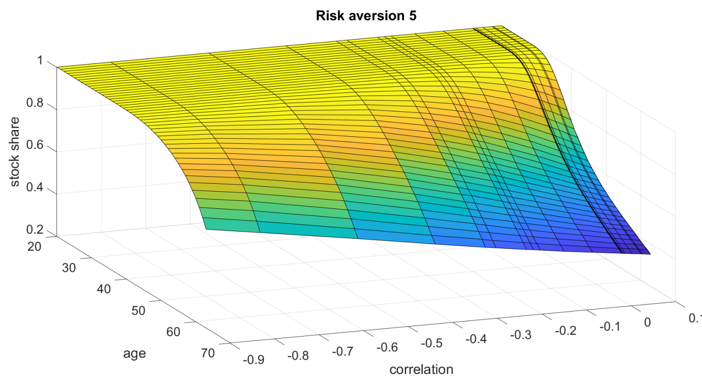


Figure 15. Optimal stock share: Risk aversion of 5

(a) Correlation $-0.9 - 0.1$



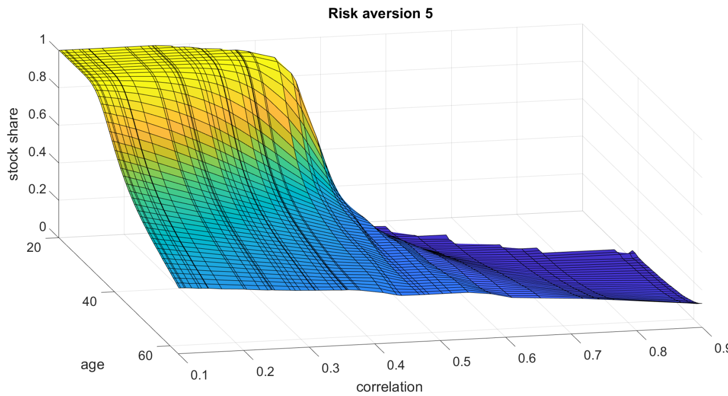
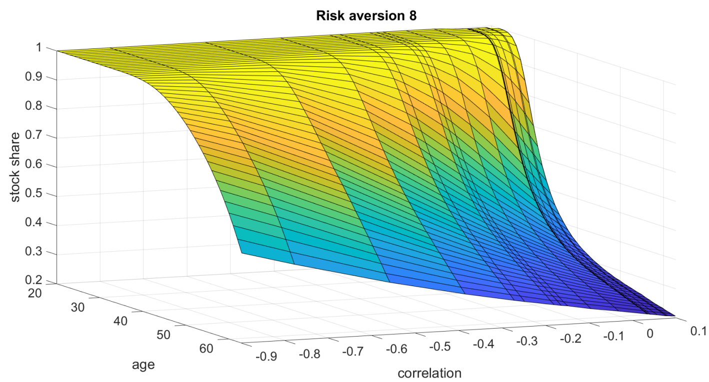

Risk aversion 5
stock share
0
0.2
0.4
0.6
0.8
1
20
40
60
age
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
correlation

Figure 16. Optimal stock share - risk aversion 8

(a) Correlation -0.9 **-** 0.1



(b) Correlation 0.1 **-** 0.9

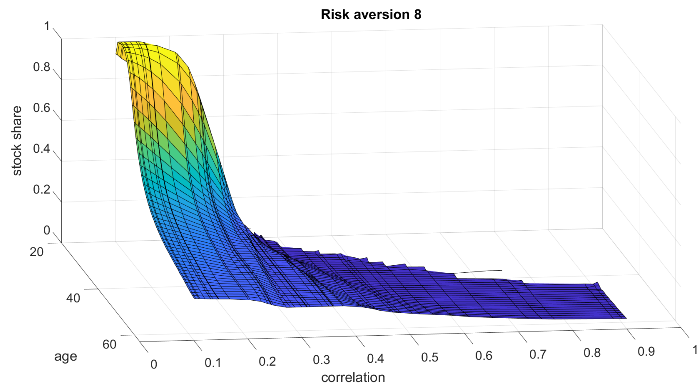


Table 1 Definitions of variables

Variable	Definition
OwnSTK	One if own stocks and zero otherwise.
OwnMF	One if own mutual funds and zero.
OwnSTKMF	One if own stocks or mutual funds and zero otherwise.
PropSTK	Financial wealth fraction invested in stocks.
PropMF	Financial wealth fraction invested in mutual funds.
PropSTKMF	Financial wealth fraction invested in stocks or mutual funds.
Ln(NetWorth)	Log of net worth.
Ln(NetIncome)	Log of net income.
Corr(d(lnInc),Rm)	Correlation between income growth rate and Dutch stock market returns.
SD(lnInc)	Standard deviation of income growth rate.
HH	Household size.
Age	Years old.
Education	One if college graduate and zero otherwise.
Male	One if male and zero otherwise.
Unemployed	One if unemployed and zero otherwise.
Retired	One if retired and zero otherwise.
Health	Health rating (1-5) with 5 being good.
Fin. Literacy	One if knowledgeable about financial assets.
Risk aversion	Perception of risk (rating from 1 to 7) where 7 is belief that investing in stocks is very risky.

Table 2 Summary statistics. Full Sample

This table reports the summary statistics for the variables used for the empirical analysis. The data are from the Dutch National Bank Household Survey and cover all waves for the period 1993–2019. N denotes the total number of observations, n indicates the number of individuals, and T represents the average number of years in which those individuals participated in the survey. Definitions of the variables are provided in Table 1.

Variable	Mean	Standard Deviation	p10	Median	p90	N (n x T)
						27445
OwnSTK	0.05	0.23	0	0	0	27445
OwnMF	0.17	0.37	0	0	1	27445
OwnSTKMF	0.32	0.47	0	0	1	22275
PropSTK	0.02	0.12	0	0	0	22275
PropMF	0.06	0.18	0	0	0.22	22275
PropSTKMF	0.09	0.22	0	0	0.40	22275
Ln(NetWorth)	11.71	1.71	9.05	12.30	13.16	16695
Ln(NetIncome)	9.79	0.87	8.84	9.98	10.56	21084
Corr(d(LnInc),Rm)	-0.06	0.27	-0.40	-0.06	0.26	12204
SD(d(LnInc))	0.41	0.49	0.07	0.23	1.03	50868
HH size	2.38	1.20	1	2	4	27403
Age	55.82	14.19	36	56	74	27401
Education	0.51	0.49	0	1	1	27401
Male	0.61	0.49	0	1	1	27401
Unemployed	0.14	0.35	0	0	1	27427
Retired	0.13	0.33	0	0	1	27401
Health	3.87	0.70	3	4	5	23911
Fin. Literacy	0.36	0.48	0	0	1	27443
Risk Aversion	4.57	2.08	1	5	7	22945

Table 3 Correlation Parameters from DHS. Summary Statistics on Full Sample. The table reports summary statistics of the estimated correlations between stock market return and different specifications of labor income shocks: the total income shocks, the stochastic component, the deterministic component, the permanent component estimated by using the minimum distance methodology (MD) both at the individual and the cluster levels, the permanent component obtained using the Kalman Filter and computing the correlation at T (Ex-post KF), the permanent component obtained using the Kalman Filter and updating the correlation at each t (Revised KF). We also report summary statistics about different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (Permanent (BKK) and Transitory (BKK), respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. In the last column, we report the t-test for the statistical significance of the parameter and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DNB Household Survey and cover waves from 1993–2019.

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Total	-0.062	0.269	-0.402	-0.065	0.260	-4.905***
Deterministic	-0.015	0.244	-0.325	-0.025	0.318	-2.006 **
Stochastic	-0.058	0.269	-0.404	-0.065	0.264	-4.566***
Permanent (MD cluster)	0.257	0.436	-0.379	0.248	0.868	25.611***
Permanent (MD individual)	0.057	0.502	-0.804	0.049	0.899	4.913***
Permanent (Ex-Post KF)	0.276	0.532	-0.544	0.365	0.899	22.522***
Permanent (Revised KF)	0.303	0.583	-0.643	0.459	0.944	18.459***
Permanent (BKK)	-0.091	0.269	-0.455	-0.105	0.276	-5.044***
Transitory (BKK)	-0.021	0.260	-0.359	-0.010	0.296	-1.191
Beta (Guvenen et al. (2017))	0.004	0.021	-0.015	0.003	0.031	8.915***
Beta (C&V(2002))	0.006	0.028	-0.021	0.005	0.028	10.100***

Table 4 Probit Estimates for Stock Market Participation
The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation and the correlation between stock market return and different specifications of labor income shocks: the labor income growth rate (Tot), the stochastic component of labor income growth rate (Stoc), the deterministic component of labor income growth rate (Det), the permanent component of labor income shocks estimated by using the minimum distance methodology (permMD80). Additional control variables are individual demographic characteristics detailed in table 1. We also control for year-fixed effects. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively.The data are from the DNB Household Survey and cover waves from 1993–2019.

	OWNSTKMF					
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.274***	-0.281***	-0.147**	-0.251***	-0.259***	-0.153***
	(0.043)	(0.043)	(0.017)	(0.075)	(0.059)	(0.038)
Corr(Tot,Rm)	-0.273***			-0.549***		
	(0.051)			(0.077)		
Corr(Det,Rm)		0.156**			0.028	
		(0.051)			(0.084)	
Corr(Stoc,Rm)		-0.291***			-0.571***	
		(0.057)			(0.077)	
Corr(PermMD80,Rm)			-0.313***			-0.267**
			(0.058)			(0.033)
(log)-Income				-0.095*	-0.097*	-0.005
				(0.042)	(0.043)	(0.027)
(log)-Wealth				0.301***	0.301***	0.289***
				(0.020)	(0.020)	(0.013)
HH Size				-0.122***	-0.117***	-0.106***
				(0.019)	(0.020)	(0.013)
Age				0.008	0.007	0.019**
				(0.011)	(0.011)	(0.007)
Education				0.236***	0.249***	0.181***
				(0.043)	(0.043)	(0.029)
Sex				0.196**	0.190***	-0.022
				(0.057)	(0.057)	(0.037)
Unemployed				-0.001	0.008	-0.013
				(0.104)	(0.105)	(0.075)
Retired				0.029	0.036	0.054
				(0.071)	(0.072)	(0.045)
Health				0.004	0.001	0.013
				(0.030)	(0.030)	(0.020)
Risk Aversion				-0.355***	-0.354***	-0.334***
				(0.010)	(0.010)	(0.007)
Year Dummy	YES	YES	YES	YES	YES	YES
N	9,217	9,186	27,445	5,719	5,709	12,758
Pseudo R^2	0.016	0.016	0.026	0.289	0.290	0.277

Table 5 Probit Estimates for Stock Market Participation (Benchmarking)
The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the cluster-based correlation between stock market return and the permanent component of labor income shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), and different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (tranBKK and permBKK, respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. We include the same control variables as in table 4 and year-fixed effects, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

	OWNSTKMF					
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.480***	-0.504***	-0.141	-0.153	-0.139***	-0.152***
	(0.117)	(0.118)	(0.037)	(0.038)	(0.038)	(0.038)
Corr(permMD80,Rm)		-0.185***		-0.276***		-0.261***
		(0.065)		(0.037)		(0.035)
Corr(tranBKK,Rm)	0.084	0.059				
	(0.113)	(0.114)				
Corr(permBKK,Rm)	-0.175	-0.164				
	(0.104)	(0.105)				
Beta (Guvenen et al. (2017))			0.221	-0.063		
			(0.086)	(0.093)		
Beta (C&V(2002))					-0.228**	-0.054
					(0.087)	(0.088)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	3,286	3,286	12,758	12,758	12,758	12,758
Pseudo R^2	0.320	0.322	0.272	0.276	0.272	0.277

Table 6 Tobit Estimates for Asset Allocation

The table reports the Tobit regression results for asset allocation decision. The dependent variable is the portfolio shares in stocks either directly or through mutual funds (propSTKMF). The main independent variables are the standard deviation and the correlation between stock market return and different specifications of labor income shocks: the labor income growth rate (Tot), the stochastic component of labor income growth rate (Stoc), the deterministic component of labor income growth rate (Det), the permanent component of labor income shocks estimated by using the minimum distance methodology (permMD80). Additional control variables are individual demographic characteristics detailed in table 1. We also control for year-fixed effects. Regressions (7)-(8) report the estimates from the Heckman model, in which we use the same control variables for the selection and the asset allocation regressions. Heckman model coefficients are estimated using maximum likelihood. We report in parentheses the Tobit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves 1993–2019.

	PropSTKMF						Heckman	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
St. Dev.	-0.125***	-0.131***	-0.095**	-0.052*	-0.059*	-0.033*	-0.011***	0.001
	(0.023)	(0.023)	(0.012)	(0.024)	(0.024)	(0.014)	(0.002)	(0.005)
Corr(Tot,Rm)	-0.173***			-0.175***				
	(0.028)			(0.026)				
Corr(Det,Rm)		0.147***			0.098**			
		(0.030)			(0.028)			
Corr(Stoc,Rm)		-0.184***			-0.182***			
		(0.027)			(0.027)			
Corr(PermMD80,Rm)			-0.217***			-0.085**	-0.049***	-0.018***
			(0.012)			(0.013)	(0.003)	(0.005)
(log)-Income				-0.029*	-0.033*	-0.007		-0.005
				(0.014)	(0.014)	(0.010)		(0.005)
(log)-Wealth				0.105***	0.105***	0.107***		0.011***
				(0.007)	(0.007)	(0.005)		(0.004)
HH Size				-0.043***	-0.041***	-0.037***		-0.011***
				(0.007)	(0.020)	(0.005)		(0.002)
Age				0.001	-0.002	0.017**		-0.001
				(0.004)	(0.004)	(0.003)		(0.001)
Education				0.083***	0.086***	0.075***		0.028***
				(0.015)	(0.015)	(0.011)		(0.006)
Sex				0.078**	0.079***	0.013		0.004
				(0.020)	(0.020)	(0.014)		(0.005)
Unemployed				-0.001	0.004	-0.001		-0.003
				(0.036)	(0.035)	(0.029)		(0.009)
Retired				-0.001	0.004	0.027		0.028***
				(0.024)	(0.024)	(0.017)		(0.007)
Health				0.010	0.009	0.007		0.005
				(0.010)	(0.010)	(0.008)		(0.003)
Risk Aversion				-0.127***	-0.126***	-0.134***		-0.043***
				(0.004)	(0.004)	(0.003)		(0.002)
Year Dummy	YES	YES	YES	YES	YES	YES	YES	YES
N	8,208	8,184	22,275	5,612	5,604	12,498	10,519	10,519
Pseudo R^2	0.012	0.015	0.022	0.319	0.322	0.292		
Lambda							-0.057***	0.074***
							(0.008)	(0.014)

Table 7 Tobit Estimates for Asset Allocation (Benchmarking)

The table reports the Tobit regression results for asset allocation decision. The dependent variable is the portfolio shares in stocks either directly or through mutual funds (propSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the cluster-based correlation between stock market return and the permanent component of labor income shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), and different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (tranBKK and permBKK, respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. We include the same control variables as in table4 and year-fixed effects, but we suppress coefficients to save in space. We report in parentheses the Tobit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

	OWNSTKMF					
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.173***	-0.185***	-0.029*	-0.033*	-0.028*	-0.033*
	(0.037)	(0.038)	(0.014)	(0.014)	(0.014)	(0.014)
Corr(permMD80,Rm)		-0.064**		-0.086***		-0.082***
		(0.067)		(0.014)		(0.013)
Corr(tranBKK,Rm)	0.075*	0.070				
	(0.036)	(0.036)				
Corr(permBKK,Rm)	-0.086*	-0.079*				
	(0.033)	(0.034)				
Beta (Guvenen et al. (2017))			0.079*	-0.011		
			(0.039)	(0.037)		
Beta (C&V(2002))					-0.086*	-0.026
					(0.034)	(0.035)
Controls	YES	YES	YES	YES	YES	YES
N	3,224	3,224	12,498	12,498	12,498	12,498
Pseudo R^2	0.366	0.368	0.289	0.292	0.289	0.292

Table 8 Frequency of market participation: Poisson regression estimates
The table reports results from cross-sectional Poisson regressions. The dependent variable is the number of waves in which respondents reported investing in stock either directly or through mutual funds (FreqSTKMF). The main independent variables are the standard deviation, the correlation between stock market return and different specifications of labor income shocks as described in table 10, and different measures of the relationship between labour income growth rate and stock market returns used in previous papers as described in table 11. We use the clustered correlation obtained with minimum distance estimation as ex-ante correlation parameter of the state-space model to reconstruct the PI shocks using the Kalman filter at the individual level (KF80). We include the same control variables as in table10 and year-fixed effects, but we suppress coefficients to save in space. We report in parentheses the robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

	FreqSTKMF					
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.289***	-0.303***	-0.055*	-0.343**	-0.060*	-0.060*
	(0.067)	(0.067)	(0.027)	(0.110)	(0.027)	(0.027)
Corr(Tot,Rm)	-0.333***					
	(0.071)					
Corr(Det,Rm)		0.091				
		(0.078)				
Corr(Stoc,Rm)		-0.334***				
		(0.072)				
Corr(PermMD80,Rm)			-0.162***			
			(0.052)			
Corr(KF80,Rm)			-0.077*			
			(0.038)			
Corr(tranBKK,Rm)				0.169		
				(0.105)		
Corr(permBKK,Rm)				-0.110		
				(0.102)		
Beta (Guvenen et al. (2017))					0.107	
					(0.072)	
Beta (C&V(2002))						-0.011
						(0.062)
Controls	YES	YES	YES	YES	YES	YES
N	411	409	1,567	207	1,567	1,567
Pseudo Rˆ2	0.359	0.360	0.257	0.361	0.256	0.256

Table 9 Correlation Parameters from DHS. Summary Statistics on Short Sample. The table reports summary statistics of the estimated correlations between stock market return and different specifications of labor income shocks: the total income shocks, the stochastic component, the deterministic component, the permanent component estimated by using the minimum distance methodology (MD) both at the individual and the cluster levels, the permanent component obtained using the Kalman Filter and computing the correlation at T (Ex-post KF), the permanent component obtained using the Kalman Filter and updating the correlation at each t (Revised KF). We also report summary statistics about different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (Permanent (BKK) and Transitory (BKK), respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. In the last column, we report the t-test for the statistical significance of the parameter and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DNB Household Survey and cover waves from 1993–2011.

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Total	-0.068	0.412	-0.603	-0.079	0.501	-6.165***
Deterministic	-0.003	0.315	-0.347	-0.041	0.415	-0.352
Stochastic	-0.068	0.412	-0.595	-0.085	0.503	-6.189***
Permanent (MD cluster)	0.310	0.334	-0.149	0.309	0.876	25.693***
Permanent (MD individual)	0.050	0.384	-0.389	0.037	0.551	4.998***
Permanent (Ex-Post KF)	0.276	0.532	-0.544	0.365	0.899	22.522***
Permanent (Revised KF)	0.303	0.583	-0.643	0.459	0.944	18.459***
Permanent (BKK)	-0.115	0.505	-0.752	-0.174	0.654	-6.712***
Transitory (BKK)	-0.038	0.425	-0.616	-0.049	0.519	-2.648***
Beta (Guvenen et al. (2017))	0.004	0.096	-0.016	0.001	0.049	1.694
Beta (C&V(2002))	-0.001	0.110	-0.031	-0.001	0.019	-0.316

Table 10 Probit Estimates for Stock Market Participation. Short Sample
The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stocks either directly or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation and the correlation between stock market return and different specifications of labor income shocks: the labor income growth rate (Tot), the stochastic component of labor income growth rate (Stoc), the deterministic component of labor income growth rate (Det), the permanent component of labor income shocks estimated by using the minimum distance methodology (permMD80). Additional control variables are individual demographic characteristics detailed in table A.1. We also control for year-fixed effects. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. We use data from the DNB Household Survey for the waves from 1993–2011 in columns (1) to (3), and for the waves from 1993–2007 in columns (4) to (6).

	OWNSTKMF					
	Sample: 1993-2011			Sample: 1993-2007		
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.092	-0.092	-0.172**	-0.025	-0.022	-0.074
	(0.051)	(0.051)	(0.067)	(0.059)	(0.059)	(0.056)
Corr(Tot,Rm)	-0.280***			-0.237***		
	(0.047)			(0.054)		
Corr(Det,Rm)		0.097			0.130**	
		(0.097)			(0.064)	
Corr(Stoc,Rm)		-0.316***			-0.286***	
		(0.047)			(0.055)	
Corr(PermMD80,Rm)			-0.289***			-0.227**
			(0.058)			(0.068)
(log)-Income	0.048	0.046	0.050	0.078	0.077	0.086**
	(0.036)	(0.036)	(0.044)	(0.042)	(0.042)	(0.041)
(log)-Wealth	0.277***	0.277***	0.267***	0.259***	0.259***	0.254***
	(0.016)	(0.016)	(0.016)	(0.018)	(0.018)	(0.018)
HH Size	-0.058***	-0.055**	-0.048**	-0.043*	-0.038*	-0.036
	(0.016)	(0.016)	(0.016)	(0.019)	(0.018)	(0.018)
Age	0.003	0.003	0.006	-0.008	-0.009	-0.005
	(0.009)	(0.011)	(0.009)	(0.009)	(0.011)	(0.011)
Education	0.151***	0.161***	0.189***	0.112**	0.122**	0.135**
	(0.037)	(0.037)	(0.037)	(0.042)	(0.043)	(0.042)
Sex	-0.025	-0.028	-0.076	-0.073	-0.083	-0.125**
	(0.049)	(0.049)	(0.049)	(0.058)	(0.059)	(0.058)
Unemployed	0.088	0.096	0.056	0.009	0.021	-0.023
	(0.099)	(0.098)	(0.098)	(0.112)	(0.112)	(0.111)
Retired	0.098	0.104	0.093	0.069	0.076	0.092
	(0.055)	(0.055)	(0.054)	(0.063)	(0.069)	(0.069)
Health	-0.015	-0.017	-0.005	-0.029	-0.031	-0.022
	(0.026)	(0.026)	(0.026)	(0.030)	(0.030)	(0.030)
Risk Aversion	-0.324***	-0.323***	-0.319***	-0.335***	-0.335***	-0.330***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
N	7,585	7,582	5,968	5,614	5,613	5,720
Pseudo Rˆ2	0.259	0.262	0.264	0.264	0.265	0.263

Table 11 Probit Estimates for Stock Market Participation (Benchmarking). Short Sample

The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stocks either directly or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the correlation between stock market return and the permanent component of labor income shocks estimated by using the minimum distance methodology (permMD80), and different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (transBKK and permBKK, respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. We include the same control variables as in table10 and year-fixed effects, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. We use data from the DNB Household Survey for the waves from 1993–2011.

	OWNSTKMF					
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.173**	-0.191**	-0.139**	-0.149**	-0.140**	-0.150**
	(0.067)	(0.051)	(0.067)	(0.047)	(0.048)	(0.039)
Corr(permMD80,Rm)		-0.289***		-0.284***		-0.289***
		(0.067)		(0.058)		(0.058)
Corr(tranBKK,Rm)	-0.073	-0.088				
	(0.052)	(0.052)				
Corr(permBKK,Rm)	-0.094*	-0.085*				
	(0.042)	(0.042)				
Beta (Guvenen et al. (2017))			0.190	0.152		
			(0.109)	(0.110)		
Beta (C&V(2002))					0.053	0.038
					(0.099)	(0.099)
Controls	YES	YES	YES	YES	YES	YES
N	5,677	5,677	7,745	7,745	7,745	7,745
Pseudo R^2	0.262	0.265	0.252	0.259	0.256	0.259

Table 12 Out-Of-Sample Probit Estimates
The table reports the out-of-sample Probit regression results for stock market participation decision for the waves 2012-2019. The dependent variable is a dummy variable for respondents who reported to own stock either directly or through mutual funds (OwnSTKMF), own stock only (OwnSTK) or mutual funds only (OwnMF), between 2012 and 2019. The main independent variable is the correlation between stock market returns and the PI shocks estimated at the cluster level using Minimum Distance and data up to 2011. This correlation is then assigned to each individual on the basis of the corresponding cluster to which the individual is allocated according to her personal characteristics observed in 2012. We also include other independent variables, and additional control variables as defined in table 4. N is the number of observations. The data are from the DNB Household Survey and cover waves from 1993–2019.

	OwnSTKMF	**OwnSTK**	**OwnMF**
	(1)	(2)	(3)
St. Dev. (dy)	-0.239***	-0.168**	-0.207**
	(0.078)	(0.090)	(0.075)
Corr(permMD80,Rm)	-0.207***	-0.140*	-0.136*
	(0.073)	(0.087)	(0.075)
N	4,336	4,336	4,336
Pseudo R^2	0.324	0.273	0.264

Table 13 Correlation and Beta Parameters. Summary Statistics (PSID)
The table reports summary statistics of the estimated correlations between stock market return and different specifications of labor income shocks: the total income shocks, the stochastic component, the deterministic component, the permanent component estimated by using the minimum distance methodology (MD) both at the individual and the cluster levels, the permanent component obtained using the Kalman Filter and computing the correlation at T (Ex-post KF), the permanent component obtained using the Kalman Filter and updating the correlation at each t (Revised KF). We also report summary statistics about different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (Permanent (BKK) and Transitory (BKK), respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. In the last column, we report the t-test for the statistical significance of the parameter and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the Panel Survey Income Dynamics (PSID) and cover waves from 1988–2011.

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Total	0.049	0.034	-0.697	0.065	0.733	2.915***
Deterministic	-0.013	0.208	-0.703	-0.015	0.611	-1.273
Stochastic	0.046	0.322	-0.681	0.045	0.731	2.851***
Permanent (MD cluster)	0.268	0.209	-0.001	0.268	0.539	10.356***
Permanent (MD individual)	0.048	0.357	-0.900	0.027	0.900	2.757***
Permanent (Ex-Post KF)	0.399	0.330	-0.041	0.426	0.789	39.049***
Permanent (Revised KF)	0.410	0.347	-0.060	0.456	0.823	25.244*
Permanent (BKK)	0.069	0.338	-0.657	0.074	0.733	4.125***
Transitory (BKK)	0.016	0.449	-0.823	0.026	0.825	0.719
Beta (Guvenen et al. (2017))	0.232	0.164	0.067	0.216	0.432	52.610***
Beta (C&V(2002))	0.040	0.144	-0.096	0.043	0.164	10.396***

Table 14 Probit Estimates for Stock Market Participation (PSID)
The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stocks either directly or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation and the correlation between stock market return and different specifications of labor income shocks: the labor income growth rate (Tot), the stochastic component of labor income growth rate (Stoc), the deterministic component of labor income growth rate (Det), the permanent component of labor income shocks estimated by using the minimum distance methodology (permMD) at the cluster-level. We also include different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (transBKK and permBKK, respectively), and the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. Additional control variables are individual demographic characteristics, such as the (log)-labour income, marital status, family size, age and years of schooling. We also control for year-fixed effects. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. We use data from the Panel Survey Income Dynamics (PSID) for the waves from 1988–2011.

	OWNSTKMF					
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.001	-0.001	0.015	0.015	0.015	0.014
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Corr(Tot,Rm)	-0.064					
	(0.048)					
Corr(Det,Rm)		0.050				
		(0.081)				
Corr(Stoc,Rm)		-0.131**				
		(0.047)				
Corr(PermMD,Rm)			-0.401***	-0.365***	-0.398***	-0.399***
			(0.077)	(0.077)	(0.077)	(0.077)
Corr(tranBKK,Rm)				0.123*		
				(0.057)		
Corr(permBKK,Rm)				-0.075		
				(0.043)		
Beta (Guvenen et al. (2017))					-0.027	
					(0.120)	
Beta (C&V(2002))						0.048
						(0.122)
Controls	YES	YES	YES	YES	YES	YES
N	6,804	6,804	6.804	6,804	6,804	6,804
Pseudo R^2	0.091	0.092	0.094	0.094	0.094	0.095

Table 15 Individual correlations and personal traits: Ordinary least squares regression The table reports the cross-sectional OLS regression results of individual correlations between permanent income shocks and stock market returns over the personal characteristics used to cluster individual in homogeneous groups. The dependent variable is the individual correlation parameter estimated using the minimum distance methodology. The independent variables are the observable traits that may be used to cluster the individuals in homogeneous groups. For each trait, we select the mode over time for each individual. Coefficients in bold are statistically significant at the 10% significance level.

	Independent Variables						
Correlation	Sex	Education	Risk Aversion	Urban	Financial Literacy	N	Adj R^2
	0.051	**-0.012**	**-0.003**	**0.063**	**0.023**	12,957	0.012

Table 16 Summary statistics. Short Sample

This table reports summary statistics for the variables used for the empirical analysis. The data are from the DNB Household Survey and cover all the waves from 1993–2011. N denotes the total number of observations, n denotes the number of individuals, and T is the average number of years in which those individuals participated in the survey. The definitions of the variables are provided in Appendix A.

Variable	Mean	St. Dev.	p10	Median	p90	N (n x T)
OwnSTK	0.06	0.24	0	0	0	16976
OwnMF	0.18	0.38	0	0	1	14999
OwnSTKMF	0.33	0.47	0	0	1	16976
PropSTK	0.02	0.13	0	0	0	13801
PropMF	0.06	0.18	0	0	0.24	13801
PropSTKMF	0.10	0.23	0	0	0.43	13801
Ln(NetWorth)	11.55	1.75	8.89	12.17	13.08	10116
Ln(NetIncome)	9.74	0.86	8.76	9.94	10.51	13694
Corr(d(LnInc),Rm)	-0.07	0.41	-0.60	-0.06	0.50	27664
SD(d(LnInc))	0.39	0.49	0.07	0.21	0.96	27664
HH size	2.42	1.21	1	2	4	16945
Age	53.91	13.64	35	54	72	16945
Education	0.48	0.49	0	1	1	16945
Male	0.63	0.48	0	1	1	16945
Unemployed	0.13	0.33	0	0	1	16945
Retired	0.17	0.37	0	0	1	16945
Health	3.88	0.70	3	4	5	14809
Fin. Literacy	0.36	0.48	0	0	1	16968
Risk Aversion	4.41	2.06	1	5	7	13833

Table 17 Probit Estimates for Stock Market Participation: Sub-samples

The table reports the probit regression results for stock market participation decisions considering sub-samples of the full sample of 1,884 individuals. The dependent variable is a dummy variable for respondents who reported either stocks directly or mutual funds (OwnSTKMF). The independent variables are the standard deviation and the correlation between stock market returns and different specifications of labor income shocks. For details, review Tables 4 and 5. We include the same control variables as in table 10 and year-fixed effects, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

	OwnSTKMF								
	Males			College-Graduated			Not-Retired		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
St. Dev.	-0.382***	-0.364***	-0.159***	-0.288***	-0.294***	-0.132***	-0.221***	-0.159***	-0.132***
	(0.079)	(0.079)	(0.038)	(0.087)	(0.087)	(0.045)	(0.051)	(0.076)	(0.076)
Corr(tot,Rm)	-0.478***			-0.352***			-0.432***		
	(0.083)			(0.091)			(0.079)		
Corr(det,Rm)		-0.337***			0.207*			0.096*	
		(0.093)			(0.105)			(0.086)	
Corr(stoc,Rm)		-0.495***			-0.351***			-0.446***	
		(0.083)			(0.091)			(0.079)	
Corr(permMD80,Rm)			-0.209***			-0.242***			-0.255***
			(0.039)			(0.040)			(0.034)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	4,760	4,752	11,037	3,697	3,692	8,091	5,356	5,353	12,534
Pseudo R^2	0.304	0.307	0.277	0.292	0.294	0.291	0.292	0.293	0.274

Table 18 Probit Estimates for Stock Market Participation: Stocks or Mutual Funds The table reports the probit regression results for stock market participation decisions. The dependent variable is a dummy variable for respondents who reported either owning stock (OwnSTK), mutual funds (OwnMF), or stocks only (OwnSTKnoMF). The independent variables are the standard deviation and the correlation between stock market returns and different specifications of labor income shocks. For details, review Tables 4 and 5. We include the same control variables as in table 10 and year-fixed effects, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***,**,* over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

	OwnSTK			OwnMF			OwnSTKnoMF		
	(1)	(2)	(3)	(4)	(5)	(6)			
St. Dev.	-0.022	-0.016	-0.009	-0.242***	-0.256***	-0.173***	-0.022	-0.002	-0.037
	(0.081)	(0.082)	(0.039)	(0.071)	(0.071)	(0.037)	(0.093)	(0.093)	(0.045)
Corr(tot,Rm)	-0.551***			-0.322***			-0.406***		
	(0.090)			(0.075)			(0.106)		
Corr(det,Rm)		-0.248**			0.088			-0.198*	
		(0.097)			(0.082)			(0.111)	
Corr(stoc,Rm)		-0.600***			-0.283***			-0.546***	
		(0.090)			(0.075)			(0.106)	
Corr(permMD80,Rm)			-0.178***			-0.242***			-0.252***
			(0.039)			(0.032)			(0.043)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	6,059	6,049	14,485	6,059	6,049	14,485	6,059	6,049	14,485
Pseudo R^2	0.304	0.307	0.272	0.209	0.210	0.209	0.192	0.197	0.153

Table 19 Calibration parameters

Description	Parameter	Value
Working life (max)	T	20 -65
Retirement (max)	$t_0 + K$	65 -100
Discount factor	β	0.96
Risk aversion	γ	5 and 8
Replacement ratio	λ	0.68
Riskless rate	r	0.02
Excess returns on stocks	μ^s	0.04
St. dev returns on stocks	σ_s	0.18