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: July 15th, 2022

Abstract

This paper estimates the individual correlation between permanent earnings shocks and an aggregate shock affecting both earnings and stock market returns. These correlation estimates (on average equal to 0.2-0.3 in both Dutch and U.S. data) imply larger hedging motives than previously thought. They retain statistical significance in predicting both portfolio choice and participation when other measures do not. They explain participation both out-of-sample and for the same individual over time. The data consistently support the theoretical prediction that portfolio holdings of equities respond to such correlations.

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, as well as participants in the 2021 CeRP Workshop, the 2020 Consob ESMA Bocconi webinar, the 2021 Inquire Europe webinar, the 2022 European Meetings of the Econometric Society, 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. However, the evidence of correlation as a driver of observed portfolios is less clear-cut (see e.g. Heaton and Lucas (2000), Campbell, Cocco, Gomes, and Maenhout (2001), Campbell and Viceira (2002), Cocco, Gomes, and Maenhout (2005a), Angerer and Lam (2009), Bonaparte, Korniotis, and Kumar (2014), Catherine, Sodini, and Zhang (2020)). Both stock market participation and portfolio choice are 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 individual correlation coefficients are often estimated to be close to zero, so that they are unlikely to account for moderate risk taking in equity markets. This paper uses a novel approach to estimate the individual correlation between PI shocks and an aggregate shock affecting both earnings and stock market returns. These new estimates (averaging 0.2-0.3 in both Dutch and U.S. data) imply much larger hedging motives than previously thought. They retain statistical significance in predicting both portfolio choice and participation when other measures of correlation between PI shocks and stock market returns do not. They are also able to predict participation out-of-sample. Since unobserved heterogeneity may plague cross-sectional results, we show that individual participation over time is sensitive to updates of correlation estimates. For the first time, the data consistently align with the theoretical prediction that portfolio holdings of equities respond to such correlations.

Our approach posits that individuals display heterogeneous exposure to aggregate risk, as discussed in Campbell and Viceira (2002) and Guvenen, Schulhofer-Wohl, Song, and Yogo (2017). Accordingly, we allow for correlation between an aggregate shock and individual PI shocks in a labor income process that is commonly used (see Carroll and Samwick, 1997 and Cocco et al., 2005a) in the literature. The resulting implied moment conditions exploit 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 noisy exposure to aggregate risk, in proportion to their correlation coefficients. More precisely, individual log labor income is modeled as the sum of a function of demographic and personal characteristics (e.g., education) and a stochastic trend hit by permanent and transitory shocks. Only the former permanently affect the level of individual earnings. Moreover, individual PI shocks feature both idiosyncratic and systematic components. In our setting, the latter co-moves with the aggregate shock according to the individual's correlation and, through it, with the stock market return. That is, an individual's PI shocks noisily co-move with both the aggregate shock and stock market returns. While previous methods focus on each individual earnings' growth over time to pin down correlation with the stock return, we exploit the co-movements between any two individuals' earnings shocks to capture exposure to the aggregate shock. Since the covariance of the idiosyncratic earnings shocks across different agents is zero, the observable co-movements between any two individuals' earnings shocks are due to each shock's dependence on aggregate risk. It is this individual correlation with aggregate risk that will help predict conditional portfolio shares and participation in the stock market.

We estimate the parameters characterizing the joint distribution of earnings and the aggregate shock with data from the Dutch National Bank Household Survey (DNB), an annual panel covering years from 1993 to 2019. The rich information about individuals is the reason why a closely related paper (Bonaparte et al., 2014) prefers it to US data sets such as the Panel Study of Income Dynamics (PSID). We exploit such rich information on participants to perform a minimum distance (MD) estimation on 80 clusters (MD80), grouping individuals according to time-invariant observable characteristics, such as education, risk aversion, sex and financial literacy. Finally, we study whether such MD80 correlation estimates pin down the variation in equity investments. We repeat our analysis on the 1993-2011 waves. This allows for a direct comparison with (Bonaparte et al., 2014), that finds support for income hedging motives using a moving-average method. It also enables to perform an out-of-sample experiment on the remaining waves. Finally, we apply our method on PSID data on stock market participation.

We find that the propensity to participate in the stock market is highly sensitive to the MD80 correlation between PI shocks and the aggregate shock. For instance, the probability of individual participation in the equities market decreases by 12% when the correlation coefficient increases from -0.6 to 0.5, which respectively correspond to the 10-th and the 90-th percentiles of the empirical distribution of the correlations between income shocks and stock market returns. An equivalent switch in the correlation between realized income growth and stock returns using previous time series methods (Bonaparte et al., 2014) would decrease the probability of participation by about 5%, only. Results concerning asset allocation confirm the economic importance of hedging motives, with the total equity share decreasing by 10.70% when the MD80 correlation changes from -0.6 to 0.5. Moreover, the explanatory power of the MD80 correlation estimates for portfolio holdings is robust, also relative to several competing methods, not only in long samples ($T = 27, 18$) but also also in shorter ones ($T = 14$). We can therefore implement an out-of-sample analysis on the remaining waves. Even out-of-sample, correlation estimates predict participation.

A concern regarding the time series approach in (Bonaparte et al., 2014) is that the estimated correlations based on past income shocks may not reflect future hedging needs. This concern emerges in Catherine et al. (2020), that find a change in the sign of the relationship between an individual's equity shares and correlation when this is solely based on the individual's future income shocks. Our estimation method is immune from this problem, as it simultaneously uses inter-temporal conditions at multiple lags and cross-sectional conditions. Moreover, it pins down the correlation parameter for a given agent relying on conditions involving the co-movement in income growth between pairs of other agents.

Another concern in cross-sectional studies is the possibility that both the individual's earnings risk and portfolio choice are driven by unobserved characteristics. To address this concern, we add to the literature an analysis of changes in individuals' risk-taking decisions over time. We estimate such revised correlations after reconstructing the dynamics of individual PI shocks using a Kalman filter. 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 and the MD80 exposure to aggregate risk. These revised correlations display explanatory power, adding to the impact of hedging needs on risk-taking when compared to previous estimation approaches. Increasing correlation between PI shocks and aggregate shocks from -0.6 to 0.5 reduces participation in the equities market by a minimum of 21 months according to our estimates and by 14 months according to the method described in Bonaparte et al. (2014).

Previous results go through when we model the stock market returns as a linear combination of the aggregate risk factor and an idiosyncratic noise, and compute correlations between PI shocks and such noisy returns (see the Online Appendix). In sum, our results consistently overturn the commonly held view (see e.g. recently Catherine et al. (2020)) that heterogeneous individual exposures to aggregate risk (and therefore to stock returns) do not matter for hedging permanent income shocks.

Our result are also relevant for 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, 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 a latent aggregate shock. 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 in this literature, our approach precisely pins down the heterogeneous exposures to aggregate risk. We use a single factor representation for aggregate risk, that may seem restrictive. Even if the mentioned results of the empirical analysis are comforting, the Online Appendix addresses this concern applying a principal component analysis (PCA) to the covariance matrix of individuals' income growth and estimating the individual loadings associated to each principal component. The high correlation (0.6) between the individual MD correlation estimates and the sum of such loadings indicates a close relationship between the non-parametric multi-factor representation of aggregate risk and our parsimonious single factor representation.

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 high (0.32-0.5). In Cocco, Gomes, and Maenhout (2005b), 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 results support the presence of a bias that is reduced by clustering, as the Online Appendix formally shows. We indeed find that the mean MD correlation between PI shocks and the aggregate shock is as high as $0.2 - 0.3$ when we group individuals, while it drops to 0.06 when we do not. This happens both on Dutch and on US data. This difference rules out the possibility that the shift to the right in the distribution of MD80 correlations owe to the restrictions imposed by our parsimonious model. Both MD individual and MD80 correlations indeed rely on the same model.[1]. 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).[2]

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

[1] Aside from reducing an estimation bias, imposing the same correlation within a cluster may also reveal the common exposure to aggregate risk that each individual occasionally shields through a new job (as in Low, Meghir, and Pistaferri (2010)) or informal insurance (Guvenen and Smith (2014))

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

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.[3] 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 co-move with that household's risky share. Against this background, Bonaparte et al. (2014) identifies the role of the individual sample correlation between labor income shocks and stock market returns in explaining stock market participation, a role that is however questioned in (Catherine et al., 2020). Our work contributes an original approach to the assessment of hedging motives, that resurrects the broad take-away of Bonaparte et al. (2014). Our results imply that heterogeneous exposure to aggregate risk is able to predict both participation in the stock market and the observed equity share in individuals' portfolios, consistent with life-cycle theory.

A recent strand of research shows that adding non-linear income risk makes it possible to replicate the age profile of low equity investment ((Bagliano, Fugazza, and Nicodano, 2019), (Catherine, 2022), (Galvez and Paz Pardo, 2022), (Shen, 2021)) while other mechanisms, such as participation costs, contribute to explain non participation in the equity market. In these papers the individual correlation between labour income shocks and stock market returns either is ignored or plays a limited role on the optimal stock share, when measured with traditional time series methods (see e.g. (Catherine et al., 2020)). Our results suggest reconsidering the role played by the individual correlation between PI shocks and aggregate shocks to track equity market participation also out-of-sample, while reinforcing results regarding the optimal portfolio share.

The rest of the paper is organized as follows. In Section 2, we introduce the model for the individual

[3]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), among others cited in Gomes, Haliassos, and Ramadorai (2020).

labor income process, the aggregate shock, 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 portfolio shares to such correlation. 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, repeats the analysis using PSID data, and reports our out-of-sample analysis. Section 6 presents our concluding remarks.

2 The Model and the Implied Moment Conditions

This section presents our parsimonious model of the joint distribution of earnings, the aggregate shock and the stock return. The stochastic process of the individual labor income will capture the notion that agents are differently exposed to the aggregate shock through the permanent component of their labor income shocks.

In section 2.1, we derive the implied moment conditions. We exploit the restrictions imposed by the model not only on the time-series distribution of individual income shocks, as in prior literature, but also on the cross-sectional one. The time-series restriction is the zero inter-temporal covariance of the idiosyncratic transitory shock for each agent. The cross-sectional restriction is the zero covariance of the idiosyncratic component across different agents. In section 2.2, we present the sample counterparts of the implied moment conditions and formalize the MD estimator for the unknown parameters. These are the correlations between the individual permanent shocks and the aggregate shock, the variance of the permanent shock and the variance of the transitory shock.

To begin, consider an economy with N individuals, indexed by i, each working for T years. At each time t, each individual receives the labor income, $Y_{i,t}$. Following Cocco et al. (2005a), the log-labor income process is 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 $u_{i,t} = \sigma_u W^p_{i,t}$, and $\epsilon_{i,t} = \sigma_\epsilon W^q_{i,t}$, where $W^p_{i,t}$ and $W^q_{i,t}$ 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 further assume that that $W^p_{i,t}$ is correlated with the aggregate shock W_t:[4]

$$W^p_{i,t} = \rho_i W_t + \sqrt{1-\rho_i^2}\, W_{i,t},$$

where ρ_i denotes the correlation coefficient and W_t and $W_{i,t}$ are two standard normal random variables.

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

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

Thus the PI shock is 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.

The main body of the paper focuses on the individual correlations of PI shocks with the systematic

[4] The Online Appendix benchmarks results obtained with this single factor representation of the aggregate shock against those obtained through a multi-factor representation.

component (or aggregate risk factor). In the Online Appendix we consider correlations with stock market returns, assuming they are a linear combination of the aggregate factor and an idiosyncratic noise, with the latter that follows a standard normal distribution, such that $r_t \sim \mathcal{N}(0, \sigma_r^2)$, where σ_r denotes the standard deviation of the stock market returns.

2.1 Model-Implied Moment Conditions

This section derives two sets of moment restrictions that will be used to estimate the variance of the transitory shock, the variance of the PI shock and the individual correlations between PI shocks and the latent aggregate shock. One set pins down the two above-mentioned variances from the observable inter-temporal covariance of the time variation in total income shocks (DTS), for each individual. The other shows that the observable covariance between any two individuals' DTS is a linear function of the two individuals' correlation coefficients. This owes to the dependence of each individual's PI shock on the aggregate risk factor. This second set gives rise to $N(N-1)/2$ conditions that will allow to estimate the N correlation coefficients. The dimension of the first set would be equal to $T(T-1)/2$ if we only relied on the intertemporal restrictions on the familiar DTS over one period. Instead, we also consider the DTS over more than one period thereby enlarging the number of moment conditions.

Formally, the total shocks (TS) to labor income for individual i at time t, $e_{i,t}$, are defined in equation (2). Let DTS denote the time variation in TS for each individual over a time interval of length d:

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

where $d = \{1, 2, ..., D\}$, and D is the maximum length of the time interval.

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

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

.

For instance, when $d = 1$, $\Delta_1 e_{i,t}$ is the familiar 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 build two sets of variance-covariance matrices of DTS.

In the first set, we construct D matrices, one for each length d. Each matrix, with dimension $[(T-d) \times (T-d)]$, has the cross-sectional variance of the DTS for each point in time on the main diagonal. This is the variance of the N-dimensional vector containing the DTS of N individuals at each time t, with t going from 1 to $(T-d)$. Off the main diagonal are the intertemporal covariances; that is, the covariances between the N-dimensional vectors containing the DTS of N individuals at different lags. Then, each $[(T-d) \times (T-d)]$ matrix features the following symmetrical form:

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

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. For instance, when $d = 3$ and $l = 1$:

$$\Delta_d e_{i,t+l} = e_{i,t+4} - e_{i,t+1}.$$

Hence, when $l = 0$, $C_d(t, t+l)$ is the cross-sectional variance—at time t—of the DTS of length 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.

We now build 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:

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

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 $i = j$:

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

because the covariance between individuals is due to the correlation with the aggregate shock:

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

$$cov(\sigma_u \rho_i W_{t+1}, \sigma_u \rho_j W_{t+1}) = \sigma_u^2 \rho_i \rho_j,$$

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 $cov(W_{t+1}, W_{t+1}) = var(W_{t+1}) = 1$.

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.[5] This feature of the economy enables exploitation of the data's cross-sectional dimension to infer 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 $(N(N-1))/2$.[6]

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 DTS, $(\Delta_d e_{i,t})$, are the first differences of regression residuals for each individual and are denoted as *dres*. Then, we construct two sets of variance-covariance matrices of the *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' *dres* and we fill the on-diagonal entries by computing the variance of each individual's *dres*.

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 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.

[5] We can expand this matrix using the condition on the covariance between each individual DTS time series and the stock market returns, obtaining a $[(N+1) \times (N+1)]$ matrix, in which the last row and the last column contain the N covariance terms between the N individuals' DTS and the stock market returns. However, results are substantially unaltered.

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

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

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

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:

$$Q(\theta) = (g_M - G_M(\theta))' I_M (g_M - G_M(\theta)) \tag{14}$$

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.[7]

Because we have $N+2$ unknown model parameters and $M > N+2$, the model is (highly) over-identified. Specifically, using the inter-temporal restrictions, we obtain at each d a number of moment conditions equal to:

$$M_d^{int} = \frac{(T-d)(T-d-1)}{2} + (T-d) = \frac{(T-d)(T-d+1)}{2},$$

where the first term is the number of conditions off the main diagonal and the second term is the number of conditions on the main diagonal, respectively, of each $[(T-d)x(T-d)]$ matrix.

[7] 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.

Furthermore, using the cross-sectional restrictions, we obtain an additional number of moments conditions equal to:

$$M^{crs} = \frac{N(N-1)}{2} + N = \frac{N(N+1)}{2}$$

from the NxN matrix containing $(N(N-1)/2)$ covariance terms between the individuals' *dres* off the main diagonal and the variance of the N individuals' *dres* on the main diagonal. As a result, we use an overall set of moment conditions of size equal to:

$$M^n = M^{int} + M^{crs},$$

that we can summarize as a function of the D inter-temporal lags, the length of the time-series, T, and the number of individuals in the sample, N:

$$M^n = M(D, T, N)$$

For instance, with 27 yearly observations, 1884 individuals, and using at most 18 inter-temporal lags, we obtain roughly 1.776 millions of moment conditions to estimate 1886 parameters–that is,

$$M^n = M(18, 28, 1884) = 1,776,662$$

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.

Several reasons dictate using the DNB survey as baseline 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 at the individual level over the entire time span. Second, the DNB is the reference dataset in the assessment of hedging motives by Bonaparte et al. (2014), which is the natural benchmark for comparing results. Furthermore, the large cross-sectional dimension of these data may enable precise measurement of the income-return correlation parameter, also when the time dimension is short. Finally, 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 when estimating the correlation between individual labor income growth and stock market returns.

[Table 1 about here.]

We provide details on 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. [8] 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. [9]

[Table 2 about here.]

3.1 Relating Correlations to Observable Characteristics

This section estimates the individual correlation coefficient between PI shocks and the aggregate shock, assuming it is common within groups of agents with similar time-invariant characteristics. Then it compares the distribution of our cluster-based estimates to those obtained using the methods in previous literature, including some cluster-based ones (Campbell and Viceira (2002) and Guvenen et al. (2017)).

We embed the restriction that the correlation coefficient, defined in equation (4), be cluster-specific rather than individual-specific. Accordingly, the equation describing the individual PI shocks becomes:

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

[9] We also present descriptive statistics for the short sample up to 2011 in the Online Appendix, in which we report very similar figures to the longer sample in terms of personal characteristics of the individuals.

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

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

To ensure consistency with both the model and the empirical strategy, the clustering variables must fulfill two conditions. First, the observable variables must be stable 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, a simple regression analysis supports this clustering procedure.

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.[10] Figure 1 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 MD80 correlations to the one of sample correlations

[10]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).

between stock market returns and total income shocks, TS. The right-hand panel plots the MD80 correlations against the correlations between stock market returns and PI shocks obtained with the approach of Bonaparte et al. (2014). We also report summary statistics of the correlation parameter estimates in Table 3.

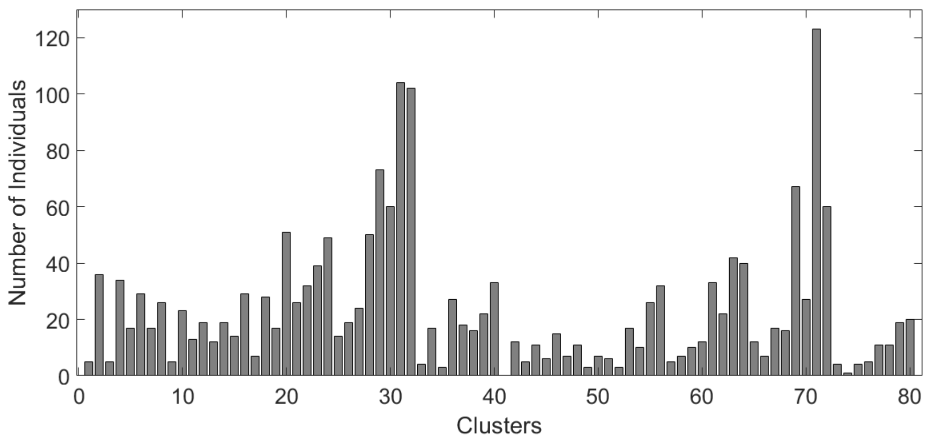
Results are striking. The distribution of MD80 correlation parameters shifts to the right with respect to both alternatives, indicating larger average hedging needs. The positive mean correlation signals that the PI shocks of most agents have the same sign of business cycle movements (which are, in turn, reflected in the stock market return). The difference with respect to the correlation between PI shocks and stock market returns in Bonaparte et al. (2014) can be traced back to their use of realized labor income shocks, since the PI shock at time t is measured as the equally-weighted average of labor income growth at time $t-1$, t, and $t+1$. This distribution of estimated correlations turns out to be widely dispersed around its mean, like the one between total income shocks and stock returns. In contrast, the distribution of MD80 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).

Thus, it appears that the MD80 restrictions based on the covariance matrix of contemporaneous income innovations capture 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) and by our Online Appendix, ignoring the differential exposure across clusters of workers to aggregate risk leads to underestimating that exposure. Furthermore, imposing the same correlation for each agent within a cluster isolates the common exposure to the aggregate shock that each individual may occasionally shield through a new job (as in Low, Meghir and Pistaferri (2020)) or informal income support (Guvenen and Smith (2014)).

A reader may be tempted to attribute the shift in the distribution of estimated correlations to the restrictions imposed by our parsimonious model. This is not the case. In the Appendix A we report the same distribution for individual, non cluster-based, correlations that are based on the same model. The mean correlation drops from 0.257 to 0.057 (see 3). [11] Moreover, in the Online Appendix we estimate the cluster-based correlations with stock market returns, explicitly assuming that the latter are noisily correlated with the aggregate risk factor. The resulting correlations are

[11] Clustering also reduces the number of unknown parameters, thereby increasing efficiency.

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).



very similar to the ones reported in this section.

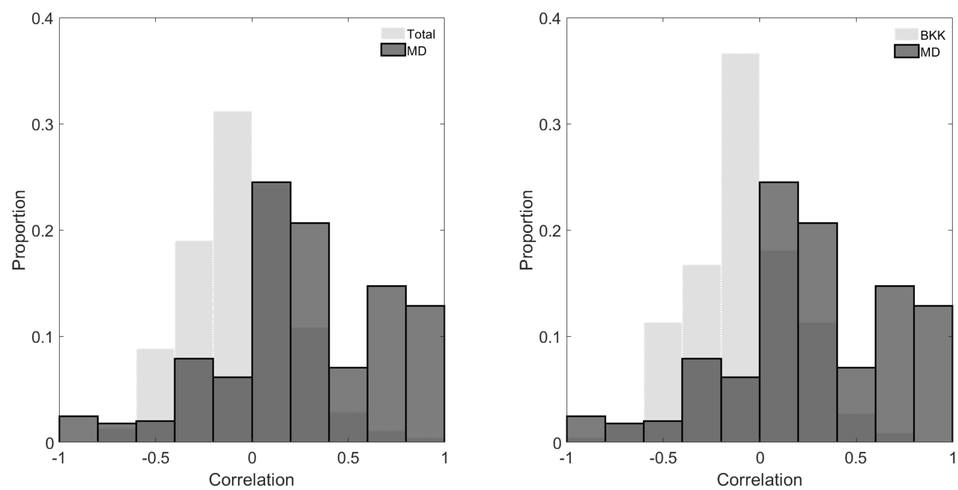
[Table 3 about here.]

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 compare results based on our parametric approach to estimation of PI shocks with the moving average one in Bonaparte et al. (2014). To this end, we use the same specification but for the individual correlation coefficients. 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 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

Figure 2. Distribution of cluster-based correlations
The figure's left panel compares—after clustering the individuals in 80 homogeneous groups—the distribution of the estimated correlations between the permanent labor income shocks and the aggregate shocks estimated through the minimum distance (MD) methodology with the empirical distribution of the 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.



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.

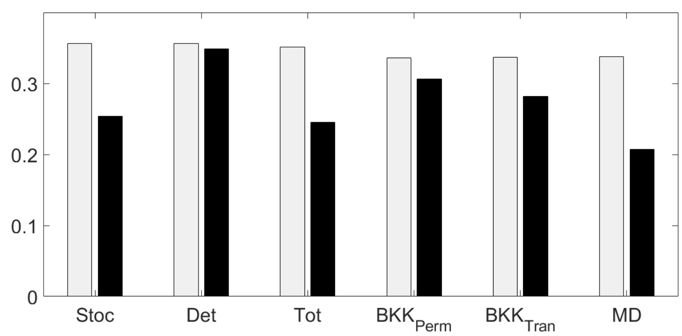
4.1 Stock Market Participation

[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

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 aggregate shocks and PI shocks estimated using the minimum distance (MD) methodology as well as 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). 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.



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.

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 hypothesis. 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 aggregate shocks estimated using our MD80 approach. 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 estimates. 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)). [12] 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 et al. (2017) (columns (3-4)) and Campbell

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

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 betas display statistical significance (see columnn 3 and 5). However, this disappears when we also allow in the regression the MD80 correlation estimates.

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 can gauge the economic significance of hedging motives implied by alternative estimates by computing the marginal effect of the variables on the propensity to invest. The probability of participation in the stock market, according to the Probit model, is given by

$$P(\mathcal{I}_{i,t}=1)=\Phi\left(\sum_{k=1}^{K}\beta_k X_{k,i,t}\right), \tag{15}$$

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. Therefore, the marginal effect of the k-th variable on $P(\mathcal{I}_{i,t}=1)$ is given simply by

$$\frac{\partial P(\mathcal{I}_{i,t}=1)}{\partial X_{k,i,t}}=\frac{\partial\Phi(\beta' X)}{\partial X_{k,i,t}}=\phi(\beta' X)\beta_k, \tag{16}$$

where ϕ indicates the probability distribution function of a standard normal variable, and $\beta' X$ is the equivalent in matrix notation of the argument of Φ in (15). Thus, the marginal effect also depends on the original probability. For instance, consider two individuals with original propensity to participate in the stock market equal to 50%, which corresponds to $\beta' X=0$ because

$$\beta' X = \Phi^{-1}(P(\mathcal{I}_{i,t} = 1))$$

where Φ^{-1} denotes the inverse of the cumulative distribution function of a standard normal variable.

We ascribe the two individuals to the 10-th and the 90-th percentiles of the empirical distribution of the correlations between TS and stock market returns, which correspond to ρ=-0.6 and ρ=0.5, respectively. Using the results of column (6) in Table 4, the first individual's propensity to participate is 12% higher than the second individual's propensity to participate $((-0.6-0.5)*(-0.267)*\phi(\hat{\beta}'X))$. However, using the results of column (1) in Table 5, we conclude that an equivalent difference in the correlation between the two individuals makes the first individual's probability of participating only 5% higher compared to that of the second individual $((-0.6-0.5)*(-0.175)*\phi(\hat{\beta}'X))$, according to the approach described in Bonaparte et al. (2014).

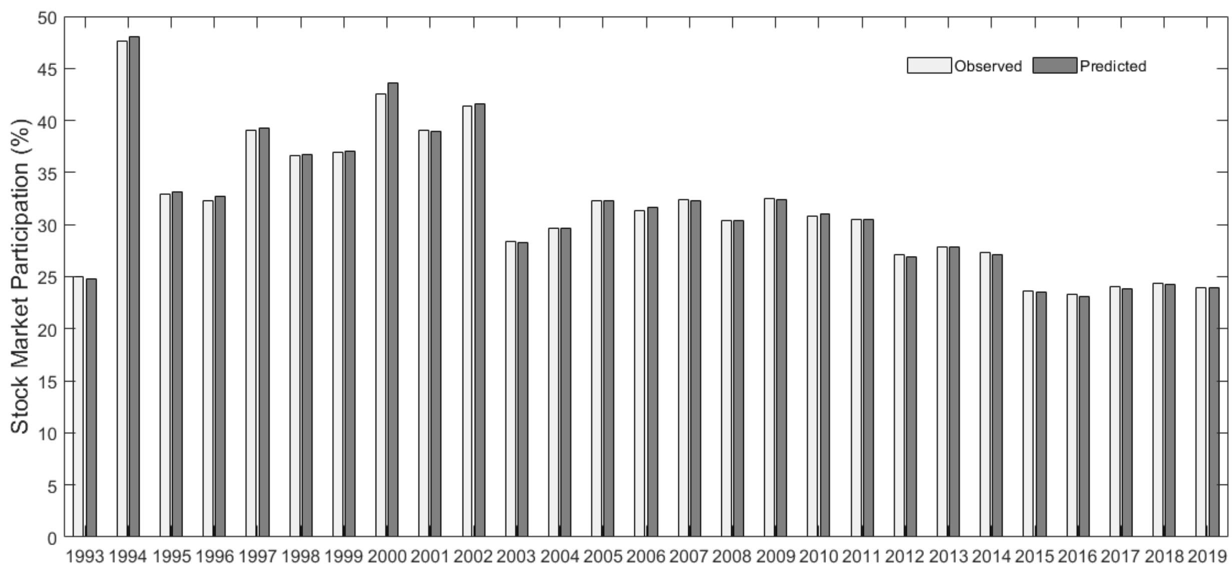
We then assess the overall model fit of our probit estimates. We compare the model-implied probability of participation in the stock market and the observed stock market participation rate, for each year of our sample. Using equation (15), we compute $P(\mathcal{I}_{i,t} = 1)$ for each year and for each 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 MD80 estimates aligns with the observed one for each year of our sample.

These results go through when we model the stock market returns as a linear combination of the aggregate risk factor and an idiosyncratic noise, and compute correlations between PI shocks and such noisy returns. Indeed, we see that the coefficients equal to -0.313 and -0.267 in Table 4, Column 3 and 6, respectively become -0.321 and -0.249 (see Table 21 of the Online Appendix.). Similarly, the coefficients in Table 5, column 4 and 6, change from -0.276 and -0.261 to -0.248 and -0.241 in Table 21 in the Online Appendix. There is hardly any change in statistical significance.

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

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 (15) 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.



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.[13]

Our results are in line with prior empirical results in the literature.

[13] 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.

Table 7 reports the benchmarking analysis. 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.

The previous results concerning asset allocation hold when we compute correlations between Pi shocks and stock market returns that are a combination of the aggregate risk factor and noise. For instance, the coefficients equal to -0.217 and -0.0.085 in Table 6, Column 3 and 6, respectively become -0.218 and -0.073 (see Table 22 of the Online Appendix.). Similarly, the coefficients in Table 7, column 4 and 6, change from -0.086 and -0.082 to -0.07 and -0.069 in Table 22 of the Online Appendix.

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. This evidence is based on replicas of existing Probit, Tobit and Heckman regression specifications in Bonaparte et al. (2014) that focus on the cross-section. In the following section, we 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

Hedging choices change over time because earnings and return realizations affect them through both budget constraints and learning (as in Chang, Hong, and Karabarbounis, 2018). 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 the aggregate shock.

To this end, we first reconstruct the sequence of unobservable PI shocks at the individual level using a Kalman filter, and identifying the latent aggregate shock with the stock market returns. The Kalman filter thus exploits the assumed relationship between labor incomes and stock market returns, their realizations and the MD80 parameter estimates. 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 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:

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

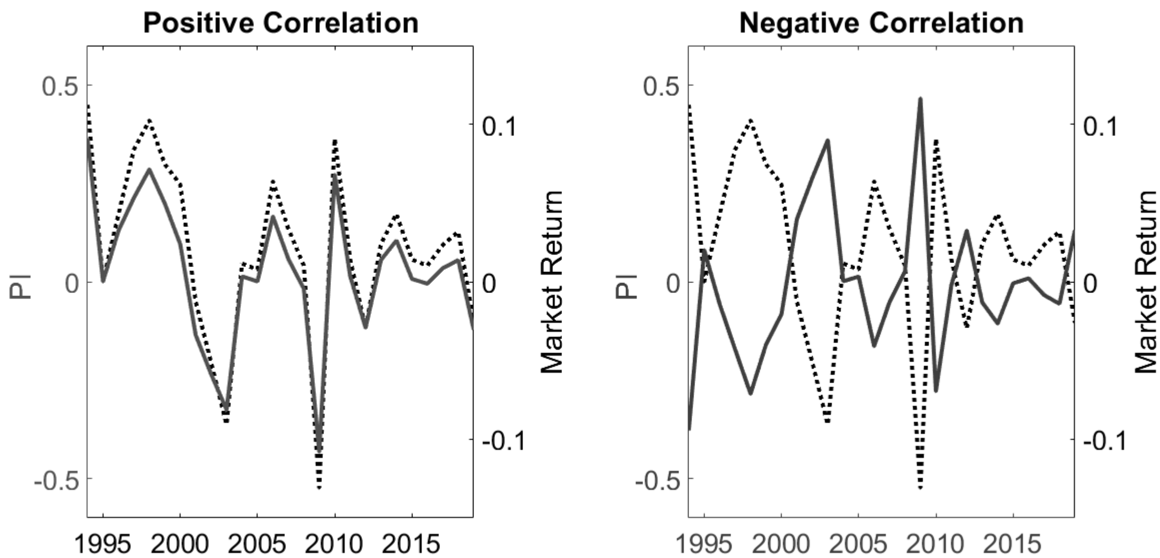
where,

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

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

Figure 5. Individual permanent income shocks
The left panel compares the dynamics of stock market returns (dotted line) and permanent labor income (PI) shocks reconstructed using the Kalman filter (solid line) 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 (dotted line) with the PI shocks reconstructed using the Kalman filter (solid line) 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.

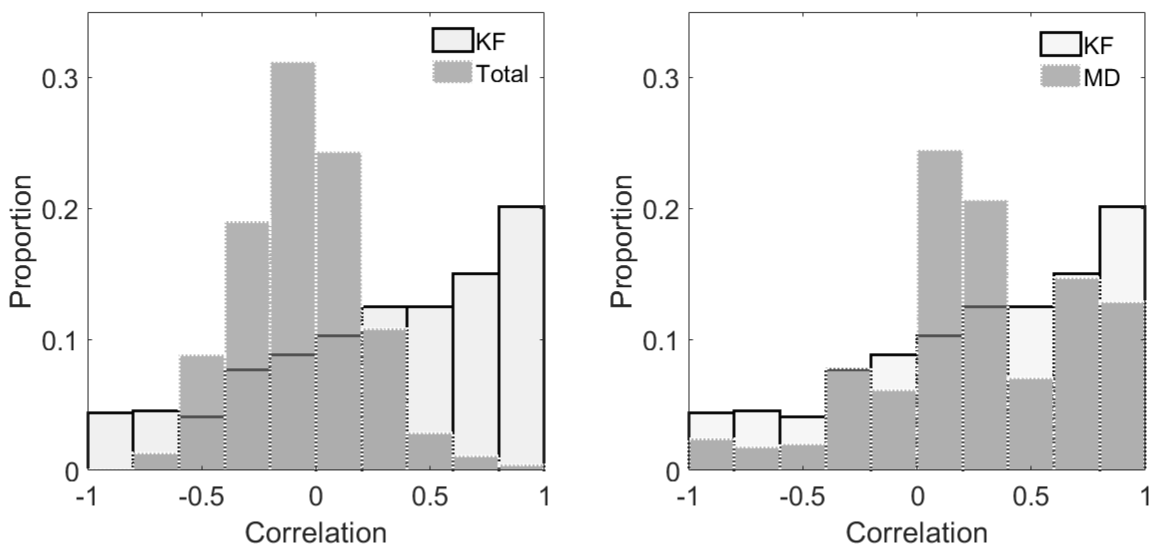


shocks by computing the first differences of the random walk. Consequently, for each individual, we obtain the dynamics of the permanent shocks to the labor income over the entire time series.

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

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.



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, as in (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 8 reports results concerning the decision to invest in stocks either directly or through mutual funds. 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.

[Table 8 about here.]

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 MD 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 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.

The statistical significance of this last result becomes weaker when we allow for stock returns that respond not only to aggregate risk but to noise, as well. However, revised correlations retain explanatory power when MD correlations with stock returns are omitted, with a large coefficient (-0.281) that remains non-negligible (-0.163) when MD correlations are inserted (see Table 23 of the Online Appendix).

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 9 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 9 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 MD 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 ($(-0.6 - 0.5) \times 0.11 \times 10$).

These results are free from unobserved heterogeneity concerns that may plague cross-sectional results. They support both our approach to the measurement of hedging needs and the role of correlation implied by life-cycle theory.

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. Last but not least, Section 5.3 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 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 10 and the associated figure displaying the distribution of cluster-based correlations confirm both the high average MD correlation parameter estimates (0.3) and the marked shift to the right of the distribution.

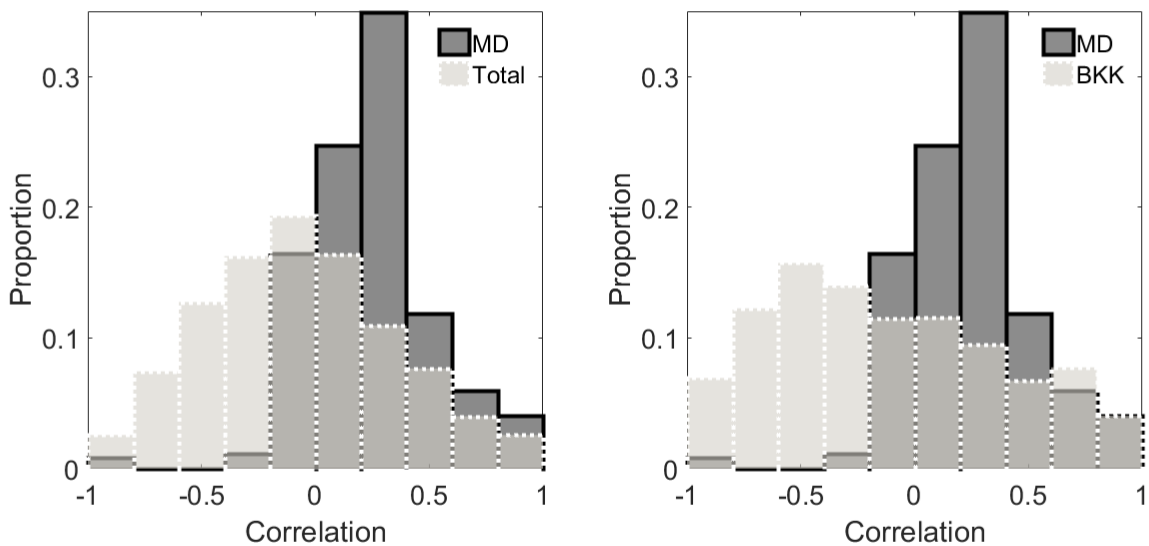
[Table 10 about here.]

Table 11 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 11 about here.]

Similarly, results of benchmarking in Table 12 confirm the relative strength of the MD method. In the sample including waves up to 2011, the BKK-correlation estimates with PI shocks have

Figure 7. Distribution of cluster-based correlations
The figure's left panel compares—after clustering the individuals in 80 homogeneous groups—the distribution of the estimated correlations between the permanent labor income shocks and the aggregate shocks estimated through the minimum distance (MD) methodology with the empirical distribution of the 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.



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 MD 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 12 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 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 correlation parameter estimates at the cluster-level, using data up to 2011.

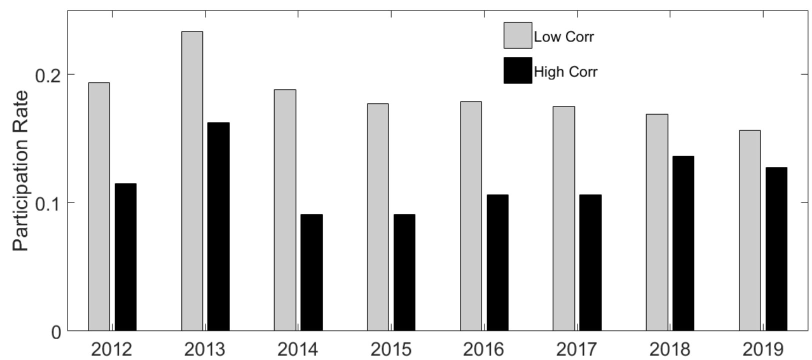
When considering data up to 2019, we allocate each new survey participant to one of the 80 clusters (defined in Section 3.1) on the basis of their personal characteristics. We then attribute to each individual the correlation parameter of the corresponding cluster estimated in the previous step using data up to 2011. That is, we use *ex-ante* measures of future income risk, as suggested by Guiso, Jappelli, and Terlizzese (1996).

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.

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 aggregate shocks and PI shocks using the minimum distance (MD) methodology 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.



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 parameter assigned to individuals 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 13 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) is 8% less likely to participate in the stock market than an individual allocated to a low correlation cluster (see equation (15) in Section 4.1).

[Table 13 about here.]

5.3 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 15, where the dependent variable refers to ownership of equities or mutual funds. Results are comparable to the ones in Table 11-12, as the sample covers the same years and the main independent variables are the same.[14] In the probit 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 14. 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.[15]

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

[15] Let us note that when we use only three education clusters to compute β_{CV}, we obtain values that are very in

[Table 14 about here.]

[Table 15 about here.]

6 Summary and Conclusions

This paper proposes a parsimonious model to assess the individuals' hedging motives associated with aggregate shocks, considering that aggregate shocks drive both stock market returns and differently affect groups of individuals. Our estimates of the mean individual exposure to aggregate shocks lie in the range 0.2-0.3. This result appears both in the Dutch Household Survey and in the US PSID data, despite the non-homogeneous groups of individuals covered by these surveys, and is insensitive to the length of the sample. Such exposures to aggregate risk exceeds on average the ones estimated without clustering, also in prior research. This difference supports the view that clustering corrects a bias (Guvenen et al., 2017).

Since stock returns co-move with aggregate shocks, the individual with the mean correlation hedges labour income risk by reducing exposure to the equity market with respect to an individual without exposure to aggregate risk. Such reduction exceeds the one implied by the only recent study that indicates the presence of hedging motives related to correlation of permanent income shocks with stock returns. This implies that observed portfolios are closer than previously thought to the implications of portfolio choice theory, without resorting to alternative mechanisms.

Results concerning equity investments support our measurement of hedging motives, that exploits not only the time series dimension of the data, as in prior research, but also the cross-sectional one. Not only the sign but also the size and the precision of the estimated effects confirm the theoretical prediction that equity investments fall when permanent income is exposed to aggregate risk. 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. Since earnings risks is a central issue in the economics of incomplete markets, these advances in the measurement of hedging motives may prove useful beyond the boundaries of financial economics.

line with those estimated by Campbell and Viceira (2002) on the three education groups.

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.

Bagliano, F.C., C. Fugazza, and G. Nicodano, 2019, Life-cycle portfolios, unemployment and human capital loss, *Journal of Macroeconomics* 60, 325–340.

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.

Calvet, Laurent E, and Paolo Sodini, 2014, Twin picks: Disentangling the determinants of risk-taking in household portfolios, *The Journal of Finance* 69, 867–906.

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., 2022, Countercyclical income risk and portfolio choices over the life-cycle, *Review of Financial Studies.* 35, 4016–4054.

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, 2005a, Consumption and portfolio choice over the life cycle, *Review of Financial Studies* 18, 491–533.

Cocco, Joao F, Francisco J Gomes, and Pascal J Maenhout, 2005b, Consumption and portfolio choice over the life cycle, *The 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.

Galvez, J., and G. Paz Pardo, 2022, Richer earnings dynamics, consumption and portfolio choice over the life cycle, *Netspar Conference* 1–64.

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

Guiso, L., T. Jappelli, 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.

Guvenen, Fatih, and Anthony A Smith, 2014, Inferring labor income risk and partial insurance from economic choices, *Econometrica* 82, 2085–2129.

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.

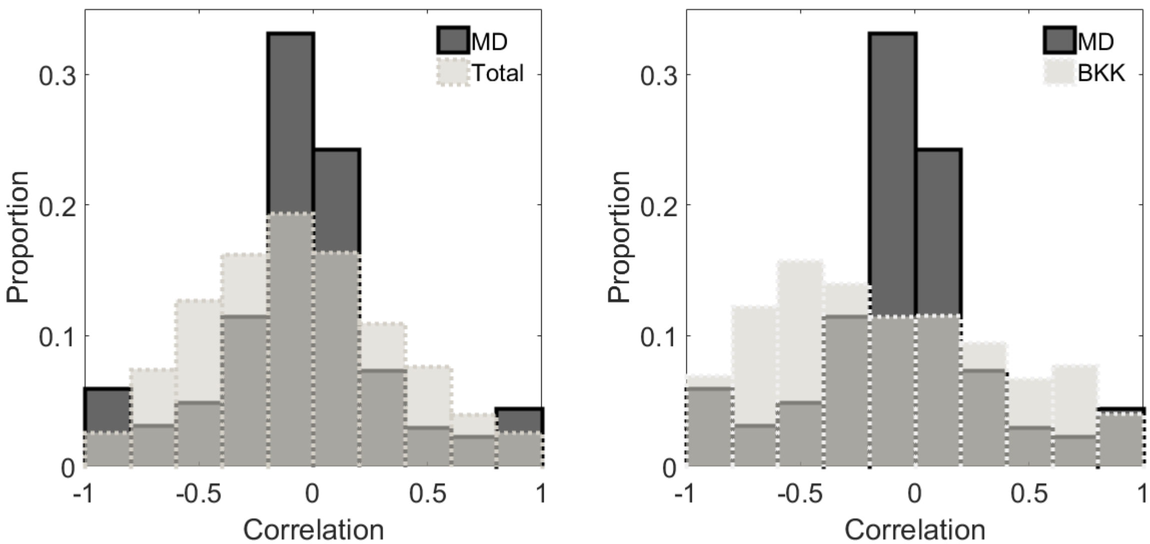
Shen, J., 2021, Countercyclical risks, consumption and portfolio choice: Theory and evidence, *University of Missouri* .

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* .

Figure 9. Distribution of individual correlations
The figure's left panel compares the distribution of the estimated correlations between the permanent labor income shocks and the aggregate shocks estimated through the minimum distance (MD) methodology, at the individual-level, with the empirical distribution of the 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.



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 9 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).

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 16. 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 16 about here.]

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 the aggregate shocks and the permanent component of labor income shocks estimated by using the minimum distance methodology (MD) both at the individual and the cluster levels, as well as 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 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, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), 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). 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)			-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 Rˆ2	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 correlation between the permanent component of labor income shocks and the aggregate 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)		-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, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), 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). 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)			-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 correlation between the permanent component of labor income shocks and the aggregate 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.

	PropSTKMF					
	(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)		-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 Rˆ2	0.366	0.368	0.289	0.292	0.289	0.292

Table 8 Probit Estimates for Participation Revision

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 either directly or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation, the correlation between permanent income shocks and aggregate shocks, the correlation between stock market return and different specifications of labor income shocks as described in table 4, and different measures of the relationship between labour income growth rate and stock market returns used in previous papers as described in table 5. The correlation and the beta coefficients used here as independent variables are computed on an expanding time window going from the first wave up to the t-th wave, where t goes from 10 to the 19 (the last wave). 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 Kalman filter at the individual level. Then, we compute the yearly change in the correlation between these PI shocks and the stock market returns, where the first change is computed with respect to the clustered correlation estimated using minimum distance (KFexp80). 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					
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.207***	-0.206***	-0.202**	-0.203***	-0.187***	-0.187***
	(0.046)	(0.046)	(0.043)	(0.057)	(0.060)	(0.042)
Corr(Tot,Rm)	-0.039					
	(0.032)					
Corr(Det,Rm)		0.064				
		(0.034)				
Corr(Stoc,Rm)		-0.067*				
		(0.042)				
Corr(PermMD80)			-0.288***			
			(0.037)			
Corr(KFexp80,Rm)			-0.270**			
			(0.131)			
Corr(tranBKK,Rm)				-0.031		
				(0.034)		
Corr(permBKK,Rm)				-0.047		
				(0.036)		
Beta (Guvenen et al. (2017))					0.068	
					(0.061)	
Beta (C&V(2002))						-0.080
						(0.066)
Controls	YES	YES	YES	YES	YES	YES
N	9,345	9,299	10,423	7,586	10,423	10,423
Pseudo Rˆ2	0.277	0.277	0.285	0.280	0.272	0.272

Table 9 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 permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), the correlation between stock market return and different specifications of labor income shocks as described in table 4, and different measures of the relationship between labour income growth rate and stock market returns used in previous papers as described in table 5. 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 table 4 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)			-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 10 Correlation Parameters from DHS. Summary Statistics on Short Sample. The table reports summary statistics of the estimated correlations between the aggregate shocks and the permanent component of labor income shocks estimated by using the minimum distance methodology (MD) both at the individual and the cluster levels, as well as 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 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 11 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 stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), 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). 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. 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)			-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 R^2	0.259	0.262	0.264	0.264	0.265	0.263

Table 12 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 stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the correlation between the permanent component of labor income shocks and the aggregate 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) (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 table11 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)		-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 13 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 the PI shocks and the aggregate shocks estimated at the cluster-level using the minimum distance methodology (permMD80) at the cluster-level 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)	-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 14 Correlation and Beta Parameters. Summary Statistics (PSID)
The table reports summary statistics of the estimated correlations between the aggregate shocks and the permanent component of labor income shocks estimated by using the minimum distance methodology (MD) both at the individual and the cluster levels, as well as 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 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 15 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 stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), 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). 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)			-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 16 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 aggregate shocks 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