Hedging Permanent Income Shocks.*

and 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: December 8th, 2023

Abstract

This paper solves the long-standing non-participation puzzle by predicting individual participation to the equity market, also out-of-sample and for the same individual over time. It connects non-participation to co-movements of individual income shocks with an aggregate shock (or, equivalently, with stock market returns), thereby supporting the hedging motives explanation for non-participation. Such new insight owes to the large number of income shocks co-movements across clusters of individuals, which we exploit to identify the parameters driving individual hedging motives.

Keywords: Stock market participation, Earnings Co-Movements, Aggregate shocks

JEL Classification: G10,G11, D14, C15

*We are grateful to Alex Michaelides, Laurent Bach, Marie Brière, Laurent Calvet, Bernard Dumas, Davide Fiaschi, Christian Gollier, George Korniotis, Lorenz Kueng, Roberto Marfé, Ignacio Monzon, Henriette Prast, Theo Nijman, Alberto Quaini, Olga Goldfayn-Frank, as well as participants in the 2023 Paris Financial Risks International Forum, the 2023 ZEW Mannheim Conference on Ageing and Sustainable Finance, the 2022 CEPR European Conference on Household Finance, the 2022 European Meetings of the Econometric Society, 2021 CeRP Workshop, the 2021 Inquire Europe webinar, the 2020 Consob ESMA Bocconi webinar, the OEE Conference 2020, 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), and seminars at the European Central Bank, Imperial College London, the University of Pisa and USI for helpful suggestions. We are also grateful to the Institute for Quantitative Investment Research (INQUIRE) Europe, the Observatoire de l'Epargne Européenne, and the Italian MIUR ("Department of Excellence" 2018–2022, 2023-2027) for funding.

1 Introduction

According to portfolio theory, 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 on such individual hedging as a driver of observed equity portfolios is weak (see e.g. Heaton and Lucas (2000), Campbell, Cocco, Gomes, and Maenhout (2001), Campbell and Viceira (2002), Cocco, Gomes, and Maenhout (2005), Angerer and Lam (2009), Bonaparte, Korniotis, and Kumar (2014), Catherine, Sodini, and Zhang (2020)). Importantly, hedging motives appear unable to account for non-participation in the stock market. - a phenomenon known as the stock market participation puzzle (Haliassos and Michaelides, 2003). [1]

This paper solves this disconnect between portfolio theory and the data by predicting individual participation to the equity market, over different data sets and their sub-samples, also out-of-sample and for the same individual over time. Our results robustly connect non-participation to the joint co-movements of several individual income shocks with an aggregate shock or the stock market returns. This contribution stems from an innovation in both the modelling and the estimation of the earnings processes. Specifically, we let the labor income shocks of any two individuals co-move, due to their noisy exposure to an aggregate shock (or, equivalently, the stock market return). Such co-movement

[1] Alternative mechanisms suggested to explain non-participation include a fixed participation cost, the degree of trust in the stock market, liquidity differences between stocks and bonds, ambiguity aversion, a stochastic interest rate possibly correlated with earnings shocks, mean reversion in stock market returns. See the recent survey in Gomes, Haliassos, and Ramadorai (2021).

is proportional to their individual correlation coefficient between permanent income (PI) shocks and aggregate shocks (correlation from now on). A minimum distance (MD) estimation can then exploit such co-movement information to identify correlations and other parameters. Extant models and methods instead measure the correlation over time between each individual's earnings growth and the stock market returns, ignoring co-movement information. We otherwise follow prior literature in letting groups of individuals display heterogeneous exposure to aggregate risk (as in Campbell and Viceira (2002), Guvenen, Schulhofer-Wohl, Song, and Yogo (2017)) and distinguishing between permanent and transitory shocks to labor income (see Carroll and Samwick, 1997 and Cocco et al., 2005). We abstract from non-linear income risk (see e.g. Guvenen, Karahan, Ozkan, and Song (2021), Guvenen, Pistaferri, and Violante (2022)) since it cannot explain non-participation, while being able to account for the flat life-cycle profile of conditional equity holdings (Bagliano, Fugazza, and Nicodano (2019), Catherine (2022), Galvez and Paz-Pardo (2022), and Shen (2023)).

Our empirical analysis uses as baseline data the Dutch National Bank Household Survey (DHS), an annual panel with rich information about both individual characteristics and portfolios, and the US Panel Study of Income Dynamics (PSID) for robustness. In our full DHS sample (27 yearly observations for 1884 individuals) we rely on over 1.7 million conditions to identify the 80 correlation coefficients characterizing the agents' clusters. Such MD correlations between PI shocks and the aggregate shock average 0.2-0.3, or above, in both Dutch and U.S. data, implying a shift to the right of the distribution relative to the ones obtained with the time-series methods in Bonaparte et al. (2014),Campbell and Viceira (2002) and Guvenen et al. (2017).

To validate this new approach, we compare these alternative estimates in several ways based on their ability to shed light on hedging-based participation. The results support the innovation. First, the observed participation rate decreases in estimated correlations based on our MD method, while it is mostly flat when based on prior time-series methods. Second, the MD estimates retain predictive power for both participation and the equity portfolio shares in the Probit and the Tobit analyses proposed in (Bonaparte et al., 2014). On the contrary, prior estimates do not when MD estimates are present. Moreover, individual choice is twice as sensitive to the MD than to other correlation estimates.

As a second check, we examine shorter samples ($T = 18, 14$) in both the DHS and the PSID. The MD correlation estimates retain predictive power, while the ones based on prior methods no longer predict participation with one exception in DHS. The robustness of our approach in shorter samples owes to cross sectional information on top of the time series one. Importantly, such robustness makes it possible to implement an out-of-sample analysis that was never undertaken despite its relevance for portfolio management. Even out-of-sample, MD correlation estimates predict participation. In sum, these results robustly connect portfolio investments to co-movements of earnings shocks and aggregate shocks for the first time. They also imply larger hedging motives than previously thought and explain non-participation in equity markets.

Our third check exploits the panel dimension of the data to address the concern that, in cross-sectional studies, both the individual's earnings risk and portfolio choice may be driven by unobserved characteristics. Based on our simple parametric model for PI shocks We compute revised correlation for each individual over time, based on the dynamics of their PI shocks which is reconstructed using a Kalman. We then relate changes in individuals' risk-taking decisions over time to their revised correlations. The results show that the decision to remain in the market is informed by both this revised correlation and the initial exposure to aggregate risk.

A final worry is that results wash away when measuring correlations between PI shocks and stock market returns, since the latter are a combination of the aggregate risk factor and idiosyncratic noise. All results instead go through. In sum, this paper consistently overturns the commonly held view (see e.g. recently Catherine et al. (2020)) that heterogeneous individual exposures to aggregate risk or stock returns do not matter for observed portfolios.

Our paper contributes a solution to the non-participation puzzle thanks to correlation estimates based on a new approach. In portfolio theory, a positive correlation coefficient between earnings shocks and stock market returns is needed to generate a negative hedging demand and a (below) zero optimal investments in equities (see e.g. Campbell and Viceira (2002) and Bagliano et al. (2014)). Yet, the empirical evidence on the size of and the sensitivity to correlation is weak, also compared with the one regarding earnings volatility (Fagereng, Guiso, and Pistaferri (2018),Angerer and Lam (2009), Betermier, Jansson, Parlour, and Walden (2012)), making equity market non-participation a puzzle. Indeed, in Campbell and Viceira (2002) the correlation between labor income shocks and contemporaneous stock returns is too low (0.06-0.1), even if the one with lagged stock returns is higher (0.32-0.5). Similarly, the correlation is negligible in Cocco et al. (2005) and also within educational groups in Munk and Sorensen (2010). In Angerer and Lam (2009), the relationship between covariance measures and the risky asset share is insignificant both statistically and economically. In Massa and Simonov (2006), hedging motives cannot explain portfolio tilts away from the market portfolio. In Arrondel, Pardo, and Oliver (2010), a proxy for both correlation and earnings uncertainty explains the decision to hold risky assets for a subset of households, 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. Bonaparte et al. (2014) show that sample correlation between labor income shocks and stock market returns predict stock market participation. A concern is however that the estimated correlations based on past income shocks may not reflect future hedging needs (Catherine et al. (2020)). Our new method is immune from this problem, as it uses both 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 income growth co-movement between pairs of other agents. Such enhanced information allows the cluster-based MD correlation estimates to robustly predict non-participation for the first time.

The need of a long time series of earnings data to identify PI shocks is well known (e.g., Carroll, Hall, and Zeldes, 1992; Carroll, 1997; Meghir and Pistaferri, 2004). Guvenen (2009) circumvents such requirement with a parsimonious model, relative to the richer one in Guvenen et al. (2021), to get MD estimates of the persistence of income shocks. We estimate the parameter needed to sign the hedging demands for stocks, the correlation, through a parsimonious model that restricts other parameters including the variance of the PI shocks. Unreported analysis shows that the same results obtain without both such variance restriction and the split between permanent and transitory income shocks. What matters to our results are both the clustering and the MD estimation of co-movements. The latter explains the robustness of our results relative to alternative methods. Clustering matters for the shift towards positive values of the correlation distribution. In both the DHS and the PSID, the mean MD correlation between PI shocks and the aggregate shock is above

0.2 when we group individuals, while it is small (around 0.05) when we do not. This evidence also indicates that the shift in the cluster-based correlation does not depend on the parametric restrictions imposed by our model, since both individual and cluster-based MD estimates rely on the same restrictions. At the same time, the size of the mean MD correlation estimates does not depend on the use of specific clusters, since it appears in both databases that provide different personal information. Thus our results support the view, in Guvenen et al. (2017), that traditional estimation of correlations underestimates systematic risk by ignoring the heterogeneous exposure of workers' clusters. The multi-factor versus single factor representation of aggregate risk does not instead appear critical. In an unreported robustness check, we estimate the loading associated to each principal component of the covariance matrix of individuals' income growth. The high correlation (0.6) between the MD correlation estimates and the sum of such loading indicates a close relationship between the non-parametric multi-factor representation of aggregate risk and our single factor one.

The rest of the paper is organized as follows. After presenting our model (Section 2), we estimate correlations with several methods (Section 3). We then examine their ability to predict participation in cross sectional regressions performed over different samples (Section 4). In Section 5, we update correlation estimates and study participation revisions of each individual over time. Section 6 repeats the analysis focusing on co-movements between earnings shocks and stock market returns instead of the latent aggregate shock. Conclusions follow. The Appendix presents the details of our estimation strategy. The Online Appendix examines the implied dynamics of PI shocks for different clusters, and the sensitivity of optimal life-cycle portfolios based on our correlation estimates.

2 Individual Exposure to Aggregate Shocks: A New Approach

This section presents a parsimonious framework of the joint distribution of earnings, a latent aggregate shock and the stock market returns. We model income growth co-movements across pairs of agents, arising from their individual-specific exposure to aggregate risk. This novelty considerably increases the number of moment restrictions relative to the number of parameters to be estimated. In other words, the income co-movements between pairs of the other $N-1$ agents help identify the exposure of the N-th agent to aggregate risk. We also impose other common restrictions, e.g. both a zero inter-temporal covariance of the idiosyncratic transitory shock for each agent and a zero covariance of the idiosyncratic component across different agents. We then explain how we compute the sample counterparts of the implied moment conditions to perform a minimum-distance (MD) estimation of the unknown model parameters.

2.1 A Parsimonious Model of Income Shocks Co-Movements

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. (2005), 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 now depart from prior work, assuming that that $W^p_{i,t}$ is correlated with the aggregate shock W_t:

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

where ρ_i denotes the individual correlation coefficient and W_t and $W_{i,t}$ are two standard normal random variables. We can therefore express the PI shock $u_{i,t}$ as the sum of a systematic component, $\xi_{i,t}$, and an idiosyncratic component, $\omega_{i,t}$:

$$u_{i,t} = \xi_{i,t} + \omega_{i,t},$$

where

$$\xi_{i,t} = \sigma_u(\rho_i W_t) \sim \mathcal{N}(0, \sigma_u^2 \rho_i^2),$$

and

$$\omega_{i,t} = \sigma_u \left(\sqrt{1-\rho_i^2} W_{i,t} \right) \sim \mathcal{N}\left(0, \sigma_u^2(1-\rho_i^2)\right). \tag{4}$$

Thus the PI shock is a linear combination of two normally distributed random variables:

$$u_{i,t} = \sigma_u \left(\rho_i W_t + \sqrt{1-\rho_i^2} W_{i,t} \right) \sim \mathcal{N}(0, \sigma_u^2 \rho_i^2 + \sigma_u^2(1-\rho_i^2)), \quad \text{i.e.} \quad u_{i,t} \sim \mathcal{N}(0, \sigma_u^2) \tag{5}$$

Thus, the variance of the PI shocks is the sum of systematic and idiosyncratic variances, where the relative weight of the two components is given by the correlation between PI shocks and the aggregate shocks. We finally assume that also the stock market return is 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. While the main body of the paper focuses on the case of a negligible variance of the noise, Section 6 allows for positive variance.

2.2 Implied Moment Restrictions and Estimation Approach

Two sets of model-implied moment conditions are used in the estimation of the model parameters. The first set, derived in Appendix A.1, pins down the variances of both the transitory and the PI shocks from the observable inter-temporal covariance of the time variation in total income shocks (DTS). [2] The second set, derived in Appendix A.2, exploits the dependence of each individual's PI shock on the aggregate risk factor to compute the covariance between any two individuals' DTS as a linear function of the two individuals' correlation coefficients. By doing so, we obtain $N(N-1)/2$ conditions that will allow to estimate the N correlation coefficients.

Next, we obtain the sample counterparts of these model-implied moment conditions in Appendix A.3.

We can now formalize the MD estimator of the vector θ, which contains the unknown parameters of the labor income process:

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

[2] The dimension of this set would be equal to $T(T-1)/2$ if we only relied on the familiar inter-temporal restrictions on the DTS over one period. Instead, we also consider the DTS over more than one period thereby enlarging the number of moment conditions.

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

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

Similarly, we denote the set of M empirical counterparts as $\{g_m\}_{m=1}^{M}$, 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:

$$Q(\theta) = (g_M - G_M(\theta))' I_M (g_M - G_M(\theta)) \tag{6}$$

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 a MD estimator that weighs moments with an identity matrix is asymptotically consistent and normal.

Because we have $N+2$ unknown model parameters and $M > N+2$, the model is (highly) over-identified. Specifically, we use an overall set of moment conditions of size equal to:

$$M^n = M^{int} + M^{crs} = M(D, T, N),$$

that is a function of the D inter-temporal lags, the length of the time-series, T, and the number of individuals in the sample, N, as specified in equations (13) and (14).

3 Measuring Individual Exposure to Aggregate Shocks

Our baseline data set derives from the DNB Household Survey (or DHS), which reports annual labor income for a representative sample of the Dutch population over 1993-2019. It provides information on financial investments of 1884 individuals throughout the sample, along with their detailed personal characteristics. 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.]

Table 1 provides details on the variables used in our analysis and the descriptive statistics for our sample. The average age is around 56, half of the individuals 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. [3] 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.

3.1 Cluster-Based Correlation Estimates

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 alternative methods in previous literature. These include the moving average approach in Bonaparte et al. (2014) where the individual PI shock at time t is the equally-weighted average of the own realized labor income growth at time $t-1$, t, and $t+1$; and the beta coefficients of the regression of the

[3] The DHS provides no information on either direct holdings of individual stocks or indirect holdings through pension funds.

labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002) at the cluster level. We obtain the former, 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. [4] To obtain the latter, 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).

When the correlation coefficient is cluster-specific rather than individual-specific, equation (5), describing the individual PI shocks, becomes:

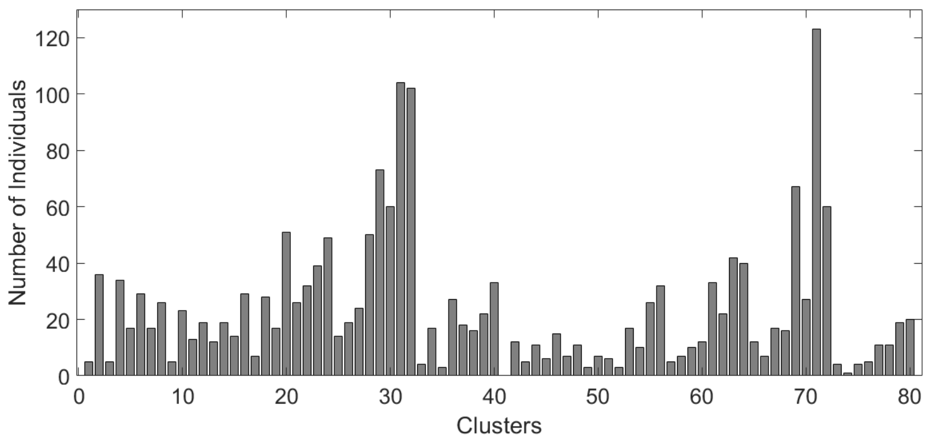
$$u_{i,t} = \sigma_u \left(\rho_k W_t + \sqrt{1-\rho_k^2} W_{i,t}^p \right),$$

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

To ensure consistency with the model, the clustering variables must be stable because the cluster-specific correlation does not change over time. We therefore 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. *Urbanization* is a dummy variable equal to 1 if the individual lives in an urban area. *Risk aversion*

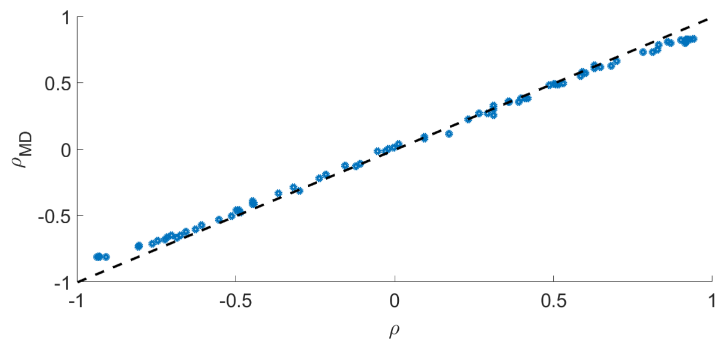
[4] We cannot account for exposure to both employer- and industry-level risk when using the DHS because of data availability.

Figure 1. Number of individuals in each cluster
The figure displays the number of individuals belonging to each of the 80 clusters, which group individuals based on combinations of the following personal traits: Education (5 groups), Sex (2), Urbanization (2), Risk aversion (2), and Financial literacy (2). The data are from DHS, 1993–2019.



is a dummy variable equal to 1 if the individual displays high risk aversion, identified as a value of the DHS risk aversion variable greater than 5. *Financial Literacy* is a dummy variable equal to 1 if the individual reports being knowledgeable with respect to financial investing. Each survey wave records these variables for each individual. While they are mostly constant over time, we input the value of the mode in case of time variation. 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). Figure 1 shows how individuals are distributed into their 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.Then, with 27 yearly observations ($T = 27$), 1884 unique individuals ($N = 1884$) and using at most 18 inter-temporal lags ($D = 18$), we obtain $M^n = M(18, 28, 1884) = 1,776,662$ moment conditions (see equations 13 and 14) to estimate 80 cluster-specific correlation coefficients, the volatility of the transitory shock and the

Figure 2. Minimum Distance estimation. Monte Carlo Simulation
The figure displays results from Monte Carlo simulation on the Minimum Distance (MD) estimator. We simulate a panel of individual income shocks using a time series of 30 observations (T=30), 1880 individuals (N=1880), and 80 clusters (K=80). To generate income shocks, we use the model equations and the following set of parameters: σ_u=0.1, σ_ϵ=0.3, and a random set of K clustered correlation coefficients ρ from a uniform distribution defined on the support $[-1, 1]$. Then, we apply the MD estimator to the simulated data to infer the model parameters. In the figure, we report on the x-axis the true, arbitrary set of ρ and on the y-axis the ρ_{MD} estimated using MD. We obtain ρ_{MD} as the average across 100 Monte Carlo simulation results. The dashed line is the 45-degrees line representing perfect inference.

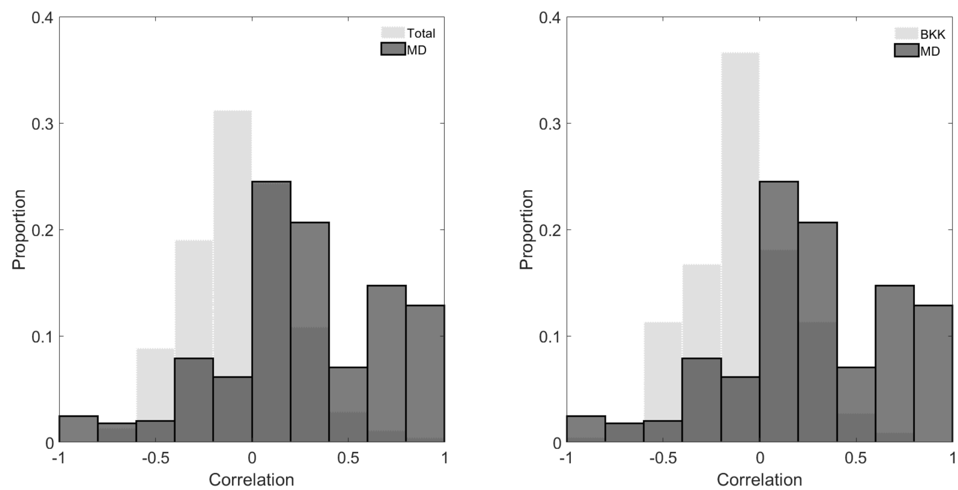


one of the permanent shock.

As a first check, we test the accuracy of our estimation methodology using standard Monte Carlo simulation techniques. We generate a panel of simulated, individual income shocks using the model equations and an arbitrary set of model parameters. In particular, we draw a random set of K clustered correlation coefficients from a uniform distribution defined on the support $[-1, 1]$. For consistency with the actual data, we simulate a time series of 30 observations (T=30), for 1880 individuals (N=1880) grouped into 80 clusters (K=80). Then, we apply the MD estimator to the simulated data to infer the model parameters. In Figure 2, we show that the MD xcluster-based correlations match the true, arbitrary set of correlation parameters.

In Figure 3, we display the distribution of the cluster-based Minimum Distance correlations (MD)

Figure 3. Distribution of cluster-based correlations
This figure displays the distribution of the cluster-based MD correlations between the permanent labor income shocks and the aggregate shocks. It compares it in the left-hand panel with the distribution of the sample correlations between stock market returns and total income shocks (Total); and in the right-hand panel 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 DHS, 1993–2019.



against the one of sample correlations between stock market returns and total income shocks in the left-hand side panel. The right-hand panel plots the MD correlations against the correlations between stock market returns and PI shocks obtained with the approach of Bonaparte et al. (2014). The latter distribution is widely dispersed around a slightly negative mean, like the one between total income shocks and stock returns. In contrast, the distribution of MD correlations is heavily concentrated around positive values, with the proportion of individuals characterized by a negative correlation between PI shocks and stock market returns dropping from 36%, when estimated using the approach of Bonaparte et al. (2014), to 11%. When we turn to summary statistics of the correlation parameter estimates in Table 2, such striking results become stronger. The distribution of MD correlation parameters shifts to the right with respect to all alternatives. The mean MD correlation (0.257) indicates that the PI shocks of most agents are positively exposed to the aggregate shock. This represents a radical change with respect to previous mean correlation estimates, that are either slightly negative (-0.091 as in Bonaparte et al. (2014)) or close to zero (0.004 as in Guvenen et al. (2017) and 0.006 as in Campbell and Viceira (2002)).

[Table 2 about here.]

Before proceeding, let us analyze the possible causes for such a marked difference in the distribution of estimated correlations. First, our focus is on correlations with an aggregate shock rather than a stock return which has both an aggregate and an idiosyncratic component. Section 6 addresses the latter extension revealing that the MD correlations between PI shocks and a noisy stock return are similar to the ones reported in this section, with a mean value of 0.189 as opposed to 0.257 (see Table 11).

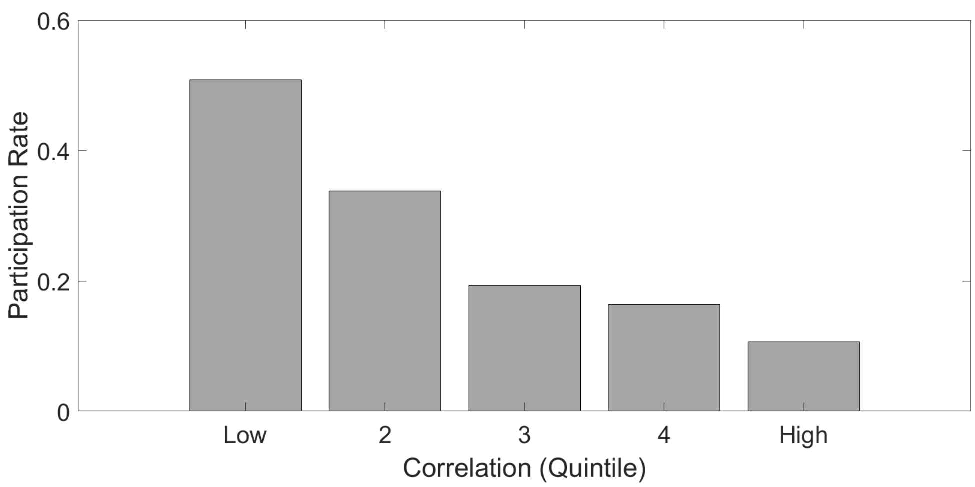
Second, the sample period may be peculiar. In a shorter sample ($1993-2011$) for the Dutch economy the mean MD correlation estimate is even larger (0.310) while estimates based on previous methods are unchanged (see Table 5).

Third, the shift in the distribution of estimated correlations may be due to the restrictions imposed by our parsimonious parametric model for PI shocks. This is not the case. In Table 2 we report the same distribution for individual, non cluster-based, correlations that are based on the same parametric model. The mean correlation drops from 0.257 to 0.057. Thus neither noisy stock returns, nor the sample period, nor simplifying restrictions explain the shift to the right in the distribution while clustering does matter. As suggested by Guvenen et al. (2017), ignoring the differential exposure across clusters of workers to aggregate risk leads to underestimating that exposure. It therefore misinterprets the residual from the wage regression as purely idiosyncratic when in fact it contains aggregate risk. 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)). Of course, clustering also reduces the number of unknown parameters, thereby increasing efficiency. This explains the lower dispersion in the distribution of cluster-based MD correlations relative to the one estimated with the method in Bonaparte et al. (2014). It is however the shift to the right in the distribution of correlations that opens up the possibility to explain observed non-participation in the equity market through hedging motives (see Bagliano et al. (2014), Gomes et al. (2021)), since individuals with positive correlation have a negative hedging demand for stocks according to portfolio theory.

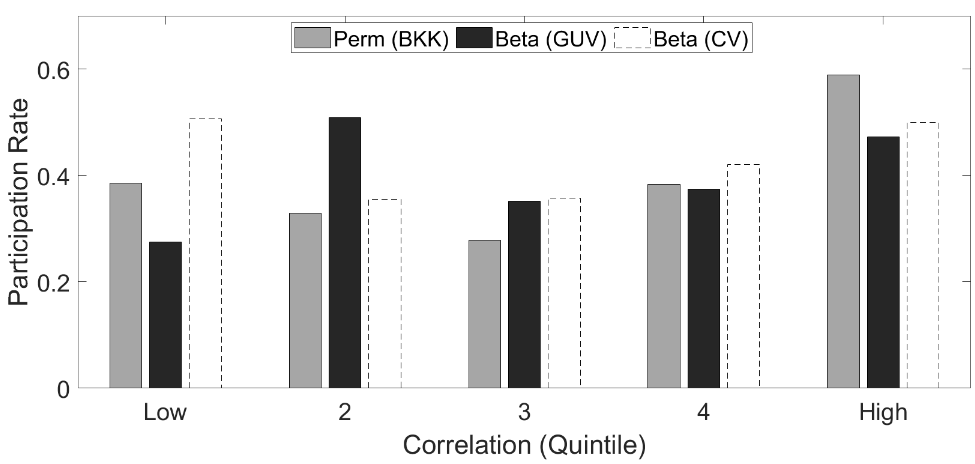
A reader might however expect such shift to disappear with a change in clustering characteristics. When we use PSID data and the same clusters as in Guvenen et al. (2017), the mean correlation estimate based on their method (0.232) gets closer to the mean cluster-based MD correlation (0.268) and all others mean correlation estimates increase. This evidence on the one hand confirms the importance of clustering for capturing systematic risk components. On the other hand, the added robustness of the MD estimates across different model specifications, samples and clustering characteristics stems from our proposed approach. Thus, it appears that the restrictions based on the covariance matrix of contemporaneous income innovations capture co-movements of individual PI shocks with the aggregate shock that instead escape methods relying on the time series of

Figure 4. Stock market participation: sorting on correlation.
This figure reports the average stock market participation rate, i.e. the mean value of the dummy variable describing individual's stock market participation, in each estimated correlation quintile. We sort individuals into quintiles according to either the minimum distance cluster-based estimates (MD) of correlation between aggregate shocks and PI shocks (**Panel A**); the correlation between stock market returns and the permanent earnings shocks estimated according to Bonaparte et al. (2014) (*Perm (BKK)*); the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) (GUV) and Campbell and Viceira (2002) (CV) at the cluster-level (**Panel B**). Low (high) is defined as the bottom (top) quintile of the estimated correlation. The data are from the DHS, 1993–2019.

Panel A. MD Estimates



Panel B. Benchmarking



individual income shocks, only. The shift to the right of the correlation distribution is a necessary condition to explain non-participation in the equity market through negative hedging demands for stocks. However it is not sufficient. [5] We therefore display the relationship between individual stock market participation observed in the data against their individual exposure to the stock market return (or the aggregate shock). We sort individuals into quintiles according to the estimated correlation. Then we compute the stock market participation rate in each quintile as the mean value of the dummy variable $\mathcal{I}_{i,t}$, which takes a value equal to 1 if the individual i reports investing in the stock market at time t, and zero otherwise. Panel A in Figure 4 highlights a consistent, monotonic pattern across MD80 correlation quintiles: the participation rate declines substantially between the low- and the high-correlation quintiles, in line with the income hedging hypothesis. This is not the case when considering, in Panel B, the other measures of correlation between the stock market returns and alternative specifications of labor income shocks.

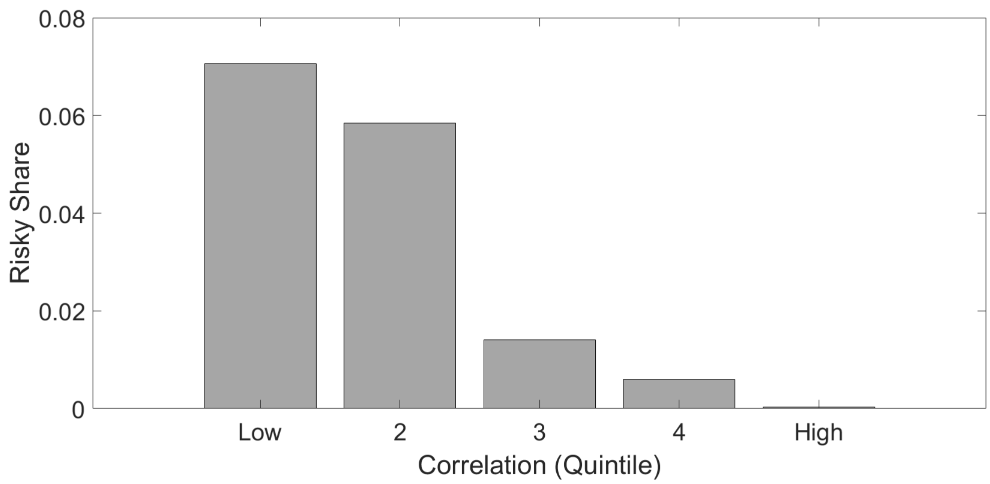
To gain additional insight, we similarly compute in each correlation quintile the average share of the individual's portfolio invested in stocks either directly or through mutual funds. Figure 5 supports again a clear, negative pattern across correlation quintiles: the average risky share of the financial portfolio decreases substantially between the low- and the high-correlation quintiles. Once more, this is not the case when considering different measures of correlation between the stock market returns and alternative specifications of labor income (Figure 5).

This descriptive evidence encourages studying the variation of participation explained by individual exposure to aggregate risk in multivariate regressions.

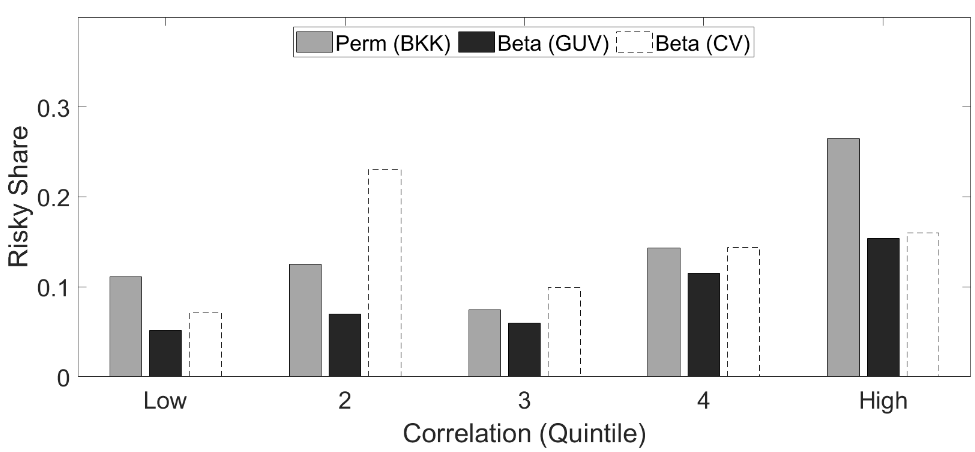
[5] The Online Appendix shows the endogenous emergence of stock market non-participation in a calibrated life-cycle model when correlation is positive.

Figure 5. Asset allocation: sorting on correlation.
The figure reports the average risky share in each estimated correlation quintile. We sort individuals into quintiles according to either the correlation between aggregate shocks and PI shocks estimated using the minimum distance (MD) methodology (**Panel A**), or the correlation between stock market returns and different components describing labor income shocks: the total labor income growth rate (Tot), the transitory and the permanent shocks estimated according to Bonaparte et al. (2014) (*Perm (BKK)* and *Tran (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) (GUV) and Campbell and Viceira (2002) (CV) at the cluster-level (**Panel B**). Then we compute in each quintile the average share of the individual's portfolio invested in stocks either directly or through mutual funds, across individuals. Low (high) is defined as the bottom (top) quintile of the estimated correlation. The data are from the DHS, 1993–2019.

Panel A. MD Estimates



Panel B. Benchmarking



4 Hedging Exposure to Aggregate Shocks

For hedging motives to account for the observed individuals' non-participation in the equity market, estimates of their exposure to aggregate shocks should be positive. Furthermore, the variation in exposure across individuals should explain variation in both observed equity market participation and asset allocation. This section investigates whether the new measure of individual exposure to aggregate shocks, that satisfies the first requirement, also predicts variation in individual portfolio choice in multivariate regressions. We also benchmark existing alternatives even if they do not satisfy the first requirement based on the DHS baseline sample.

We analyze the decision to participate in the stock market in Section 4.1 and the asset allocation in Section 4.2. In each section, we use the same dependent variable. Apart from the individual correlation estimates, the control variables, that are inspired by the analysis in Bonaparte et al. (2014), are the same in all sections.

4.1 Hedging Motives and Participation

We estimate the probability of participating in the stock market through a Probit regression, where the dependent variable is the dummy $\mathcal{I}_{i,t}$ introduced in the previous section. The main independent variables are the standard deviation of the income growth rate and the metrics for exposure to aggregate shocks as well as.

[Table 3 about here.]

In Table 3, we first note that the cluster based MD correlations between PI shocks and the aggregate shock (column 1) display greater predictive ability than the correlations between stock market returns and the transitory [6] and the permanent shocks to labour income estimated as in Bonaparte et al. (2014), tranBKK and permBKK respectively, in column (2). A similar, but less extreme, pattern emerges with respect to the betas coefficients estimated at the cluster level following Guvenen et al. (2017) and Campbell and Viceira (2002), in columns (4) and (6) respectively. Second, all competing metrics lose their predictive power when they appear together with the MD correlations, in columns (3,5,7). The predictive power of MD estimates of correlation is instead unchanged.

The economic significance of hedging motives implied by the MD correlations is also remarkable, as their coefficient is large and above 2.6 in most cases. We can gauge the difference in the probability to invest in the stock market [7] for two individuals with ρ=-0.6 and ρ=0.5, respectively. These correspond to the 10-th and the 90-th percentiles of the sample distribution of the correlations between total income shocks and stock market returns. Using the estimated coefficient of MD correlations in column (1) of Table 3 the first individual's propensity to participate is 12% higher than the one of the second individual. Using instead the coefficient in column (2), measured with the approach described in Bonaparte et al. (2014), an equivalent difference in the correlation makes the first individual's probability of participating only 5% higher compared to that of the second individual.

Among the controls, we include personal characteristics such as income and wealth levels, age, education, sex, risk aversion, family size, retirement or unemployment, and health status. Details

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

[7] The derivation of these marginal effects for the Probit model is in Appendix C.

about the control variables are in Table 13 of Appendix B. This table also replicates in our sample the known negative relationship between stock market participation and the sample correlation between income shocks and the stock market return.

We address potential endogeneity issues arising from including 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. Second, we construct dummy variables for quartiles of wealth and use these dummies as control variables in our regressions instead of the original continuous variable, following the approach of Van Rooij, Lusardi, and Alessie (2011). In both cases, we obtain results that are quantitatively equivalent to those reported in the regression tables.

We also repeat the empirical analysis using the individual's total income rather than labour income alone. The total income includes additional income components such as transfers from other members of the household. All the results still hold. These robustness checks are repeated in subsequent regression, without noteworthy changes in results.

As main independent variable we experiment with the *covariance* instead of the *correlation* between the PI shocks and the aggregate shocks. We compute this covariance using the clustered correlations between PI shocks and the aggregate shocks estimated with the MD methodology and either (i) the standard deviation of the PI shocks estimated with the MD methodology or (ii) the sample standard deviation of the total income shocks. Note that while (i) is assumed to be constant across individuals, (ii) is computed at the individual level. Nevertheless, using either (i) or (ii) yields very similar results: the marginal impact of the covariance is analogous to the one estimated for the correlation, both in terms of sign and statistical significance, but with a larger magnitude. This is not surprising since the size of the covariance is generally lower than the size of the correlation.

These results confirm the relevance of exposure to aggregate risk for stock market participation, when it is measured through co-movements of income shocks across groups of individuals. The co-movement based approach is in fact able to robustly explain the variation in non-participation associated with hedging motives.

4.2 Hedging Motives and Asset Allocation

This section exploits the DHS information on the shares of wealth invested in stocks to study the sensitivity of asset allocation to alternative estimates of exposure to aggregate risk.

Table 4 reports results from a Tobit specification, displayed with the same structure of Table 3. The dependent variable is the portfolio share invested in stocks, held by the individual either directly or through mutual funds. We see that the MD correlations between PI shocks and aggregate shocks (column 1) matters for the asset allocation decisions, as they always predicts the fraction of wealth invested in stocks with great precision (0.1% significance level). The coefficients associated with the other measures of exposure to aggregate risk are slightly smaller, with a lower level of statistical significance (5%) in columns (2),(4) and (6). When we add the MD correlation term to these alternative metrics in columns (3),(5) and (7), only one alternative correlation metric keeps its explanatory power.

Similar unreported results hold when the dependent variable alternatively measures just the equity share held directly by the individual or the one held through mutual funds. Unreported statistics for control variables show that wealthier, more educated and less risk-averse individuals generally invest larger fractions of their wealth in stocks, either directly or through mutual funds or both. As in Bonaparte et al. (2014), we find that high income-risk individuals prefer to directly allocate wealth to stocks, rather than through mutual funds. This analysis shows that the MD, co-movement based approach, is able to robustly explain variation in individuals' asset allocation, on top of their equity market participation choice.

[Table 4 about here.]

Vissing-Jorgensen (2002), Fagereng, Gottlieb, and Guiso (2017) and Bonaparte et al. (2014) consider simultaneously the market participation and the 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, adding lagged financial wealth and lagged squared financial wealth to the control variables as in Bonaparte et al. (2014), and report the results of Heckman regression estimates in columns (7)-(8) of Table 14. We find that the MD correlation estimates still predict asset allocation with the same precision (0.1% significance level), but the coefficient is much smaller than the one in the regression combining both participants and non-participant. This result is in line with Bonaparte et al. (2014) and is likely due to the limited size of the market participants sub-sample, which appears to be non-randomly drawn from the population based on the statistical significance of the lambda.

4.3 Robustness: Out-of-Sample Analysis and US Non-Participation

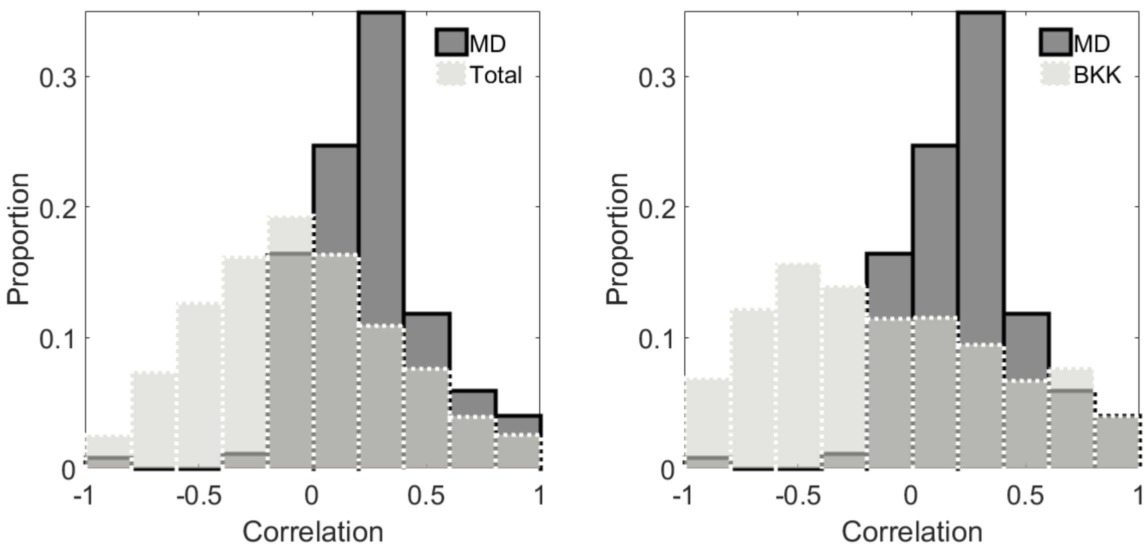
Based on the DHS sample, we have documented that hedging demands are negative for a large part of individuals who abstain from participating and that the mean Dutch individual hedges permanent income shocks by avoiding stock investments. Importantly, variations in the cluster-based correlation estimates obtained using our co-movement methodology have a large economic impact on portfolio choice. This evidence could be sample specific, though. For this reason we will now change the sample, in two ways. Section 4.3.1 will show that our results hold when we use a smaller number of waves rather than the waves, from 1993 until 2019. 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, that is what matters in portfolio management. In turn, Section 4.3.2 confirms the results on both the size of hedging needs and the variation of participation associated with individual's hedging on U.S. data. 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.

4.3.1 Predicting Portfolio Choice Out-Of-Sample

[Table 5 about here.]

Figure 6 and Table 5 display the distribution of alternative estimates of individual exposure to stock market returns, replicating the structure of Figure 3 and Table 2 respectively, based on the

Figure 6. Distribution of cluster-based correlations- short DHS sample
This figure displays the distribution of the cluster-based MD correlations between the permanent labor income shocks and the aggregate shocks. It compares it with the distribution of the sample correlations between stock market returns and total income shocks (Total) in the left-hand panel; and with the distribution of correlations between stock market returns and permanent income shocks estimated according to Bonaparte et al. (2014)(BKK) in the right-hand panel. The data are from the DHS, 1993–2011.



shorter DHS sample from 1993 to 2011. This is the sample used in Bonaparte et al. (2014). [8] Results are similar to the ones obtained using both the baseline sample and a sample truncated in 2007 (unreported). The mean cluster-based MD correlation estimates are higher (0.3) than both the negative ones obtained with prior methods and the non-cluster-based individual MD correlation estimates.

Table 6 reports the Probit regression results for stock market participation in the upper panel and the Tobit regression results for the share invested in stocks in the lower panel. This table confirms the ability of the MD correlation coefficients in predicting both dependent variables that

[8] Unreported descriptive statistics for this sample reveal individuals' characteristics similar to those in Table 1.

we found in previous sections using surveys up to and including 2019. Moreover, it shows the relative strength of the MD method relative to prior ones. The economic significance of hedging motives, when assessed through the coefficient of MD correlation estimates, becomes even larger than in the full sample.

[Table 6 about here.]

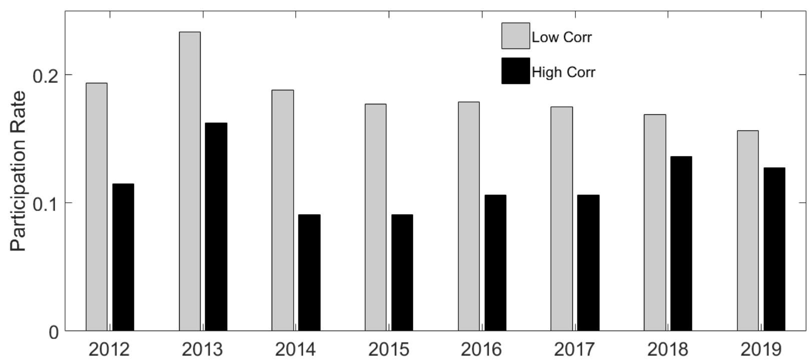
Thus, our results on both the size of co-movements with aggregate shocks and their impact on portfolio choice hold irrespective of the length of the sample. Similar results hold on sub-samples by education and retirement status and when focusing the asset allocation analysis on different dependent variables (Only Stocks or Only Mutual Funds).

We can therefore exploit the robustness of our estimates in short samples to implement an out-of-sample analysis. This relies on the MD correlation estimates, using data up to 2011, to predict portfolio choice up to and including 2019. This exercise reveals whether it is possible to use MD correlation estimates for portfolio management, that can only rely on past information.

We allocate each new survey participant past 2011 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 MD correlation parameter of the corresponding cluster estimated using data up to 2011. Then, we relate these correlation estimates to the decision to invest in equities in the years 2012 to 2019.

In a preliminary visual analysis of participation, we 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 7, display a systematic pattern: individuals belonging

Figure 7. 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 the sample correlation between income growth and stock market returns. The data are from the DHS, 1993–2019.



to clusters with lower estimated correlation parameters in 2011 participate in future years more than individuals allocated to clusters displaying higher correlation. The difference between the top and bottom quartiles is remarkable. This difference shrinks when we step away from the estimation time window. Nonetheless, 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 would be stronger were the clustered correlations estimated using data up to 2017.

The out-of-sample response of equity investments to correlation is examined through both a Probit analysis of participation and a Tobit analysis of the portfolio equity share in Table 7. In Table 7, equity investment is alternatively defined as the total one (in columns 1 and 4), held both directly or indirectly through mutual funds, or just the direct one (in columns 2 and 5) or just the indirect one (in columns 3 and 6). In the Probit analysis (columns $1-3$) the dependent variable is a dummy, indicating participation in the stock market between 2012 and 2019, while in the Tobit (columns $4-6$) it is the portfolio share held in stocks during the same years. Results demonstrate that the clustered correlations, estimated using data up to 2011, predict both stock market participation and the portfolio share allocated to equities for the period 2012–2019 with the negative sign implied by portfolio theory. The economic significance of our estimates is also remarkable, in line with the in-sample results. For instance, the coefficient in column (1) of Table 7 suggests 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 (ρ=-0.6) cluster (see equation (15)).

[Table 7 about here.]

In conclusion, MD correlation coefficients predict participation and asset allocation also out-of-sample, suggesting their use for improving on the current design that does not consider hedging demands. This exercise also confirms the relevance of hedging motives for portfolio choice in general and non-participation in particular also out-of-sample.

4.3.2 Accounting for Non-Participation in the US Stock Market

In household finance, the PSID database has been a workhorse data set for estimating earnings processes for U.S. individuals as it allows to measure equity market participation. [9] For this reason, this section examines whether our results carry over to the US economy using PSID data.

[9] Large administrative data are now available for earnings studies (see (Guvenen et al., 2022)).

Some adjustments in clustering are necessary, since there is no information about individuals' risk aversion and financial literacy and the sample is almost entirely composed by men. On the other hand, information about both the industry and the education level of the household head is available. Thus, we form 48 clusters based on 4 education groups and 12 industries, as in Campbell and Viceira (2002), to estimate the correlation parameters at the cluster-level with the MD methodology. We also estimate both β_{GUV} and β_{CV} at the cluster-level as in previous sections.

We report the mean correlation and beta coefficient estimates in Table 8. The MD one (0.268) is similar to those presented in Tables 2 and 5 based on the DHS over the full (0.257) and shorter (0.310) samples respectively. This confirms that also the mean US household's head has a negative hedging demand for stocks. The same insight now derives for the mean estimate of β_{GUV} (0.232). The other benchmark mean correlation estimates [10] are non-negative and therefore larger when compared to those estimated on the Dutch data. However, both the BKK mean correlation estimate (0.069) and the β_{CV} (0.040) are too small to deliver negative hedging demands that could account for observed stock market non-participation.

Given the relevance of industry-based correlations, that appear to be priced in Eiling (2013), Table **??** also reports the estimated correlations for different industries. For many industries, the MD estimates are in line with the ones presented in Campbell and Viceira (2002) and reported in the second column of the table.

[Table 8 about here.]

[10] When we use only three education clusters to compute β_{CV}, our MD estimates are in line with the OLS estimates in Campbell and Viceira (2002).

[Table 9 about here.]

Next, we check whether cross-sectional variation in both the MD correlation and β_{GUV} estimates based on PSID also explain variation in equity market non-participation. We hence repeat the Probit analysis of stock market participation in Table 6, reporting results in Table 9. Results are comparable, as the sample covers the same years and the main independent variables are the same.[11] While the statistical significance of the coefficients associated with the MD correlation parameters is maintained, the one of the β_{GUV} coefficient disappears.

In conclusion, our estimates of individuals' heterogeneous hedging motives explain cross-sectional variation in US stock market non-participation, showing robustness across both different clustering methods and sub-samples.

5 Learning about Income Hedging Needs

In cross-sectional analyses, unobserved characteristics may drive both the individual's earnings risk and portfolio choice. To rule out this possibility, this section studies whether revised correlations between the PI and the aggregate shocks explain changes in participation for each individual over time. Learning about earnings and return realizations may indeed affect hedging choices (as in Chang, Hong, and Karabarbounis, 2018).

To obtain revised correlation estimates, we first reconstruct the unobserved PI shocks at the individual-level using a Kalman filter. This exploits the assumed relationship between labor in-

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

comes and stock market returns, their realizations and the MD80 parameter estimates. [12] We then construct the sequence of updated correlation coefficients between the PI shocks and stock market returns over an expanding window. These revised correlation estimates will belong to the set of independent variables explaining the individual's probability to participate in the stock market in each period.

5.1 Updating Permanent Income and Correlation Estimates

In Section 2.1, equation (2) relates the observable total income shocks at time t, $e_{i,t}$, to the latent permanent component, $v_{i,t}$. In turn, equation (3) describes the dynamics of such latent component as a function of the unobserved PI shock at time t, $u_{i,t}$.

Using equations (4) and (5), we can relate the unobserved permanent income shocks to the realized stock market returns as follows:

$$u_{i,t} = \sigma_u \rho_i (r_t/\sigma_r) + \omega_{i,t}, \tag{7}$$

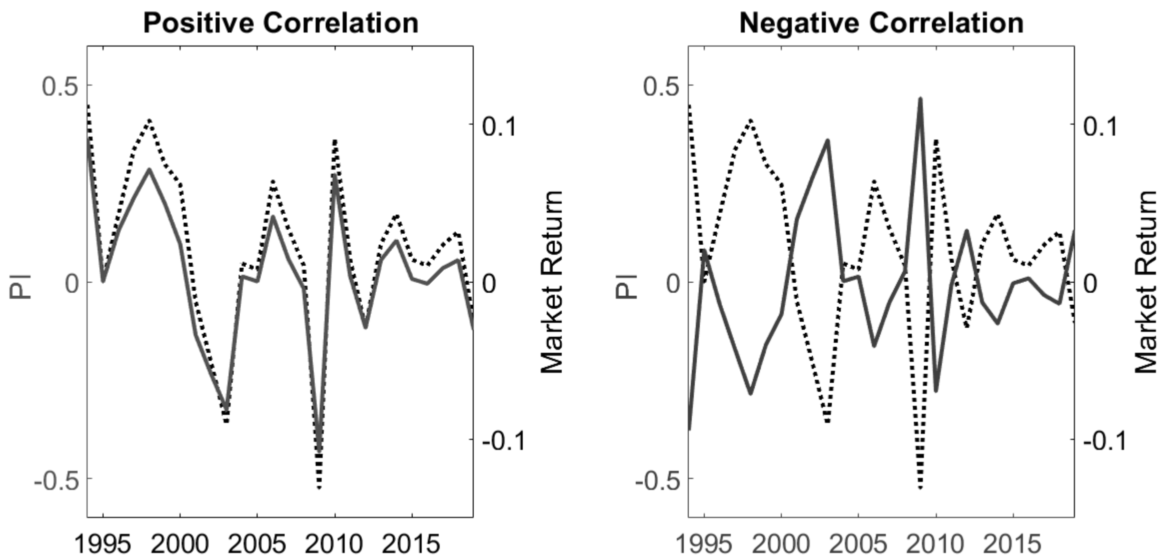
when the aggregate shock and the stock market return are perfectly correlated, $r_t = \sigma_r W_t$. [13]

We track the random walk $v_{i,t}$ using a linear Kalman Filter (KF). To implement this filter, we use estimates from Section 4 of the paper 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, using (7). We next form a prediction of the total income shock $e_{i,1}$, $\hat{e}_{i,1}$, by computing the expected value of $e_{i,1}$

[12] The Online Appendix inspects the properties of the resulting sequence of PI shocks across age, risk aversion and cohort.

[13] Section 6 relaxes this assumption.

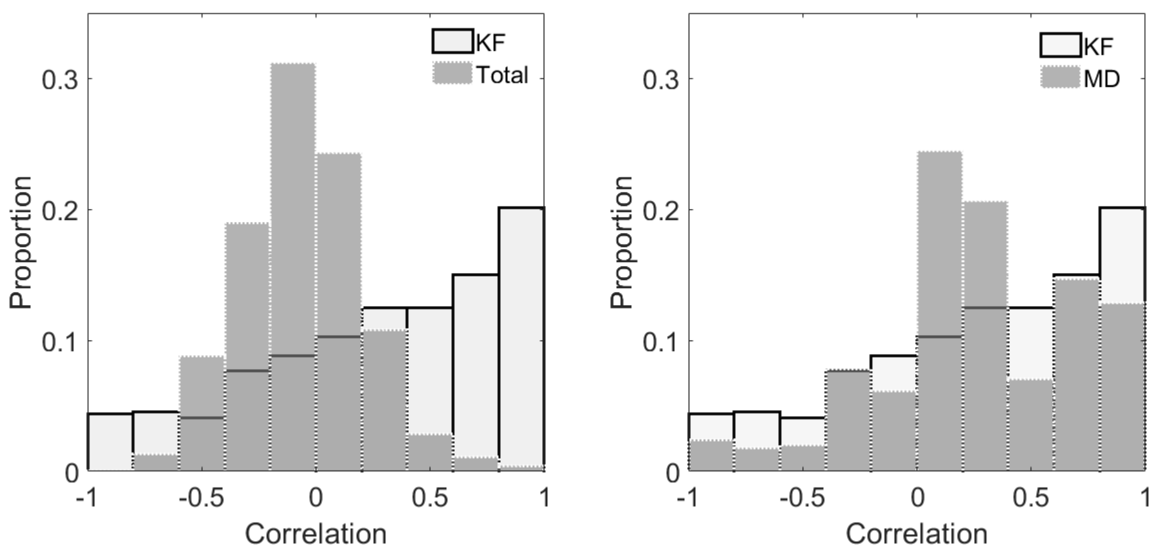
Figure 8. Individual permanent income shocks
The left panel compares the dynamics of stock market returns (dotted line) and permanent labor income (PI) shocks reconstructed using the Kalman filter (solid line) for the individual with the maximum sample correlation between stock market returns and PI 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.



conditional on $\hat{v}_{i,1}$. The difference between the actual and the predicted total income shocks is the measurement error that is used to compute the posterior for $v_{i,1}$, which turns to be the prior for the next point in time. We iterate the system up to T and we reconstruct the PI shocks by computing the first differences of the random walk. Consequently, for each individual, we obtain the dynamics of the permanent shocks to the labor income over the entire time series.

To represent our results, Figure 8 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.

Figure 9. 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). To compute the sample correlation based on the KF methodology, we use the correlation obtained with MD estimation at the cluster-level as ex-ante correlation parameter of the state-space model to reconstruct the PI shocks at the individual level, then we compute the sample correlation between these PI shocks and the stock market return (KF80). The data are from the DNB Household Survey and cover all waves for the period 1993–2019.



Meanwhile, Figure 9 plots the full distribution of the sample correlations between stock market returns and PI shocks reconstructed using the KF and compares it with both the distribution of the sample correlations between stock market returns and TS (left-panel) and with the distribution of individual correlations between stock market returns and PI shocks obtained by MD estimation (right-panel). In the left panel, the distribution of TS is widely dispersed over the entire set of the correlation values, as expected for sample realizations. However, the distribution obtained using the

KF is oriented slightly to the right, signaling higher frequency of positive values for the correlation between stock market returns and PI shocks compared to TS. In the second case, the distribution obtained using the KF is more dispersed than the distribution obtained using the MD estimates, which does not embed information on realized individual earnings and stock market returns.

5.2 Revising Participation and Hedging Needs

We now investigate whether individuals' decision to enter and exit the stock market is explained by their revised income hedging motives. To this end, we sequentially compute, on an expanding window, 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. This becomes our main independent variable, together with the MD correlation coefficient, in 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.

[Table 10 about here.]

In Table 10, we see that the lower is the revised correlation between stock market returns and PI shocks, the higher is 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 on the one hand exploit the information embedded in the variance-covariance matrices of total income shocks, both the inter-temporal one for each agent and the contemporaneous one across agents, 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 relying on the realization of both stock market returns and idiosyncratic earnings shocks over time. Thus, the latter information increase the relevance of MD correlation estimates.

In an unreported regression, 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 control variables. Results show that individuals remain in the market longer if their labor income shocks are negatively correlated with stock market returns, as indicated by the coefficients of both the MD and the KF correlation estimates. When we assess the economic significance of this effect, we find that increasing the revised correlation between PI shocks and stock market returns from -0.6 to 0.5 reduces participation by 21 months (38 months when considering the coefficients of both MD and KF correlation terms). Instead, participation is 14-months lower if we use the method in Bonaparte et al. (2014) to estimate correlations.

Results concerning participation revisions are free from unobserved heterogeneity concerns that may plague cross-sectional results. They support a strong role of hedging needs in explaining non-participation in the stock market, in line with portfolio theory.

6 Hedging Exposure to Stock Market Shocks

Previous sections document that the individuals' exposure to the latent aggregate shock shapes their hedging motives, which in turn explain observed non participation. This section shows that exposure to the stock market return, which is the focus of much prior literature, delivers the same insight. To this end, we estimate the correlations between the PI shocks and the stock market returns substituting them to the ones with the aggregate risk factor in the analysis.

Section 6.1 adds to both the model and the moment conditions while the subsequent ones replicate the empirical analysis with the new measure of correlation. We will see minor changes in results.

6.1 Model and Moment Conditions involving the Stock Market Return

Let the stock market return be a linear combination of the aggregate factor and an idiosyncratic noise:

$$r_t = \sigma_r W_{r,t}^p = \sigma_r(\rho_r W_t + \sqrt{1-\rho_r^2} W_{r,t}),$$

where $W_{r,t}$ is a Standard Normal random variable and ρ_r is the correlation between the stock return and the aggregate shock W_t . Therefore:

$$r_t \sim \mathcal{N}(0, \sigma_r^2 \rho_r^2 + \sigma_r^2(1-\rho_r^2)), \quad \text{i.e.} \quad r_t \sim \mathcal{N}(0, \sigma_r^2). \tag{8}$$

This specification nests the baseline case in Section 2.1 when ρ_r=1. Given equation (8), the vector θ of the unknown parameters now includes ρ_r and σ_r:

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

The set of implied moment conditions expands to include $N+1$ additional conditions. These involve both the covariance between the time variation in individual income shocks and the stock market returns, which is given by: [14]

$$cov(\Delta_d e_{i,t-1}, r_t) = \sum_{s=t}^{t+d} u_{i,s} + \epsilon_{i,t+d} - \epsilon_{i,t} = cov(u_{i,t}, r_t) =$$
$$cov(\sigma_u \rho_i W_t, \sigma_r \rho_r W_t) = \sigma_u \sigma_r \rho_i \rho_r.$$

and the variance of the stock market returns, which is equal to:

$$var(r_t) = \sigma_r^2.$$

As a result, the set of cross-sectional restrictions, is now of size equal to:

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

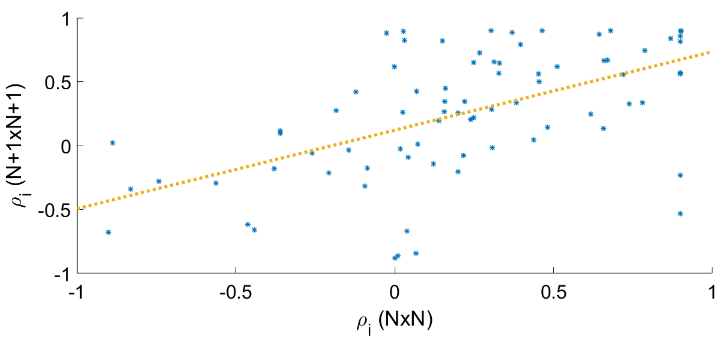
We then implement the MD estimation of θ as in Section 2.2.

6.2 Participation and Asset Allocation Choice.

We rely on the DHS data between 1993 and 2019 using the same clusters presented in Section 3.1. We report summary statistics of the cluster-based correlation parameter estimates between PI shocks and the stock market return in the upper panel of Table 11, while Figure 10 shows that these new estimates are in line with the baseline ones. This is perhaps unsurprising, since the additional moment conditions used to estimate correlations between individual PI shocks and stock market returns are but a small share of the conditions implied by the baseline specification.

[14] In the first line we use both the orthogonality of the transitory income shocks to the stock market returns and independence between the stock market returns at t and the permanent income shocks at any point time $s \neq t$. Similarly, we use in the second line the orthogonality of the idiosyncratic component of the permanent shocks to the stock market returns.

Figure 10. Cluster-based correlations: A comparison across model specifications
The figure compares the correlations between the permanent labor income shocks and the aggregate shocks estimated through the minimum distance (MD) methodology at the cluster-level in the baseline case (x-axis) with the correlations between the permanent labor income shocks and the aggregate shocks estimated through the minimum distance (MD) methodology at the cluster-level using the model specification of the stock market return presented in 6 (y-axis). The data are from the DNB Household Survey and cover waves for the period 1993–2019.



Next, we relate the correlation between income shocks and stock market returns to both equity market participation and asset allocation choices. Columns (1-4) and columns (7-10) in Table 11 show results akin to the ones presented in Tables 3 and 4, respectively. The large and robust economic impact on portfolio choice of hedging co-movements is confirmed. [15]

[Table 11 about here.]

6.3 Updating Correlations with the Stock Market Return

Let us derive the aggregate shock from equation (8):

$$W_t = \frac{r_t}{\sigma_r \rho_r} - \frac{\sqrt{1-\rho_r^2} W_{r,t}}{\sigma_r \rho_r},$$

[15]While the mean correlation coefficient is slightly lower in Table 11 than in Panel A of Table 2) ((0.189 versus 0.257), both the participation and the asset allocation choice are more sensitive to the correlation with the stock market returns (−0.442 in Panel B of Table 11 instead of −0.267 in Table 3; −0.129 in Panel B of Table 11 instead of −0.085 in Table 4).

and embed W_t in the equation (5) describing the dynamics of the PI shocks:

$$u_{i,t} = \sigma_u \left(\rho_i W_t\right) + \omega_{i,t} = \sigma_u \left(\rho_i \frac{r_t}{\sigma_r \rho_r} - \frac{\sqrt{1-\rho_r^2} W_{r,t}}{\sigma_r \rho_r}\right) + \omega_{i,t},$$

where $\omega_{i,t} = \sigma_u \left(\sqrt{1-\rho_i^2} W_{i,t}^p\right) \sim \mathcal{N}(0, \sigma_u^2(1-\rho_i^2))$. Then, we can predict the permanent component of labour income, $v_{i,t}$, based on its past value and the contemporaneous stock market return:

$$E[v_{i,t}|(v_{i,t-1}, r_t)] = v_{i,t-1} + E[u_{i,t}|(v_{i,t-1}, r_t)] = v_{i,t-1} + \sigma_u \rho_i \frac{r_t}{\sigma_r \rho_r}$$

As in Section 5, we can build the dynamics of the permanent shocks to the labor income over the entire time series for each individual. Then we sequentially compute the sample correlation between stock market returns and PI shocks up to a given wave $t < T$, where T is the total number of available waves in our sample. Finally, we use these sample correlations in a Probit analysis of stock market participation over time and we report results in Table **??**, Panel B, columns 5-6. Results are similar to the ones obtained when we use estimates of KF-based correlations between PI shocks and the aggregate risk factor. That is, revised correlations retain explanatory power for participation revisions provided that MD correlations with stock returns are omitted. Finally, we replicate the out-of-sample probit analysis of Table 7 in Table 12.

[Table 12 about here.]

7 Summary and Conclusions

This paper supports hedging motives as an explanation for the large, and so far puzzling, share of individuals who do not invest in the stock market. It shows that the mean individual has positive exposure to aggregate shocks and hence a negative hedging demand for stocks. Furthermore, the increase in such exposure is robustly associated to a decrease in stock market participation both in the cross section and in the time series.

We succeed in linking non-participation to hedging motives by modeling the co-movements of earnings growth across groups of individuals, co-movements that arise due to their heterogeneous exposure to aggregate shocks. We then exploit these co-movements to assess the individuals' hedging motives. Our estimates of the mean individual exposure to aggregate shocks lie above 0.2. The mean correlation estimate of permanent income shocks with the stock market return is also positive and close to 0.2, as opposed to previous estimates that are close to zero. This result appears in both Dutch and US data, that cover non-homogeneous groups of individuals. It is insensitive to the length of the sample and does not depend on restrictions implied by the model of income shocks. The mean individual therefore hedges labour income risk by reducing exposure to the equity market with respect to an individual, like the one implied in previous estimates of correlations, with no exposure to aggregate risk. This implies that observed portfolios are closer than previously thought to the implications of portfolio choice theory, without resorting to alternative mechanisms. [**Almost** 80% **of non-participating individuals has income shocks that positively co-move with systematic shocks, and therefore a negative hedging demand for stocks, based on our approach. This compares with shares of** 50%**,** 33% **and** 21% **based on previous estimates.**]

Results concerning equity investments further support our robust measurement that exploits both the time series dimension of the data, as in prior research, and the cross-sectional one. Not only the sign but also the size and the precision of the estimated effects align with the prediction that equity investments fall when permanent income is more 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 heterogeneous exposures to aggregate risk is a central issue in incomplete markets, these advances in their measurement may prove useful beyond the boundaries of household finance.

References

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

Arrondel, L., H.C. Pardo, and X. 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, S., T. Jansson, C. Parlour, and J. 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, L.E., and P. Sodini, 2014, Twin picks: Disentangling the determinants of risk-taking in household portfolios, *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*.

Carroll, C.D., 1997, Buffer-stock saving and the life cycle/permanent income hypothesis, *Quarterly Journal of Economics* 112, 1–55.

Carroll, C.D., R.E. Hall, and S.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, 2005, Consumption and portfolio choice over the life cycle, *Review of Financial Studies* 18, 491–533.

Eiling, E., 2013, Industry-specific human capital, idiosyncratic risk and the cross-section of expected stock returns, *Journal of Finance* 68(1), 43–84.

Fagereng, A., C. Gottlieb, and L. Guiso, 2017, Asset market participation and portfolio choice over the life cycle, *Journal of Finance* 72(2), 705–750.

Fagereng, A., L. Guiso, and L. Pistaferri, 2018, Portfolio choices, firm shocks, and uninsurable wage risk, *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., M. Haliassos, and T. Ramadorai, 2021, Household finance, *Journal of Economic Literature* 59, 919–1000.

Guvenen, F., 2009, An empirical investigation of labor income processes, *Review of Economic Dynamics* 12, 58–79.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song, 2021, What do data on millions of U.S. workers reveal about life-cycle earnings dynamics?, *Econometrica* 89, 2303–2339.

Guvenen, F., L. Pistaferri, and G.L. Violante, 2022, Global trends in income inequality and income dynamics: New insights from GRID, *Quantitative Economics* 13, 1321–1360.

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, F., and A.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, J.J., 1979, Sample selection bias as a specification error, *Econometrica* 153–161.

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

Meghir, C., and L. 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.

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

Shen, J., 2023, Countercyclical risks, consumption and portfolio choice: Theory and evidence, *Management Science*.

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 portfolio choice heterogeneity: Nonfinancial income and participation cost structures, *National Bureau of Economic Research*.

Appendix A Model and Empirical Conditions

In this Appendix, we detail the two sets of moment conditions obtained by using the model restrictions described in Section 2. First, we present the inter-temporal conditions, then we characterize the restrictions that hinge upon the co-movements between individuals' income shocks. Next, we outline the sample counterparts of the two sets of model-implied conditions.

A.1 Inter-Temporal 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:

$$\Delta_d e_{i,t} = e_{i,t+d} - e_{i,t} \tag{9}$$

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, since the random walk component $v_{i,t}$ goes away:

$$e_{i,t+d} - e_{i,t} = v_{i,t} + \sum_{s=t+1}^{t+d} u_{i,s} + \epsilon_{i,t+d} - (v_{i,t} + \epsilon_{i,t}) = \sum_{s=t+1}^{t+d} u_{i,s} + \epsilon_{i,t+d} - \epsilon_{i,t}, \tag{10}$$

.

For instance, when $d = 1$, $\Delta_1 e_{i,t}$ is the familiar first difference of TS and is equal to:

$$\Delta_1 e_{i,t} = u_{i,t+1} + \epsilon_{i,t+1} - \epsilon_{i,t} \tag{11}$$

In addition, we can also compute the DTS at different lags l:

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

that is equal to, by using (10):

$$\Delta_d e_{i,t+l} = \sum_{s=t+l+1}^{t+l+d} u_{i,s} + \epsilon_{i,t+l+d} - \epsilon_{i,t+l}.$$

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

$$\Delta_d e_{i,t+l} = e_{i,t+4} - e_{i,t+1} = \sum_{s=t+2}^{t+4} u_{i,s} + \epsilon_{i,t+4} - \epsilon_{i,t+1}.$$

Thus, the covariance between the DTS of the N individuals at two different lags l, for a given time interval of length d, is given by:

$$cov\left(\Delta_d e_t, \Delta_d e_{t+l}\right) = cov\left(\sum_{s=t+1}^{t+d} u_s + \epsilon_{t+d} - \epsilon_t, \sum_{s=t+l+1}^{t+l+d} u_{i,s} + \epsilon_{t+l+d} - \epsilon_{t+l}\right), \tag{12}$$

where $\Delta_d e_t$ and $\Delta_d e_{t+l}$ are two N-dimensional vectors containing the DTS of N individuals at two different points in time:

$$\Delta_d e_{t+s} = \{\Delta_d e_{1,t+s}, \Delta_d e_{2,t+s}, ..., \Delta_d e_{N,t+s}\}.$$

When $l = 0$, the inter-temporal covariance in (12) is simply equal to the cross-sectional variance of the DTS at each time t, for a given time interval of length d:

$$cov\left(\Delta_d e_t, \Delta_d e_{t+l}\right) = var\left(\Delta_d e_t\right) =$$
$$= var\left(\sum_{s=t+1}^{t+d} u_s + \epsilon_{t+l+d} - \epsilon_{t+l}\right) = d \;\; var(u_t) + 2\sigma_\epsilon^2,$$

where $var(u_t)$ is the cross-sectional variance of the PI shocks across the N individuals, at each point in time t. [16]

[16]This variance depends on both the cross-sectional variance of the individual correlation coefficients, ρ_i, and the unobserved realization of the aggregate shock at time t, W_t. This complication disappears when grouping individuals

When $l > 0$, the inter-temporal covariance depends on the relationship between the lag l, over which we are computing this covariance, and the length d of the time interval, over which we are computing the DTS. Specifically, if $d > l$:

$$cov\left(\Delta_d e_t, \Delta_d e_{t+l}\right) = (d-l)\ \ var(u_t),$$

since there are $(d-l)$ contemporaneous terms u_{t+s} appearing in both vectors $\Delta_d e_t$ and $\Delta_d e_{t+l}$, zero contemporaneous terms ϵ_{t+s}, and we exploit the independence over time of both permanent and transitory shocks. Instead, if $d = l$, only one contemporaneous term ϵ_{t+s} appears in both vectors $\Delta_d e_t$ and $\Delta_d e_{t+l}$, with opposite sign, so that:

$$cov\left(\Delta_d e_t, \Delta_d e_{t+l}\right) = -\sigma_\epsilon^2.$$

Finally, if $d < l$, there is no contemporaneous term appearing in both vectors $\Delta_d e_t$ and $\Delta_d e_{t+l}$, thus the inter-temporal covariance is equal to zero.

Using the inter-temporal model restrictions, we obtain for each length d a number of moment conditions equal to:

$$M_d^{int} = \frac{(T-d)(T-d-1)}{2} + (T-d) = \frac{(T-d)(T-d+1)}{2}. \tag{13}$$

A.2 Moment Conditions over N individuals

Next, we derive the model restrictions implied by the co-movements between individuals' income shocks. The key idea is that the co-movement between the income shocks of two individuals is due to their exposure to the common, aggregate risk factor. This exposure is heterogeneous across

into clusters (Section 3). Since individuals belonging to the same cluster have equal exposure to the aggregate shocks, there is zero cross-sectional variance of the correlation coefficients, ρ_i, within cluster. Thus $var(u_t)$ no longer depends on W_t and simplifies to σ_u^2. The complete derivation of $var(u_t)$ is available upon request.

the N individuals and described by the individual-specific correlation coefficient ρ_i. In fact, the covariance between two individuals' DTS is only given by the covariance between the the systematic components of their PIs:

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

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

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

By modeling such co-movements, we can exploit 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 $(N(N-1))/2$.

Indeed, when $i = j$, this covariance simply reduces to the variance of the i-th individual's DTS:

$$cov(\Delta_d e_{i,t}, \Delta_d e_{i,t}) = var(u_{i,t+1}) = \sigma_u^2 + 2\sigma_\epsilon^2$$

Thus, using the cross-sectional restrictions, we obtain a number of moments conditions equal to:

$$M^{crs} = \frac{N(N-1)}{2} + N = \frac{N(N+1)}{2} \tag{14}$$

where the first term is the number of conditions obtained from the covariance terms between the individuals' DTS and the second term is the number of conditions obtained from the variance of the N individuals' DTS.

A.3 Empirical Counterparts of Moment Conditions

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 are the first differences of the regression residuals for each individual and are denoted as *dres*. Then, we compute the covariance between each pair of individuals' *dres* and the variance of each individual's *dres* as sample counterparts of the moment restrictions detailed in section A.2. To obtain the sample counterparts of the inter-temporal restrictions, we use all of the $\Delta_d e_{i,t}$ up to $d = D$, where D is an arbitrary number denoting the maximum length of the time interval considered in the analysis. Then, for each d, the inter-temporal covariances of the N individuals' *dres* are the sample counterparts of the covariances between the DTS of the N individuals at different lags l. Moreover, the cross-sectional variance of the *dres* is the empirical counterpart of the inter-temporal covariance when l=0 (i.e., the cross-sectional variance of the DTS at each t).

Appendix B Probit and Tobit Analysis: Preliminary Evidence

We test in our sample results from prior literature about the income hedging motive of individual's stock market investments. We report the results from a Probit analysis for stock market participation in Table 13 and from a Tobit analysis for asset allocation in Table 14, in which we also detail the control variables included in the estimated regression described in the main body of the paper.

In Table 13, columns (1) and (4) display 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 only the latter component matters. Columns (3) and (6) show, importantly, the larger marginal effect of the correlation between PI shocks and aggregate shocks estimated using minimum-distance. Columns (4)-(6) further 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, other characteristics, such as sex, income level and employment status, do not play a significant role in column (6).

[Table 13 about here.]

In Table 13, we support a negative relationship between the individual's risky share and the income-aggregate shocks correlation, also when conditioning on a set of observable characteristics that may impact an individual's portfolio allocation. In addition, we include here findings from an Heckman model estimation. In fact, Vissing-Jorgensen (2002), Fagereng et al. (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 likely 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.[17]

[Table 14 about here.]

Appendix C Marginal effects in Probit model

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:

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

[17] 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.

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). Since the marginal effect also depends on the original probability, consider without loss of generality 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. Let us 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 estimated coefficient of MD80 correlations (column (1) in Table 3) the first individual's propensity to participate is 12% higher than the second individual's propensity to participate $((-0.6 - 0.5) * (-0.267) * \phi(\hat{\beta}' X))$. Using instead the results of column (2) in Table 3, we conclude that an equivalent difference in the correlation 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).

Table 1 List of Variables and Summary Statistics.
This table reports the definition (top panel) and the summary statistics (bottom panel) for the variables used for the empirical analysis. The data are from the DHS, 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.

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)-Net worth.
Ln(NetIncome)	(log)-Net income, where Net income is the annual job earnings net of taxes.
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.

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 2 Correlation Parameters from DHS. Summary Statistics on Full Sample. This table reports the sample correlations between the stock market return and total income shocks; the correlations between the aggregate shock and the permanent component of labor income shocks estimated with the minimum distance methodology (MD) at both the individual and the cluster levels; the estimated correlations at each t between stock market returns and the permanent component of labor income shocks obtained using a Kalman Filter (Revised KF); the correlations 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. The t-test appears in the last column and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DHS, 1993–2019.

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Total	-0.062	0.269	-0.402	-0.065	0.260	-4.905***
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 (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))	-0.082	0.183	-0.334	-0.063	0.073	-19.567***
Beta (C&V(2002))	-0.069	0.237	-0.256	-0.072	0.114	-12.674***

Table 3 Stock Market Participation
This table reports the Probit regression results for the 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 (PermMD), and different known measures of the relationship between labour income growth rate and stock market returns: 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 13. We report in parentheses the Probit-robust standard errors and ***,**,* respectively denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DHS, 1993–2019.

	OwnSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
St. Dev.	-0.153***	-0.480***	-0.504***	-0.141***	-0.153***	-0.139***	-0.152***
	(0.038)	(0.117)	(0.118)	(0.037)	(0.038)	(0.038)	(0.038)
Corr(PermMD)	-0.267***		-0.185***		-0.276***		-0.261***
	(0.033)		(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	YES
Year Dummy	YES	YES	YES	YES	YES	YES	YES
N	12,758	3,286	3,286	12,758	12,758	12,758	12,758
Pseudo Rˆ2	0.277	0.320	0.322	0.272	0.276	0.272	0.277

Table 4 Asset Allocation
This table reports the Tobit regression results for the asset allocation decision. The dependent variable is the portfolio share held 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 (PermMD), and different known measures of the relationship between labour income growth rate and stock market returns: 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 13. We report in parentheses the Tobit-robust standard errors and ***,**,* respectively denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DHS, 1993–2019.

	PropSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
St. Dev.	-0.033*	-0.173***	-0.185***	-0.029*	-0.033*	-0.028*	-0.033*
	(0.014)	(0.037)	(0.038)	(0.014)	(0.014)	(0.014)	(0.014)
Corr(PermMD)	-0.085**		-0.064**		-0.086***		-0.082***
	(0.013)		(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	YES
Year Dummy	YES	YES	YES	YES	YES	YES	YES
N	12,498	3,224	3,224	12,498	12,498	12,498	12,498
Pseudo R^2	0.292	0.366	0.368	0.289	0.292	0.289	0.292

Table 5 Correlation Parameters from DHS. Summary Statistics on Short Sample. This table reports reports the sample correlations between the stock market return and total income shocks; the correlations between the aggregate shock and the permanent component of labor income shocks estimated with the minimum distance methodology (MD) both at the individual and the cluster levels. It also reports the estimated correlations at each t between stock market returns and the permanent component of labor income shocks obtained using a Kalman Filter (Revised KF). We also report the correlations 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. The t-test appears in the last column and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DHS, 1993–2011.

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Total	-0.068	0.412	-0.603	-0.079	0.501	-6.165***
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 (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))	-0.083	0.196	-0.327	-0.066	0.076	-16.121***
Beta (C&V(2002))	-0.073	0.256	-0.304	-0.065	0.106	-10.790***

Table 6 Probit and Tobit Estimates for Stock Market Participation. Short Sample
The table replicates Tables 3 and 4 on a shorter sample. Data are from the DHS, 1993–2011.

	OwnSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
St. Dev.	-0.172**	-0.173**	-0.191**	-0.139**	-0.149**	-0.140**	-0.150**
	(0.067)	(0.067)	(0.051)	(0.067)	(0.047)	(0.048)	(0.039)
Corr(PermMD)	-0.289***		-0.289***		-0.284***		-0.289***
	(0.058)		(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	YES
Year Dummy	YES	YES	YES	YES	YES	YES	YES
N	5,968	5,677	5,677	7,745	7,745	7,745	7,745
Pseudo R^2	0.264	0.262	0.265	0.252	0.259	0.256	0.259
	PropSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
St. Dev.	-0.018	-0.016	-0.023	-0.014	-0.018	-0.014	-0.018
	(0.018)	(0.023)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Corr(PermMD)	-0.096***		-0.090***		-0.096***		–0.096***
	(0.022)		(0.024)		(0.022)		(0.022)
Corr(tranBKK,Rm)		-0.044	-0.048*				
		(0.019)	(0.052)				
Corr(permBKK,Rm)		-0.059***	-0.056***				
		(0.015)	(0.042)				
Beta (Guvenen et al. (2017))				0.022	0.018		
				(0.013)	(0.042)		
Beta (C&V(2002))						-0.018	-0.018
						(0.038)	(0.038)
Controls	YES	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES	YES
N	7,581	5,577	5,577	7,581	7,581	7,581	7,581
Pseudo R^2	0.269	0.287	0.290	0.266	0.269	0.269	0.269

Table 7 Out-Of-Sample Participation and Asset Allocation

The table reports the out-of-sample Probit and Tobit regression results for stock market participation decision and asset allocation, respectively, for the waves 2012-2019. In columns (1) to (3), 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. In columns (4) to (6), the dependent variable is the portfolio shares in stocks either directly or through mutual funds (PropSTKMF), in stocks only (PropSTK) or mutual funds only (PropMF), 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 (PermMD) 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 controls as in Table 13. N is the number of observations. We report in parentheses the Probit-robust standard errors and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DHS, 1993–2019.

	OwnSTKMF	**OwnSTK**	**OwnMF**	**PropSTKMF**	**PropSTK**	**PropMF**
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.239***	-0.168**	-0.207**	-0.060**	-0.034	-0.081**
	(0.078)	(0.090)	(0.075)	(0.026)	(0.031)	(0.032)
Corr(PermMD)	-0.207***	-0.140*	-0.136*	-0.081***	-0.069**	-0.060*
	(0.073)	(0.087)	(0.075)	(0.027)	(0.032)	(0.032)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	4,336	4,336	4,336	4,258	4,258	4,258
Pseudo Rˆ2	0.324	0.273	0.264	0.354	0.317	0.280

Table 8 Correlation and Beta Parameters. Summary Statistics (PSID)
Panel A reports the sample correlations between the stock market return and total income shocks; the MD correlations estimated both at the individual and the cluster level; the revised correlations between stock market return and the permanent component of labor income shocks obtained through a Kalman Filter (Revised KF); the correlations 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. Panel B reports the estimated MD correlations and the correlation coefficients between the labour income growth rate and the stock market returns reported in Campbell and Viceira (2002), for different industries. The MD correlation parameters at the industry level are equal to the average of the estimated correlation parameters across the clusters that include a given industry. The clustering variables are Education and Industry (two-digit code). The t-test appears in the last column and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the Panel Survey Income Dynamics (PSID). 1988–2011.

Panel A

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Total	0.049	0.034	-0.697	0.065	0.733	2.915***
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 (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***

Panel B

Sector	Minimum Distance (MD)	Campbell & Viceira (CV, 2002)
Agriculture	0.38	0.22
Construction	0.25	0.55
Manufacturing	0.33	0.35
Transports & Communications	0.35	0.19
Retail Trade	0.30	0.03
Finance	0.37	0.13
Business Services	0.35	0.19
Professional Services	0.26	0.26
Public Administration	0.01	0.42

Table 9 Stock Market Participation (PSID)

The table replicates the Probit regression analysis for the stock market participation in Table 3 using U.S. data and different clustering characteristics. Additional control variables are individual demographic characteristics, such as the (log)-labour income, marital status, family size, age and years of schooling, and year-fixed effects. We report in parentheses the Probit-robust standard errors and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. We use data from the Panel Survey Income Dynamics (PSID), 1988–2011.

	OwnSTKMF						
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
St. Dev.	0.015	0.013	0.015	0.001	0.006	0.005	0.009
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Corr(PermMD)	-0.401***		-0.365***		-0.280***		-0.390***
	(0.077)		(0.077)		(0.088)		(0.093)
Corr(tranBKK,Rm)		0.140*	0.123*				
		(0.057)	(0.057)				
Corr(permBKK,Rm)		-0.084*	-0.075				
		(0.042)	(0.043)				
Beta (Guvenen et al. (2017))				0.469	0.292		
				(0.097)	(0.112)		
Beta (C&V(2002))						0.337	0.013
						(0.113)	(0.138)
Controls	YES	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES	YES
N	6,804	6.804	6,804	6,804	6,804	6,804	6,804
Pseudo R^2	0.091	0.094	0.094	0.094	0.094	0.094	0.095

Table 10 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 and/or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation and different measures of the relationship between labour income growth rate and stock market returns used in previous papers as described in Table 3. The correlation and the beta coefficients used here are computed on an expanding time window, from the first up to the t-th wave, where t goes from 10 to the 19 (the last wave). We use the clustered MD correlation as ex-ante correlation parameter when reconstructing the PI shocks using the Kalman filter at the individual level. Then, we compute the yearly change in the correlation between these PI shocks and the stock market returns (Revised KF). We include the same control variables as in Table 13. We report in parentheses the Probit-robust standard errors and ***,**,* denote statistical significance at the 0.1%, 1%, and 5% levels. The data are from the DHS, 1993–2019.

	OwnSTKMF				
Independent Variable	(1)	(2)	(3)	(4)	(5)
St. Dev.	-0.207***	-0.202**	-0.203***	-0.187***	-0.187***
	(0.046)	(0.043)	(0.057)	(0.060)	(0.042)
Corr(Tot,Rm)	-0.039				
	(0.032)				
Corr(PermMD)		-0.288***			
		(0.037)			
Corr(Revised KF,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
Year Dummy	YES	YES	YES	YES	YES
N	9,345	10,423	7,586	10,423	10,423
Pseudo Rˆ2	0.277	0.285	0.280	0.272	0.272

Table 11 Participation and Correlations with the Stock Market Return.
This table reports in Panel A the summary statistics of the estimated correlations between the permanent component of labor income shocks and the stock market returns estimated by using the MD methodology at the cluster level (Corr(PermMD,Rm)) when allowing for imperfect correlation between the aggregate shocks and the stock market returns, and the estimated correlation between the permanent component of labor income shocks and the stock market returns when updating the correlation at each t (Corr(Revised KF,Rm)) obtained using the Kalman Filter. 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. In Panel B, we replicate the Probit regressions in Table 3 (columns 1-4) and Table 10 (columns 5-6), and the Tobit regressions in Table 4 (columns 7-10), where the correlation between the permanent component of labor income shocks and the stock market returns (Corr(permMD80,Rm)) substitutes the one with the aggregate shock (PermMD80). The data are from the DNB Household Survey and cover waves from 1993–2019.

Panel A

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
Corr(PermMD,Rm)	0.189	0.234	-0.121	0.195	0.472	35.211***
Corr(Revised KF,Rm)	0.379	0.566	-0.543	0.574	0.956	23.776***

Panel B

	OwnSTKMF						PropSTKMF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
St. Dev.	-0.144***	-0.499***	-0.144***	-0.144***	-0.182***	-0.030*	-0.182*	-0.191***	-0.030*	-0.030*
	(0.038)	(0.118)	(0.038)	(0.038)	(0.042)	(0.043)	(0.014)	(0.118)	(0.014)	(0.014)
Corr(PermMD,Rm)	-0.442***	-0.359***	-0.442***	-0.428***		-0.443***	-0.129***	-0.119***	-0.124***	-0.123***
	(0.063)	(0.134)	(0.068)	(0.064)		(0.072)	(0.024)	(0.042)	(0.027)	(0.025)
Corr(Revised KF,Rm)					-0.272**	-0.142				
					(0.117)	(0.119)				
Corr(tranBKK,Rm)		0.086						0.077*		
		(0.114)						(0.0.036)		
Corr(permBKK,Rm)		-0.172						-0.084*		
		(0.105)						(0.034)		
Beta (Guvenen et al. (2017))			0.002						0.017	
			(0.092)						(0.036)	
Beta (C&V(2002))				-0.155**						-0.065*
				(0.087)						(0.034)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	12,758	3,286	12,758	12,758	10,423		10,423	12,498	3,224	12,498
12,498										
Pseudo R^2	0.275	0.322	0.276	0.276	0.279	0.285	0.291	0.368	0.291	0.291

Table 12 Out-Of-Sample Participation and Asset Allocation
The table is a replica of the Out-of-Sample exercise in Table 7, where the correlation between the permanent component of labor income shocks and the stock market returns (Corr(permMD,Rm)) substitutes the one with the aggregate shock (PermMD).

	OwnSTKMF	**OwnSTK**	**OwnMF**	**PropSTKMF**	**PropSTK**	**PropMF**
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.244***	-0.162*	-0.217**	-0.057*	-0.029	-0.082*
	(0.072)	(0.090)	(0.075)	(0.026)	(0.031)	(0.032)
Corr(PermMD,Rm)	-0.302***	-0.009	-0.325***	-0.086***	-0.008	-0.116***
	(0.064)	(0.082)	(0.066)	(0.024)	(0.030)	(0.029)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	4,336	4,336	4,336	4,258	4,258	4,258
Pseudo Rˆ2	0.327	0.273	0.269	0.358	0.318	0.287

Table 13 Probit Estimates for Stock Market Participation. Preliminary Analysis
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 (permMD), 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(PermMD)			-0.313***			-0.267**
			(0.058)			(0.033)
(log)-Income				-0.095*	-0.097*	-0.005
				(0.042)	(0.043)	(0.027)
(log)-Wealth				0.301***	0.301***	0.289***
				(0.020)	(0.020)	(0.013)
HH Size				-0.122***	-0.117***	-0.106***
				(0.019)	(0.020)	(0.013)
Age				0.008	0.007	0.019**
				(0.011)	(0.011)	(0.007)
Education				0.236***	0.249***	0.181***
				(0.043)	(0.043)	(0.029)
Sex				0.196**	0.190***	-0.022
				(0.057)	(0.057)	(0.037)
Unemployed				-0.001	0.008	-0.013
				(0.104)	(0.105)	(0.075)
Retired				0.029	0.036	0.054
				(0.071)	(0.072)	(0.045)
Health				0.004	0.001	0.013
				(0.030)	(0.030)	(0.020)
Risk Aversion				-0.355***	-0.354***	-0.334***
				(0.010)	(0.010)	(0.007)
Year Dummy	YES	YES	YES	YES	YES	YES
N	9,217	9,186	27,445	5,719	5,709	12,758
Pseudo R^2	0.016	0.016	0.026	0.289	0.290	0.277

Table 14 Tobit Estimates for Asset Allocation. Preliminary Analysis
The table reports the Tobit regression results for asset allocation decision. The dependent variable is the share of wealth held in stocks either directly or through mutual funds (PropSTKMF). The main independent variables are the standard deviation of labor income shocks, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD), 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 ***,**,* denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DHS, 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(PermMD)			-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)