

Hedging Permanent Income Shocks

Fabio Bagliano, Raffaele Corvino, Carolina Fugazza,
Giovanna Nicodano

NETSPAR ACADEMIC SERIES

Hedging Permanent Income Shocks.*

Fabio C. Bagliano¹, Raffaele Corvino¹, Carolina Fugazza¹ and Giovanna Nicodano³

¹University of Torino and CeRP (Collegio Carlo Alberto)

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

May 28th, 2023

Abstract

The exposure of earnings to an aggregate shock implies co-movements across individuals' income shocks. We model such co-movements to estimate the individual correlations between permanent earnings shocks and a latent aggregate shock. We similarly estimate their correlations with stock market returns. These correlation estimates retain statistical significance in predicting both portfolio choice and participation when other measures do not. They explain participation both out-of-sample and for the same individual over time. Results, based on both Dutch and US data, support the theoretical prediction that individuals rely on asset markets to help hedge their earnings' shocks.

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

JEL Classification: G10,G11, D14, C15

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

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

*We are grateful to Alex Michaelides, Laurent Bach, Marie Brière, Laurent Calvet, Bernard Dumas, Davide Fiaschi, Christian Gollier, George Korniotis, 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 Collegio Carlo Alberto, the European Central Bank, Imperial College London and the University of Pisa for helpful suggestions. We are also grateful to the Institute for Quantitative Investment Research (INQUIRE) Europe, the Observatoire de l'Épargne Européenne, and the Italian MIUR ("Department of Excellence" 2018–2022) for funding.

1 Introduction

According to life-cycle theory, portfolio investments help hedge individual permanent labor income (PI) shocks which are otherwise uninsured. The theory encourages investors to reduce (Merton, 1969; Viceira, 2001) and possibly avoid (Benzoni, Collin-Dufresne, and Goldstein, 2007; Bagliano, Fugazza, and Nicodano, 2014) equity holdings when stock market returns display positive correlation with their PI shocks, implying that equities amplify earnings risk. However, the evidence is less clear-cut (see e.g. Heaton and Lucas (2000), Campbell, Cocco, Gomes, and Maenhout (2001), Campbell and Viceira (2002), Cocco, Gomes, and Maenhout (2005), Angerer and Lam (2009), Bonaparte, Korniotis, and Kumar (2014), Catherine, Sodini, and Zhang (2020)). Both stock market participation and portfolio choice are moderately sensitive to the estimated individual correlations between realized labor income growth and realized stock returns. Moreover, the statistical significance of such correlations suffers from the short length of the data. Finally, the estimated average correlation coefficients are often close to zero, so that they are unlikely to account for moderate risk taking in equity markets.

This paper models the joint distribution of labour income shocks and stock returns allowing for a latent aggregate shock driving both. It then uses a novel approach to estimate the individual correlation between PI shocks and such aggregate shock. These new estimates (averaging 0.2-0.3 or above in both Dutch and U.S. data) imply larger hedging motives than previously thought. They retain statistical significance in predicting both portfolio choice and participation when other measures of correlation between PI shocks and stock market returns do not. They are also able to predict participation out-of-sample. Since unobserved heterogeneity may plague cross-sectional

results, we show that individual participation over time is sensitive to updates of correlation estimates. These results hold when we similarly measure the correlation between PI shocks and stock market returns. For the first time, the data consistently align with the theoretical prediction that portfolio holdings of equities respond to such correlations.

Our approach posits that groups of individuals display heterogeneous exposure to aggregate risk, as discussed in Campbell and Viceira (2002) and Guvenen, Schulhofer-Wohl, Song, and Yogo (2017). Accordingly, we allow for correlation between an aggregate shock and individual PI shocks in a labor income process that is otherwise commonly used (see Carroll and Samwick, 1997 and Cocco et al., 2005) in the household finance literature. The resulting implied moment conditions exploit the large cross-sectional dimension of the data to estimate the individual correlation coefficients. This is because the labor income shocks of any two individuals co-move, due to their noisy exposure to aggregate risk, in proportion to their correlation coefficients. Previous income growth models ignore instead these co-movements implied by aggregate shocks and focus on each individual earnings' growth over time to pin down correlation with the stock return.

We use the Dutch National Bank Household Survey (DNB), an annual panel covering years from 1993 to 2019, with rich information about individuals. We exploit such information to perform a minimum distance (MD) estimation on 80 clusters (MD80), grouping individuals according to time-invariant observable characteristics, such as education, risk aversion, sex and financial literacy. Finally, we study whether such MD80 correlation estimates pin down the variation in equity investments. We also apply our method on the US Panel Study of Income Dynamics (PSID). We finally repeat the entire analysis estimating directly the correlation between PI shocks and the stock

market return.

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

A concern regarding the time series approach in Bonaparte et al. (2014) is that the estimated correlations based on past income shocks may not reflect future hedging needs (Catherine et al. (2020)). Our method is immune from this problem, as it simultaneously uses inter-temporal conditions at multiple lags and cross-sectional conditions. Moreover, it pins down the correlation parameter for a given agent relying on conditions involving the co-movement in income growth between pairs of other agents. Another concern in cross-sectional studies is that both the individual's earnings risk and portfolio choice may be driven by unobserved characteristics. To address it, we add to the literature an analysis of changes in individuals' risk-taking decisions over time as a function

of revised correlations, computed after reconstructing the dynamics of individual PI shocks using a Kalman filter. We relate these revised correlations, estimated on an expanding window, to a counting variable equal to the number of waves in which the individual has invested in stocks. The regression results show that the decision to remain in the market is informed by both this revised correlation and the initial exposure to aggregate risk. Increasing correlation between PI shocks and aggregate shocks from -0.6 to 0.5 reduces participation in the equities market by a minimum of 21 months according to our estimates and by 14 months according to previous methods in Bonaparte et al. (2014). A final worry is that results wash away when shifting to correlations with stock returns. We therefore replicate all the steps of this analysis using estimated correlations between PI shocks and stock market returns, the latter being a linear combination of the aggregate risk factor and idiosyncratic noise. All results 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 and to stock returns do not matter for hedging permanent income shocks.

Our results are also relevant for portfolio (robo-) advisors. They may assign a new investor to a cluster based on their observable characteristics. Then, they may follow the optimal glide-path after calibrating a life-cycle portfolio choice model using our estimates, as we do in the Online Appendix. When correlation between PI shocks and stock market returns is equal to 0.5, the optimal share invested in stocks shifts from 0% (47%) at the beginning of working life to 11% (25%) at retirement when relative risk aversion is equal to 8 (5). When exposure to stock market risk is zero, the optimal share invested in stocks is high when young and decreases towards retirement. These findings call

for a personalized implementation of target-date funds, a common vehicle for pension investing (Mitchell and Utkus, 2020), that usually adopt the latter glide-path, only.

The rest of the paper is organized as follows. In Section 2 we benchmark related research. In Section 3, we introduce the model and provide details concerning the estimation strategy. In Section 4, we describe the data, the clusters and the distribution of the estimated correlations between PI shocks and stock market returns. In Section 5, we link stock market participation and portfolio shares to such correlation. Section 6 confirms previous results using a shorter sample, repeats the analysis using PSID data, and reports our out-of-sample analysis. Section 7 presents our concluding remarks. Section 1 of the Online Appendix updates correlation estimates for each individual over time, linking them to equities market participation revisions. It also presents the dynamics of PI shocks by cohort, sex and risk aversion, after reconstructing them with a Kalman Filter. Section 2 displays the optimal life-cycle portfolios implied by our estimates. Section 3 repeats the entire regression analysis after estimating the correlation between PI shocks and stock market returns, confirming results in the paper.

2 Related Literature

Our paper contributes to the understanding of heterogeneous hedging motives in portfolio choice. Fagereng, Guiso, and Pistaferri (2018) find substantial sensitivity of portfolio decisions to uninsured wage risk in a long panel data of firms and their workers, through an ingenious identification strategy, also demonstrating that a large part of firm-level permanent shocks is passed on to wages. While this evidence pins down the role of earnings volatility in risk taking, in line with both (Angerer

and Lam, 2009) and the time-series results in Betermier, Jansson, Parlour, and Walden (2012), the one of correlation between earnings shocks and stock returns remains inconclusive. Importantly, this leaves equity market participation largely disconnected from hedging motives.¹

Indeed, several papers look for the empirical counterpart of the theoretical link between optimal portfolios and the exposure of earnings shocks to aggregate risk. Campbell and Viceira (2002) find that 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 between stock market returns and labor income shocks is negligible in Cocco et al. (2005) and is not statistically different from zero 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. Massa and Simonov (2006) find that hedging motives cannot explain portfolio tilts away from the market portfolio. In Arrondel, Pardo, and Oliver (2010), a survey-based proxy for both correlation and earnings uncertainty explains the decision to hold risky assets for a subset of households. In Calvet and Sodini (2014), the beta of income shocks on a household's portfolio return does not co-move with that household's risky share. Against this background, Bonaparte et al. (2014) identify the role of the individual sample correlation between labor income shocks and stock market returns in explaining stock market participation, a role that is however questioned in Catherine et al. (2020). Our work shows the empirical connection between hedging motives on the one hand, and both stock market participation

¹Alternative mechanisms suggested to explain participation include a fixed participation cost (Haliassos and Michaelides, 2003), the degree of trust in the stock market (Guiso, Sapienza, and Zingales, 2008), ambiguity aversion (Dimmock, Kouwenberg, Mitchell, and Peijnenburg, 2016), a stochastic interest rate possibly correlated with earnings shocks (Munk and Sorensen, 2010), mean reversion in stock market returns (Michaelides and Zhang, 2017), among others cited in Gomes, Haliassos, and Ramadorai (2020).

and the equity share on the other.

A recent strand of research adds non-linear income risk to track both the life-cycle earnings dynamics (Guvenen, Karahan, Ozkan, and Song (2021)) and the age profile of low equity investment (Bagliano, Fugazza, and Nicodano (2019), Catherine (2022), Galvez and Paz Pardo (2022), Shen (2021)) while other mechanisms, such as participation costs, contribute to explain non participation in the equity market. In these papers the individual correlation between labour income shocks and stock market returns is either absent or mildly affects the equity share. Our new estimates indicate that individual correlations would track equity market participation also out-of-sample, while reinforcing results regarding the equity portfolio share.

According to Guvenen et al. (2017), the traditional approach to the estimation of correlation between earnings shocks and stock market returns underestimates systematic risk by ignoring the heterogeneous exposure of workers to aggregate risk factors. It therefore misinterprets the residual from the wage regression as purely idiosyncratic when in fact it contains systematic risk. Guvenen et al. (2017) estimate the 'wage betas' by clustering individuals and regressing earnings on aggregate risk factors to make the residual closer to the concept of idiosyncratic risk. Our results support the presence of a bias that is reduced by clustering. We find that the mean MD correlation between PI shocks and the aggregate shock is in-between 0.2 – 0.3, if not higher, when we group individuals, while it drops to 0.06 when we do not. This happens on both Dutch and US data. This difference rules out the possibility that the shift to the right in the distribution of MD80 correlations owes to the restrictions imposed by our parsimonious model. Both MD individual and MD80 correlations indeed rely on the same model. Differently from Guvenen et al. (2017), we however use a single

factor representation for aggregate risk. In an unreported robustness check, we apply a principal component analysis (PCA) to the covariance matrix of individuals' income growth and estimate the loading associated to each principal component. The high correlation (0.6) between the individual MD correlation estimates and the sum of such loadings indicates a close relationship between the non-parametric multi-factor representation of aggregate risk and our single factor representation. In any case, we assess the added empirical contribution of our approach by considering alternative correlation metrics in our regression analysis following Bonaparte et al. (2014), Campbell and Viceira (2002) and Guvenen et al. (2017).

Importantly, the requirement of a long time series of labor income data to identify PI shocks with traditional methods is well known (e.g., Carroll, Hall, and Zeldes, 1992; Carroll, 1997; Meghir and Pistaferri, 2004). Guvenen (2009) adopts a parsimonious representation, relative to the richer one in Guvenen et al. (2021), to provide MD estimates of the persistence of income shocks with a limited length of the data. We precisely estimate the individual exposure to both aggregate risk and stock market returns with a short time series. This is made possible by modelling and estimating the implied co-movements of earnings growth across agents' groups, that was so far ignored in both the labour economics and the household finance literature.

3 The Model and the Implied Moment Conditions

This section presents a parsimonious model of the joint distribution of earnings, a latent aggregate shock and the stock market returns. The stochastic process of the individual labor income will capture the notion that agents are differently exposed to the aggregate shock through the permanent

component of their labor income shocks.

In section 3.1, we derive the implied moment conditions. We exploit the restrictions already imposed in previous household finance literature (see e.g. Cocco et al. (2005)), implying both a zero intertemporal covariance of the idiosyncratic transitory shock for each agent and a zero covariance of the idiosyncratic component across different agents. Importantly, we additionally exploit restrictions on income growth co-movements across pairs of agents, arising from their 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. In section 3.2, we present the sample counterparts of the implied moment conditions and formalize the MD estimator for the unknown parameters. These are the correlations between the individual permanent shocks and the aggregate shock, the variance of the permanent shock and the variance of the transitory shock.

To begin, consider an economy with N individuals, indexed by i , each working for T years. At each time t , each individual receives the labor income, $Y_{i,t}$. Following Cocco et al. (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 $u_{i,t}$:

$$v_{i,t} = v_{i,t-1} + u_{i,t}, \quad (3)$$

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

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

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

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

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

where

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

and

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

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

$$u_{i,t} = \sigma_u \left(\rho_i W_t + \sqrt{1 - \rho_i^2} W_{i,t} \right) \sim \mathcal{N}(0, \sigma_u^2 \rho_i^2 + \sigma_u^2 (1 - \rho_i^2)), \quad \text{i.e.} \quad u_{i,t} \sim \mathcal{N}(0, \sigma_u^2) \quad (4)$$

The interpretation of equation (4) is simple: the variance of the PI shocks is the sum of systematic and idiosyncratic variances, where the relative weight of the systematic and the idiosyncratic

components is given by the correlation between PI shocks and the aggregate shocks, denoted by ρ_i . 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 where the variance of the noise is negligible, Section 3 of the Online Appendix allows for positive variance.

3.1 Model-Implied Moment Conditions

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

3.1.1 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 :

$$\Delta_d e_{i,t} = e_{i,t+d} - e_{i,t} \quad (5)$$

where $d = \{1, 2, \dots, D\}$, 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:

$$e_{i,t+d} - e_{i,t} = v_{i,t} + \sum_{s=t+1}^{t+d} u_{i,s} + \epsilon_{i,t+d} - (v_{i,t} + \epsilon_{i,t}) = \sum_{s=t+1}^{t+d} u_{i,s} + \epsilon_{i,t+d} - \epsilon_{i,t}, \quad (6)$$

.

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

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

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 (6):

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

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

$$\Delta_3 e_{i,t+1} = e_{i,t+4} - e_{i,t+1} = \sum_{s=t+2}^{t+4} u_{i,s} + \epsilon_{i,t+4} - \epsilon_{i,t+1}.$$

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:

$$\text{cov}(\Delta_d e_t, \Delta_d e_{t+l}) = \text{cov}\left(\sum_{s=t+1}^{t+d} u_s + \epsilon_{t+d} - \epsilon_t, \sum_{s=t+l+1}^{t+l+d} u_{i,s} + \epsilon_{t+l+d} - \epsilon_{t+l}\right), \quad (8)$$

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:

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

When $l = 0$, the inter-temporal covariance in (8) is simply equal to the cross-sectional variance of the DTS at each time t , for a given time interval of length d :

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

where $\text{var}(u_t)$ is the cross-sectional variance of the PI shocks across the N individuals, at each point in time t .²

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 $d > l$:

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

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

²This 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 into clusters (Section 4). 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 $\text{var}(u_t)$ no longer depends on W_t and simplifies to σ_u^2 . The complete derivation of $\text{var}(u_t)$ is available upon request.

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:

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

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:

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

3.1.2 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 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:

$$\begin{aligned} \text{cov}(\Delta_d e_{i,t}, \Delta_d e_{j,t}) &= \text{cov}(u_{i,t+1}, u_{j,t+1}) = \\ &= \text{cov}(\sigma_u \rho_i W_{t+1}, \sigma_u \rho_j W_{t+1}) = \sigma_u^2 \rho_i \rho_j, \end{aligned}$$

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

This feature of the economy enables exploitation of the data's cross-sectional dimension to infer the correlation parameters. In fact, while the number of correlation parameters increases linearly with N , the number of model restrictions depending on those parameters grows exponentially with N and is specifically equal to $(N(N - 1))/2$.

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:

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

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.

3.2 The MD estimator

Now, we derive the sample counterparts of the model-implied moment restrictions before turning to the Minimum Distance estimation. Let us first identify the sample counterparts of the labor income shocks. We estimate a panel regression of the log-labor income on an age polynomial up to the fourth order and a set of observable personal characteristics, including sex, education, and their interactions. The fitted value of this regression is the deterministic component of the log-labor

income, $f(t, Z_{i,t})$, with the regression residuals representing the stochastic component.

The empirical counterparts of the DTS 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 3.1.2. 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).

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

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

We denote the set of M moment conditions implied by the model 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:

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

Next, 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:

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

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

$$Q(\theta) = (g_M - G_M(\theta))' I_M (g_M - G_M(\theta)) \quad (9)$$

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

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:

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

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

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

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

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

4 Data, Clusters and MD Correlation Estimates

Our main dataset derives from the DNB Household Survey, which provides information on annual labor income for a representative sample of the Dutch population since 1993. Several reasons dictate using this survey as baseline. The first is the availability of information on financial investments at the individual level over the entire time span. Second, the DNB is the reference dataset in

the assessment of hedging motives by Bonaparte et al. (2014), which is a natural benchmark for comparing our results. Furthermore, the large cross-sectional dimension of these data may enable precise measurement of the income-return correlation parameter, also when the time dimension is short. Finally, the rich personal characteristics —including age, education, health, risk aversion— allow to group individuals according to such features. Meanwhile, we use data from the Dutch stock market index when estimating the correlation between individual labor income growth and stock market returns.

We provide details on the variables used in our analysis in Table 1 and we report descriptive statistics for the sample up to 2019 in Table 2. The average age is around 56, half of the individuals have obtained a college degree, slightly more than half are male, one out of ten individuals is unemployed, the average health status is good, and the average level of risk aversion is moderate. One-third of the sample holds stocks either directly or through mutual funds, which aligns with participation rates in other developed countries, such as the US and the UK.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.⁴

³The DNB does not provide information on the individual stocks held by individuals. It does not give information on stocks held through pension funds as these are collective, rather than individual, holdings.

⁴Unreported descriptive statistics for the short sample up to 2011, used in Section 6, reveal very similar figures for the personal characteristics of the individuals.

4.1 Individual Correlations 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 the methods in previous literature, including some cluster-based ones (Campbell and Viceira (2002) and Guvenen et al. (2017)).

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

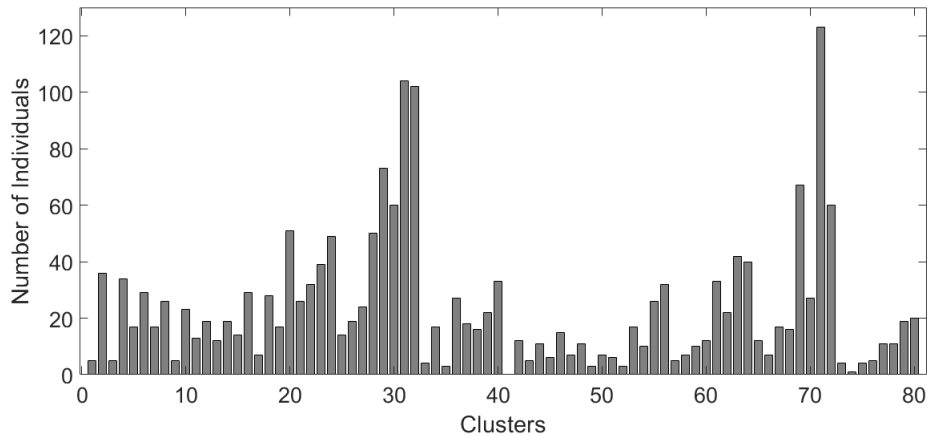
$$w_{i,t} = \sigma_u \left(\rho_k W_t + \sqrt{1 - \rho_k^2} W_{i,t}^p \right),$$

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

To ensure consistency with both the model and the empirical strategy, the clustering variables must 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* is a dummy variable equal to 1 if the individual displays high risk aversion, identified as a value of the DNB risk aversion variable greater than 5. *Financial Literacy* is a

Figure 1. Number of individuals in each cluster

The figure displays the number of individuals belonging to each of the 80 clusters formulated according to the personal traits described in the paper. Clusters are based on Education (5 groups), Sex (2), Urbanization (2), Risk aversion (2), and Financial literacy (2). The data are from the DNB Household Survey and cover waves from 1993–2019.



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 ($5 \times 2 \times 2 \times 2 \times 2$).

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, with most clusters having a similar number of individuals.

Finally, we estimate 80 correlation coefficients, which are presented in Figure 2. The left-hand panel compares the distribution of the MD80 correlations to the one of sample correlations between stock market returns and total income shocks, TS. The right-hand panel plots the MD80 correlations

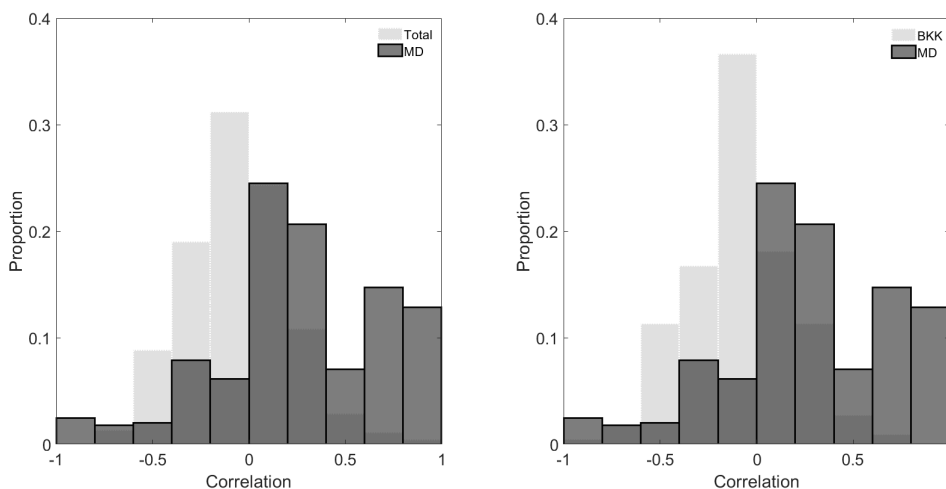
against the correlations between stock market returns and PI shocks obtained with the approach of Bonaparte et al. (2014). We also report summary statistics of the correlation parameter estimates in Table 3.

Results are striking. The distribution of MD80 correlation parameters shifts to the right with respect to both alternatives, indicating larger average hedging needs. The positive mean correlation (0.257) signals that the PI shocks of most agents have the same sign of business cycle movements (which are, in turn, reflected in the stock market return). The difference with respect to the correlation between PI shocks and stock market returns in Bonaparte et al. (2014) can be traced back to their use of realized labor income shocks, since the PI shock at time t is measured as the equally-weighted average of labor income growth at time $t - 1$, t , and $t + 1$. This distribution of estimated correlations turns out to be widely dispersed around its mean, like the one between total income shocks and stock returns. In contrast, the distribution of MD80 correlations is heavily concentrated around positive values, with the proportion of individuals characterized by a negative correlation between PI shocks and stock market returns dropping from 36%, when estimated using the approach of Bonaparte et al. (2014), to 11%. Thus, it appears that the restrictions based on the covariance matrix of contemporaneous income innovations capture co-movements in individual PI shocks that instead escape methods relying on the time series of individual income shocks only.

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

Figure 2. Distribution of cluster-based correlations

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

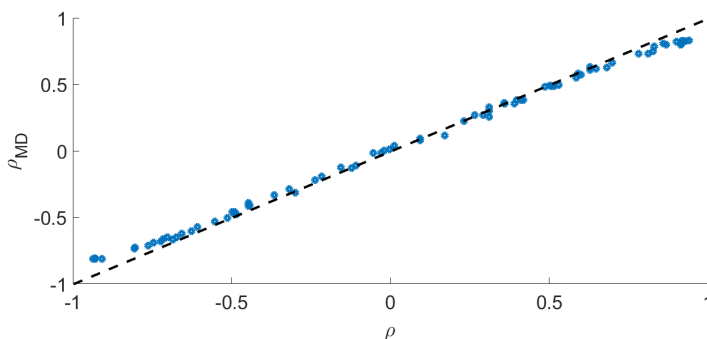


with the actual data, we simulate a time series of 30 observations ($T=30$), for 1880 individuals ($N=1880$) grouped into 80 clusters ($K=80$). Then, we apply the MD estimator to the simulated data to infer the model parameters. In Figure 3, we show that the MD80 correlations match the true, arbitrary set of correlation parameters.

A reader may be tempted to attribute the shift in the distribution of estimated correlations to the restrictions imposed by our parsimonious parametric model for PI shocks, aggregate shocks and stock market returns. This is not the case. In Table 3 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. As suggested by Guvenen et al. (2017), ignoring the differential exposure

Figure 3. 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: $\sigma_u=0.1$, $\sigma_\epsilon=0.3$, 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-degree line representing perfect inference.



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 (Güvenen and Smith (2014)).⁵

We estimate for robustness purposes the cluster-based correlations between PI shocks and stock market returns, with the method adopted in Section 3 of the Online Appendix. The resulting MD correlations are similar to the ones reported in this section, with a mean value of 0.189 as opposed

⁵Clustering also reduces the number of unknown parameters, thereby increasing efficiency.

to 0.257 (see Table 17).

5 Hedging Heterogeneous Permanent Income Shocks

We now investigate whether the new measure of hedging motives predicts financial risk-taking decisions. We also compare results based on our model of PI shocks to those using the moving average approach in Bonaparte et al. (2014). To this end, we use the same specification but for the individual correlation coefficients. We analyze the decision to participate in the stock market in Section 5.1 and the asset allocation in Section 5.2, observing the ways our estimates reveal a considerable economic impact of hedging motives on individual risk-taking.

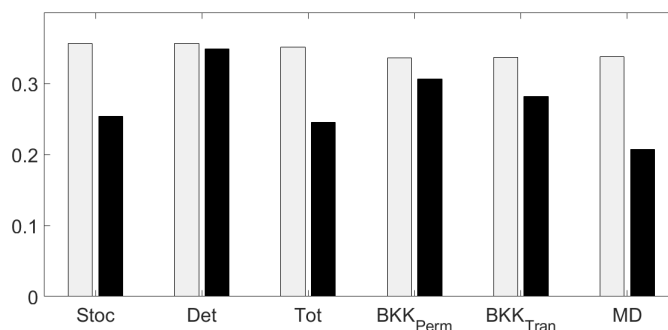
5.1 Stock Market Participation

We describe the stock market participation decisions using the dummy variable $\mathcal{I}_{i,t}$, which takes a value of 1 if the individual i invests in the stock market at time t and takes 0 otherwise. Figure 4 displays the relationship between stock market participation and measures of correlation between the stock market returns and alternative specifications of labor income: total labor income, the deterministic component (i.e., the fitted value of the panel log-labor income regression), the stochastic component (i.e., the residual of the panel log-labor income regression), the transitory and the permanent components as computed in Bonaparte et al. (2014), and the permanent component estimated using our approach.

We rank individuals according to the level of these correlation coefficients, and we plot the average participation rate for the bottom (left bar) and the top quartiles (right bar). The figure shows that

Figure 4. Stock market participation rates for low (grey bar) and high (black bar) correlation subsamples

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



individuals displaying lower correlation participate more, on average, compared to those with higher correlation, when considering both total and stochastic income shocks, in line with the income hedging hypothesis. Importantly, the figure also shows that the average participation rate varies substantially across the bottom and top quartiles when using the correlation between PI shocks and aggregate shocks estimated using our MD80 approach. This is not the case when considering the other correlation coefficients including those estimated using the approach described in Bonaparte et al. (2014).

Then, we estimate the probability of participating in the stock market by performing a probit regression, where the dependent variable is $\mathcal{I}_{i,t}$, and the explanatory variables include personal

characteristics such as income and wealth levels, age, education, sex, risk aversion, family size, and health status. We also control for situations in which the individual is retired or unemployed. The choice of control variables is motivated by the correlation that these variables may have with either direct or indirect costs of participation and align with the specification in Bonaparte et al. (2014). The results for the stock market participation probit regressions are reported in Table 4 and in Table 5.

In Table 4, columns (1) and (4) 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 large economic and statistical significance associated with the MD80 correlation estimates. In line with portfolio theory predictions, they highlight that the permanent component of the labor income shocks drives the negative relationship between income shocks and stock market participation. Estimates of the marginal effect of the control variables in 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).

In Table 5, we benchmark previous methods of measuring hedging motives that do not exploit the co-movements of individuals' income shocks. We see that the predictive power of MD80 estimates of correlation for participation holds when we control for the correlation between stock market returns and different components of income shocks computed following the methodology in Bonaparte et al.

(2014) (see columns (1)-(2)).⁶ We also include the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) (columns (3-4)) and Campbell and Viceira (2002) (columns 5-6) at the cluster-level. We obtain β_{GUV} by estimating pooled OLS regression of the (log)-real earnings growth of individual i in year t over the stock market return in year t separately by cluster, so that individuals belonging to the same cluster share the same β_{GUV} .⁷ To estimate β_{CV} , we first compute the cluster-specific average of the (log)-real earnings growth across the individuals belonging to each cluster, in order to obtain cluster-specific permanent income shocks under the restriction that transitory income shocks cancel out across individuals. Then, we regress cluster-specific permanent income shocks in year t over the stock market return in year $t + 1$, following the approach of Campbell and Viceira (2002). These cluster-based betas display statistical significance (see columnn 3 and 5). However, this disappears when we also allow in the regression the MD80 correlation estimates.

In Table 5, the regression coefficient for the correlation between PI shocks and stock market returns computed using the approach of Bonaparte et al. (2014) is smaller in magnitude compared with the regression coefficient for the correlation between PI shocks and stock market returns estimated using our MD80 approach. We can gauge the economic significance of hedging motives implied by alternative estimates by computing the marginal effect of the variables on the propensity to invest.

The probability of participation in the stock market, according to the Probit model, is given by:

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

⁷We cannot account for exposure to both employer- and industry-level risk when using the DNB survey because of data availability.

$$P(\mathcal{I}_{i,t} = 1) = \Phi \left(\sum_{k=1}^K \beta_k X_{k,i,t} \right), \quad (10)$$

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:

$$\frac{\partial P(\mathcal{I}_{i,t} = 1)}{\partial X_{k,i,t}} = \frac{\partial \Phi(\beta' X)}{\partial X_{k,i,t}} = \phi(\beta' X) \beta_k, \quad (11)$$

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 (10). 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 $\rho = -0.6$ and $\rho = 0.5$, respectively. Using the estimated coefficient of MD80 correlations (column (6) in Table 4) the first individual's propensity to participate is 12% higher than the second individual's propensity to participate $((-0.6 - 0.5) * (-0.267) * \phi(\hat{\beta}' X))$. Using instead the results of column (1) in Table 5, 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).

These results highlight both the economic and statistical significance of individuals' exposure to aggregate risk for stock market participation, when identified through co-movements of income shocks across groups of individuals. The previous Probit analysis, like most previous literature, treats each observation over time as a separate agent. Results go through when we follow each individual over time. In Section 1 of the Online Appendix we exploit the panel dimension of our data to see whether revisions in individual correlations explain revisions in portfolios choices for each individual. We reconstruct the dynamics of PI shocks of each individual based on realizations of both income and stock market returns (see an illustration in Figures (9)-(11)). We then estimate updated correlations (see Table 3, KF estimates) before linking them to revisions in both participation (Table 15) and asset allocation (Table 16). This is a second check on the quality of our approach to estimating PI shocks.

Importantly, results go through when we compute correlations between PI shocks and stock market returns (Table 17) and predict participation (Table 18) after modelling the stock market returns as a linear combination of the aggregate risk factor and idiosyncratic noise (Section 3 of the Online Appendix). In the next sections, we will show that results are not sensitive to other details of the analysis and to the data.

5.1.1 Robustness

We address potential endogeneity issues arising from using individual wealth as control variable in our regressions in two ways. First, we control for non-financial wealth only, rather than total

wealth, excluding the financial items of net worth. Non-financial wealth in DHS accounts for the individual housing property. Second, we construct dummy variables for quartiles of wealth and we include these dummies as control variables in our regressions instead of the original continuous variable, following the approach of Van Rooij, Lusardi, and Alessie (2011). In both cases, we obtain results that are quantitatively equivalent to those reported in the regression tables.

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

We also repeat the analysis using as our independent variable 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.

5.2 Asset Allocation

This section studies the asset allocation decision. The dependent variable is the share of the individual portfolio invested in stocks either directly, through mutual funds, or both. The main explanatory variables of interest are the same as for the earlier analysis concerning the decision to participate.

Table 6 reports results from the Tobit regression for individual asset allocation (see columns (1)-(6)). We see that the correlation between PI shocks and aggregate shocks matters for the asset allocation decisions, as it always significantly predicts the fraction of wealth invested in risky assets. This result holds whether or not we alternatively consider, in unreported regression, the total equity share of the individual and either direct investments in stocks or indirect investment in stocks. In columns (4)-(6), we control for the set of observable characteristics that may impact an individual's portfolio allocation. We find that richer individuals, who are more educated and less risk-averse, generally invest larger fractions of their wealth in stocks, either directly or through mutual funds or both. As in Bonaparte et al. (2014), we also find that high income-risk individuals prefer to directly allocate wealth to stocks, rather than through mutual funds.

Vissing-Jorgensen (2002), Fagereng, Gottlieb, and Guiso (2017) and Bonaparte et al. (2014) consider simultaneously the market participation and asset allocation decisions, by estimating Heckman (1979) regressions in which the control variables in the selection model (i.e, the participation regression) and the control variables in the asset allocation regression are the same. We follow the same approach and report the results of Heckman (1979) regression estimates in columns (7)-(8). We consider lagged financial wealth and lagged squared financial wealth as additional control vari-

ables, 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.⁸

Table 7 reports the benchmarking analysis. The alternative coefficients of the correlation between PI shocks and stock market return display the same order of magnitude and the same same level of statistical significance (5%) in columns (1),(3) and (5). When we add the MD80 correlation term in columns (2),(4) and (6), the beta estimated following Guvenen et al. (2017) loses explanatory power.

The previous results show that hedging motives, estimated through co-movements across agents' earnings shocks driven by a latent aggregate shock, predict heterogeneous (conditional) asset allocation decision. Similar results hold when we measure the correlation between the PI shocks and the stock market returns, letting the latter be a combination of the aggregate risk factor and noise (Table 19 in the Online Appendix).

⁸If 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.

6 Small T, Out-of-Sample Analysis and PSID Data

The previous section exploits data characterized by a relatively large number of waves, from 1993 until 2019. Often, the temporal dimension of the data that is available to the researcher or the portfolio manager is much smaller. Section 6.1 will therefore show that our results hold when we use a smaller number of waves. This experiment demonstrates the robustness of the MD estimation method that exploits both the cross-sectional and the time series dimension of the data. Such robustness allows us to set aside some observations in order to perform an out-of-sample analysis, that will be presented in Section 6.2. This will show that the estimated correlation coefficients predict participation and asset allocation also out-of-sample, suggesting their use for improving on portfolio design. Last but not least, Section 6.3 confirms the results on both the size of hedging needs and the determinants of participation on U.S. data from PSID. This confirmation also indicates that results are not an artifact of the clusters' characteristics since we necessarily have to change them based on data availability.

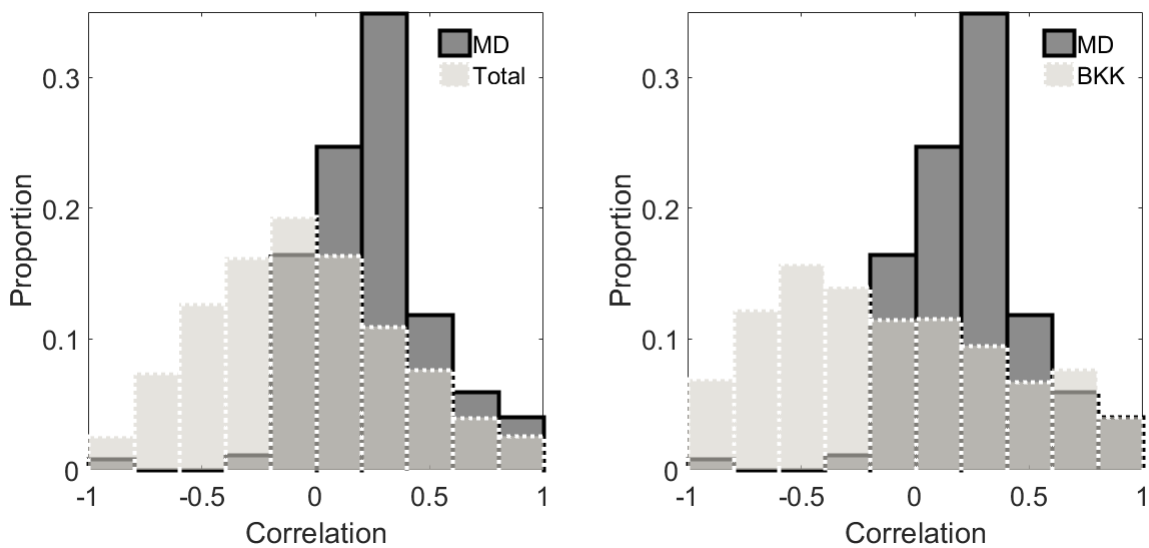
6.1 MD Estimates with small T

Table 8 display the distribution of alternative estimates of individual exposure to stock market returns. The cluster-based, MD estimates confirm both the relatively high average correlation parameter between PI shocks and aggregate risk (0.3) when compared to both prior moving average estimates and the non-cluster-based individual MD estimates.

Table 9 reports the probit regression results for stock market participation using the DNB survey waves from 1993 to 2011 in columns (1)-(3) and to 2007 in columns (4)-(6). This table shows the

Figure 5. Distribution of cluster-based correlations

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



statistical and economic significance of the MD correlation coefficients in predicting participation, similar to the one based on surveys up to and including 2019. There are minor changes such as those in the statistical significance of some control variables.

Similarly, benchmarking results in Table 10 confirm the relative strength of the MD method. In the sample including waves up to 2011, the BKK-correlation estimates with PI shocks have a statistical significant coefficient, as reported in Bonaparte et al. (2014), while competing methods lose explanatory power. The economic significance of hedging motives, when assessed through the

coefficient of MD correlation estimates, becomes even larger than in the full sample. Moreover, this coefficient is around three times larger than the coefficient for the correlation between PI shocks and stock market returns estimated according to Bonaparte et al. (2014).

This section shows that our results on both the size of hedging motives and the sensitivity of participation to them hold irrespective of the length of the sample. Further robustness checks are available. Similar results hold on sub-samples by education and retirement status, as well as by focusing on different dependent variables (Only Stocks or Only Mutual Funds). Both the portfolio share allocated to equities and revisions in the decision to participate in the stock market over time respond to correlation estimates in the short samples as they do in the longer one.

6.2 Out-of-Sample Analysis

Results presented so far show that income hedging motives are more relevant to risk-taking decisions than previously thought, also over time for each individual. The previous subsection adds that estimates of such hedging motives are precise, even given a limited time-series data dimension. This section exploits this last feature to implement an out-of-sample analysis. This exercise reveals that it is possible to use MD correlation estimates for portfolio management, using only past information.

The out-of-sample analysis starts with the 80 MD correlation parameter estimates at the cluster-level, using data up to 2011. When considering data up to 2019, we allocate each new survey participant to one of the 80 clusters (defined in Section 4.1) on the basis of their personal characteristics. We then attribute to each individual the correlation parameter of the corresponding

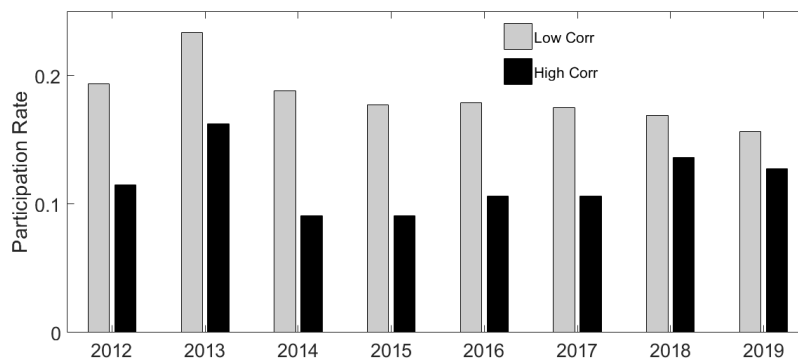
cluster estimated in the previous step using data up to 2011. That is, we use *ex-ante* measures of future income risk, as suggested by Guiso, Jappelli, and Terlizzese (1996). Then, we relate these correlation estimates to the decision to invest in equities in the years after 2011 (i.e., 2012–2019).

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 6, display a systematic pattern: individuals belonging to clusters with lower estimated correlation parameters participate more than individuals allocated to clusters displaying higher correlation. The difference between the top and bottom quartiles is remarkable. 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 will be larger when we estimate the clustered correlations using data up to 2017.

The out-of-sample response of equity investments to correlation is confirmed through both a Probit analysis of participation and a Tobit analysis of the portfolio equity share in Table 11. In Table 11, equity investment is alternatively defined as the total one (columns 1 and 4), held both directly or indirectly through mutual funds, or just the direct one (columns 2 and 5) or just the indirect one (columns 3 and 6). In the Probit, the dependent variable is a dummy, indicating participation in the stock market between 2012 and 2019, while in the Tobit it is the portfolio share held in stocks during the same years. Results demonstrate that the clustered correlations, estimated using data up to 2011, negatively and significantly predict both stock market participation and the portfolio share allocated to equities for the period 2012–2019. The economic significance of our estimates is

Figure 6. Out-of-Sample prediction of stock market participation

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



also remarkable, in line with the in-sample results. For instance, the coefficient in column (1) of Table 11 suggests that an individual allocated to a cluster displaying high correlation ($\rho=0.5$) is 8% less likely to participate in the stock market than an individual allocated to a low correlation cluster (see equation (10) in Section 5.1).

6.3 Correlation and Probit Estimates based on PSID

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

In PSID, we do not have information about risk aversion and financial literacy of individuals.

Also, the sample is almost entirely composed by men. On the other hand, information about the industry of the household head's job is available and we do have information about the education level. Thus, to estimate the correlation parameters at the cluster-level using the Minimum Distance (MD) methodology as we do with the DHS data, we form 48 clusters based on 4 education groups and 12 industries as in Campbell and Viceira (2002). We also estimate both β_{GUV} and β_{CV} at the cluster-level.

We then use these correlation parameters in a probit analysis of stock market participation. We report results from the probit regression in Table 13, where the dependent variable refers to ownership of equities or mutual funds. Results are comparable to the ones in Table 9-10, as the sample covers the same years and the main independent variables are the same.⁹ In the probit regression estimates on PSID data, the statistical significance of the coefficients associated with the MD correlation parameters is in line with the one obtained using the DHS data. Moreover, the statistical significance of the coefficients associated with alternative methods is lower, confirming results of the previous benchmarking exercise. We therefore conclude that our method captures heterogeneous hedging motives also on PSID data.

We summarize the correlation coefficient estimates in Table 12. Some parameters estimated on PSID data tend to be larger in size and generally positive compared to those estimated on the Dutch data. For instance, the PSID beta coefficients are consistent with the values obtained by β_{GUV} and β_{CV} , respectively, using US data.¹⁰ Importantly, the MD correlation coefficients

⁹Income in PSID refers to the household, differently from DHS.

¹⁰When we use only three education clusters to compute β_{CV} , we obtain values in line with those estimated by Campbell and Viceira (2002) on the same groups.

estimated on PSID are very similar to those reported in Table 8 based on the DHS over the same years.

Finally, Table 14 reports the estimated correlations for different industries, computed as the average of the MD correlation parameters across all the clusters that include a given industry. For many industries, the estimates are in line with the ones presented in Campbell and Viceira (2002) and reported in the second column of the table.

7 Summary and Conclusions

This paper models the co-movements of earnings growth across groups of individuals that arise due to their heterogeneous exposure to aggregate shocks. It then exploits these co-movements to assess the individuals' hedging motives. Our estimates of the mean individual exposure to aggregate shocks lie in, or above, the range 0.2-0.3. This result appears both in the Dutch Household Survey and in the US PSID data, despite the non-homogeneous groups of individuals covered by these surveys, and is insensitive to the length of the sample.

Since stock returns co-move with aggregate shocks, the individual with the mean correlation hedges labour income risk by reducing exposure to the equity market with respect to an individual without exposure to aggregate risk. Such reduction exceeds the one found in previous studies. This implies that observed portfolios are closer than previously thought to the implications of portfolio choice theory, without resorting to alternative mechanisms.

Results concerning equity investments support our measurement of hedging motives, that exploits

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 confirm the theoretical prediction that equity investments fall when permanent income is exposed to aggregate risk. These results are robust, holding both in-sample and in out-of-sample experiments. They also hold both in the cross section and for each individual over time. Since earnings risk is a central issue in the economics of incomplete markets, these advances in the measurement of hedging motives may prove useful beyond the boundaries of household finance.

References

- Angerer, X., and P.S. Lam, 2009, Income risk and portfolio choice: an empirical study, *Journal of Finance* 64, 1037–1055.
- Arrondel, Luc, Hector Calvo Pardo, and Xisco Oliver, 2010, Temperance in stock market participation: Evidence from France, *Economica* 77, 314–333.
- Bagliano, F.C., C. Fugazza, and G. Nicodano, 2014, Optimal life-cycle portfolios for heterogeneous workers, *Review of Finance* 18, 2283–2323.
- Bagliano, F.C., C. Fugazza, and G. Nicodano, 2019, Life-cycle portfolios, unemployment and human capital loss, *Journal of Macroeconomics* 60, 325–340.
- Benzoni, L., P. Collin-Dufresne, and R.G. Goldstein, 2007, Portfolio choice over the life-cycle when the stock and labour markets are cointegrated, *Journal of Finance* 62, 2123–2167.
- Betermier, Sebastien, Thomas Jansson, Christine Parlour, and Johan Walden, 2012, Hedging labor income risk, *Journal of Financial Economics* 105, 622–639.
- Bonaparte, J., G. Korniotis, and A. Kumar, 2014, Income hedging and portfolio decisions, *Journal of Financial Economics* 113, 300–324.
- Calvet, Laurent E, and Paolo Sodini, 2014, Twin picks: Disentangling the determinants of risk-taking in household portfolios, *The Journal of Finance* 69, 867–906.
- Campbell, J. Y., J. Cocco, F. Gomes, and P. Maenhout, 2001, Investing retirement wealth: a life-cycle model, *Risk Aspects of Investment-Based Social Security Reform* University of Chicago Press, 439–483.
- Campbell, J.Y., and L. Viceira, 2002, Strategic asset allocation: Portfolio choice for long-term investors, *Oxford University Press, Oxford, UK* .

- Carroll, Christopher D, 1997, Buffer-stock saving and the life cycle/permanent income hypothesis, *The Quarterly journal of economics* 112, 1–55.
- Carroll, Christopher D, Robert E Hall, and Stephen P Zeldes, 1992, The buffer-stock theory of saving: Some macroeconomic evidence, *Brookings papers on economic activity* 1992, 61–156.
- Carroll, C.S., and A. Samwick, 1997, The nature of precautionary wealth, *Journal of Monetary Economics* 40, 41–71.
- Catherine, S., 2022, Countercyclical income risk and portfolio choices over the life-cycle, *Review of Financial Studies*. 35, 4016–4054.
- Catherine, S., P. Sodini, and Y. Zhang, 2020, Countercyclical income risk and portfolio choices: Evidence from sweden, *Swedish House of Finance Research Paper* 20-20).
- Cocco, J., F. Gomes, and P. Maenhout, 2005, Consumption and portfolio choice over the life cycle, *Review of Financial Studies* 18, 491–533.
- Dimmock, S.G., R. Kouwenberg, O.S. Mitchell, and K. Peijnenburg, 2016, Ambiguity aversion and household portfolio choice puzzles: Empirical evidence, *Journal of Financial Economics* 199, 559–577.
- Fagereng, A., C. Gottlieb, and L. Guiso, 2017, Asset market participation and portfolio choice over the life cycle, *Journal of Finance* 72(2), 705–750.
- Fagereng, A., L. Guiso, and L. Pistaferri, 2018, Portfolio choices, firm shocks, and uninsurable wage risk, *The Review of Economic Studies* 85, 437–474.
- Galvez, J., and G. Paz Pardo, 2022, Richer earnings dynamics, consumption and portfolio choice over the life cycle, *Netspar Conference* 1–64.
- Gomes, F. J., M. Haliassos, and T. Ramadorai, 2020, Household finance, *Journal of Economic Literature* forthcoming.

- Guiso, L., T. Jappelli, and D. Terlizzese, 1996, Income risk, borrowing constraints, and portfolio choice, *The American Economic Review* 86(1), 158–172.
- Guiso, L., P. Sapienza, and L. Zingales, 2008, Trusting the stock market, *The Journal of Finance* 63, 2557–2600.
- Guvenen, F., 2009, Empirical investigation of labor income process, *Review of Economic Dynamics* 12.
- Guvenen, F., 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., S. Schulhofer-Wohl, J. Song, and M. Yogo, 2017, Worker betas: Five facts about systematic earnings risk, *American Economic Review Papers and Proceedings* 107, 398–403.
- Guvenen, Fatih, and Anthony A Smith, 2014, Inferring labor income risk and partial insurance from economic choices, *Econometrica* 82, 2085–2129.
- Haliassos, M., and A. Michaelides, 2003, Portfolio choice and liquidity constraints, *International Economic Review* 44, 144–177.
- Heaton, J., and D. Lucas, 2000, Portfolio choice and asset prices: The importance of entrepreneurial risk, *Journal of Finance* 55, 1163–1198.
- Heckman, James J, 1979, Sample selection bias as a specification error, *Econometrica: Journal of the econometric society* 153–161.
- Massa, M., and A. Simonov, 2006, Hedging, familiarity and portfolio choice, *The Review of Financial Studies* 19, 633–685.
- Meghir, Costas, and Luigi Pistaferri, 2004, Income variance dynamics and heterogeneity, *Econometrica* 72, 1–32.

- Merton, R., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, *Review of Economics and Statistics* 51, 247–257.
- Michaelides, A., and Y. Zhang, 2017, Stock market mean reversion and portfolio choice over the life cycle, *Journal of Financial and Quantitative Analysis* 52(3), 1183–1209.
- Mitchell, O.S., and S.P. Utkus, 2020, Target date funds and portfolio choice in 401(k) plans, *Working Paper Pension Research Council* .
- Munk, C., and C. Sorensen, 2010, Dynamic asset allocation with stochastic income and interest rates, *Journal of Financial Economics* 96, 433–462.
- Shen, J., 2021, Countercyclical risks, consumption and portfolio choice: Theory and evidence, *University of Missouri* .
- Van Rooij, M., A. Lusardi, and R. Alessie, 2011, Financial literacy and stock market participation, *Journal of Financial Economics* 101, 449–472.
- Viceira, L.M., 2001, Optimal portfolio choice for long-horizon investors with nontradable labor income, *Journal of Finance* 41, 433–470.
- Vissing-Jorgensen, A., 2002, Towards an explanation of household portfolio choice heterogeneity: nonfinancial income and participation cost structures, *National Bureau of Economic Research* .

Table 1 Definitions of variables

Variable	Definition
OwnSTK	One if own stocks and zero otherwise.
OwnMF	One if own mutual funds and zero.
OwnSTKMF	One if own stocks or mutual funds and zero otherwise.
PropSTK	Financial wealth fraction invested in stocks.
PropMF	Financial wealth fraction invested in mutual funds.
PropSTKMF	Financial wealth fraction invested in stocks or mutual funds.
Ln(NetWorth)	(log)-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.

Table 2 Summary statistics. Full Sample

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

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

Table 3 Correlation Parameters from DHS. Summary Statistics on Full Sample.

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

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

Table 4 Probit Estimates for Stock Market Participation

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

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

Table 5 Probit Estimates for Stock Market Participation (Benchmarking)

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

OwnSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.480*** (0.117)	-0.504*** (0.118)	-0.141*** (0.037)	-0.153*** (0.038)	-0.139*** (0.038)	-0.152*** (0.038)
Corr(PermMD80)		-0.185*** (0.065)		-0.276*** (0.037)		-0.261*** (0.035)
Corr(tranBKK,Rm)	0.084 (0.113)	0.059 (0.114)				
Corr(permBKK,Rm)	-0.175 (0.104)	-0.164 (0.105)				
Beta (Guvenen et al. (2017))			0.221* (0.086)	-0.063 (0.093)		
Beta (C&V(2002))					-0.228** (0.087)	-0.054 (0.088)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	3,286	3,286	12,758	12,758	12,758	12,758
Pseudo R ²	0.320	0.322	0.272	0.276	0.272	0.277

Table 6 Tobit Estimates for Asset Allocation

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

	PropSTKMF				Heckman			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
St. Dev.	-0.125*** (0.023)	-0.131*** (0.023)	-0.095** (0.012)	-0.052* (0.024)	-0.050* (0.024)	-0.033* (0.014)	-0.011*** (0.002)	0.001 (0.005)
Corr(Tot,Rm)	-0.173*** (0.028)			-0.175*** (0.026)				
Corr(Det,Rm)		0.147*** (0.030)			0.098** (0.028)			
Corr(Stoc,Rm)		-0.184*** (0.027)			-0.182*** (0.027)			
Corr(PermMD80)			-0.217*** (0.012)			-0.085** (0.013)	-0.049*** (0.003)	-0.018*** (0.005)
(log)-Income				-0.029* (0.014)	-0.033* (0.014)	-0.007 (0.010)		-0.005 (0.005)
(log)-Wealth				0.105*** (0.007)	0.105*** (0.007)	0.107*** (0.005)		0.011*** (0.004)
HH Size				-0.043*** (0.007)	-0.041*** (0.020)	-0.037*** (0.005)		-0.011*** (0.002)
Age				0.001 (0.004)	-0.002 (0.004)	0.017** (0.003)		-0.001 (0.001)
Education				0.083*** (0.015)	0.086*** (0.015)	0.075*** (0.011)		0.028*** (0.006)
Sex				0.078** (0.020)	0.079*** (0.020)	0.013 (0.014)		0.004 (0.005)
Unemployed				-0.001 (0.036)	0.004 (0.035)	-0.001 (0.029)		-0.003 (0.009)
Retired				-0.001 (0.024)	0.004 (0.024)	0.027 (0.017)		0.028*** (0.007)
Health				0.010 (0.010)	0.009 (0.010)	0.007 (0.008)		0.005 (0.003)
Risk Aversion				-0.127*** (0.004)	-0.126*** (0.004)	-0.134*** (0.003)		-0.043*** (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*** (0.008)	0.074*** (0.014)

Table 7 Tobit Estimates for Asset Allocation (Benchmarking)

The table reports the Tobit regression results for asset allocation decision. The dependent variable is the portfolio shares in stocks either directly or through mutual funds (propSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the correlation between the permanent component of labor income shocks and the aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), and different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (tranBKK and permBKK, respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. We include the same control variables as in Table 4, but we suppress coefficients to save in space. We report in parentheses the Tobit-robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

PropSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.173*** (0.037)	-0.185*** (0.038)	-0.029* (0.014)	-0.033* (0.014)	-0.028* (0.014)	-0.033* (0.014)
Corr(PermMD80)		-0.064** (0.067)		-0.086*** (0.014)		-0.082*** (0.013)
Corr(tranBKK,Rm)	0.075* (0.036)	0.070 (0.036)				
Corr(permBKK,Rm)	-0.086* (0.033)	-0.079* (0.034)				
Beta (Guvenen et al. (2017))			0.079* (0.039)	-0.011 (0.037)		
Beta (C&V(2002))					-0.086* (0.034)	-0.026 (0.035)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	3,224	3,224	12,498	12,498	12,498	12,498
Pseudo R ²	0.366	0.368	0.289	0.292	0.289	0.292

Table 8 Correlation Parameters from DHS. Summary Statistics on Short Sample.

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

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

Table 9 Probit Estimates for Stock Market Participation. Short Sample

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

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

Table 10 Probit Estimates for Stock Market Participation (Benchmarking). Short Sample

The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the correlation between the permanent component of labor income shocks and the aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), and different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (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 9, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. We use data from the DNB Household Survey for the waves from 1993–2011.

OwnSTKMF						
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.173** (0.067)	-0.191** (0.051)	-0.139** (0.067)	-0.149** (0.047)	-0.140** (0.048)	-0.150** (0.039)
Corr(PermMD80)		-0.289*** (0.067)		-0.284*** (0.058)		-0.289*** (0.058)
Corr(tranBKK,Rm)	-0.073 (0.052)	-0.088 (0.052)				
Corr(permBKK,Rm)	-0.094* (0.042)	-0.085* (0.042)				
Beta (Guvenen et al. (2017))			0.190 (0.109)	0.152 (0.110)		
Beta (C&V(2002))					0.053 (0.099)	0.038 (0.099)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	5,677	5,677	7,745	7,745	7,745	7,745
Pseudo R ²	0.262	0.265	0.252	0.259	0.256	0.259

Table 11 Out-Of-Sample Probit & Tobit Estimates

The table reports the out-of-sample regression results for stock market participation decision 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 (permMD80) and data up to 2011. This correlation is then assigned to each individual on the basis of the corresponding cluster to which the individual is allocated according to her personal characteristics observed in 2012. We also include other independent variables, and additional control variables as defined in Table 4. N is the number of observations. 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	OwnSTK	OwnMF	PropSTKMF	PropSTK	PropMF
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev. (dy)	-0.239*** (0.078)	-0.168** (0.090)	-0.207** (0.075)	-0.060** (0.026)	-0.034 (0.031)	-0.081** (0.032)
Corr(PermMD80)	-0.207*** (0.073)	-0.140* (0.087)	-0.136* (0.075)	-0.081*** (0.027)	-0.069** (0.032)	-0.060* (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 R ²	0.324	0.273	0.264	0.354	0.317	0.280

Table 12 Correlation and Beta Parameters. Summary Statistics (PSID)

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

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

Table 13 Probit Estimates for Stock Market Participation (PSID)

The table reports the probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stock or mutual funds (OwnSTKMF). The main independent variables are the standard deviation, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (permMD80), and the correlation between stock market return and different specifications of labor income shocks: the labor income growth rate (Tot), the stochastic component of labor income growth rate (Stoc), the deterministic component of labor income growth rate (Det). We also include different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (transBKK and permBKK, respectively), and the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. Additional control variables are individual demographic characteristics, such as the (log)-labour income, marital status, family size, age and years of schooling. We also control for year-fixed effects. We report in parentheses the Probit-robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. We use data from the Panel Survey Income Dynamics (PSID) for the waves from 1988–2011.

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

Table 14 Correlations By Industry: PSID Data.

The table reports the estimated correlations between the stock market return and the permanent component of labor income shocks estimated by using the minimum distance methodology (MD) at the cluster level and the correlation coefficients between the labour income growth rate and the stock market returns reported in Campbell and Viceira (2002), for different industries to which individuals belong to. The data are from the Panel Survey Income Dynamics (PSID) and cover waves from 1988–2011. The MD correlation parameters at the industry level are obtained by computing the average of the estimated correlation parameters across the clusters that include a given industry, where the clustering variables used for the PSID data are Education and Industry. The industry classification is the two-digit code adopted by the PSID.

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

Hedging Permanent Income Shocks.

Online Appendix

Fabio C. Bagliano¹, Raffaele Corvino¹, Carolina Fugazza¹ and Giovanna Nicodano²

¹University of Torino and CeRP (Collegio Carlo Alberto)

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

Sunday 25th June, 2023

This Appendix is organized as follows. In Section 1 we reconstruct the individual PI shocks and correlations over time with a Kalman Filter (KF). This exploits the relationship between labor incomes and stock market returns, their realizations and the MD80 parameter estimates. Despite a common initial cluster-based correlation, different realizations of individual labor income imply heterogeneous updated correlations. In section 1.3 we examine whether our estimates deliver patterns of PI shocks that align with economic intuition. Section 2 calibrates a life-cycle model. Section 3 confirms results in the paper measuring the correlation between PI shocks and stock returns, when the latter are imperfectly correlated with the aggregate shocks.

1 Learning about Income Hedging Needs

Hedging choices change over time because earnings and return realizations affect them through both budget constraints and learning (as in Chang, Hong, and Karabarbounis, 2018). In this section we follow each individual over time to assess the relationship between changes in the participation decision and revisions in the correlation between PI shocks and the aggregate shocks.

To this end, we first reconstruct the sequence of unobservable PI shocks at the individual-level using a Kalman filter. This exploits the assumed relationship between labor incomes and stock market returns, their realizations and the MD80 parameter estimates. We then construct the sequence of updated correlation coefficients between the PI shocks and stock market returns over an expanding window. These revised correlation coefficients will belong to the set of independent variables explaining the individual's probability to participate in the stock market in each period and the overall frequency of individual participation.

1.1 Updating Permanent Income and Correlation Estimates

Based on the model in Section 3 of the paper, we obtain an equation that relates the observable total income shocks to the latent permanent component:

$$e_{i,t} = v_{i,t} + \epsilon_{i,t}, \quad (1)$$

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

$$v_{i,t} = v_{i,t-1} + u_{i,t}, \quad (2)$$

where

$$u_{i,t} = \sigma_u (\rho_i (r_t / \sigma_r)) + \omega_{i,t}, \quad (3)$$

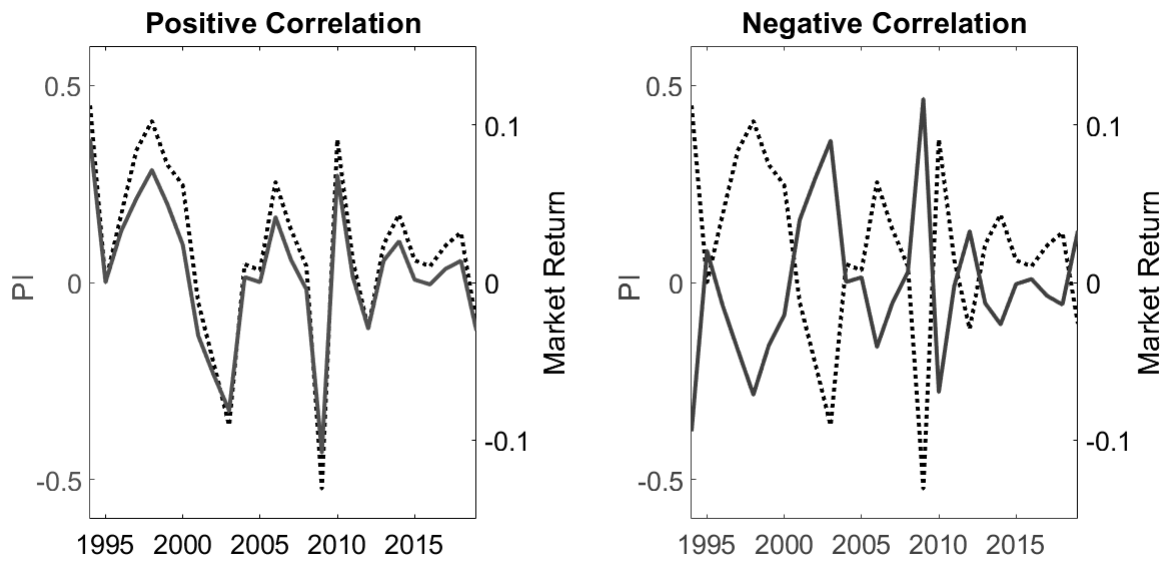
since $r_t = \sigma_r W_t$ and $\omega_{i,t} = \sigma_u \left(\sqrt{1 - \rho_i^2} W_{i,t}^p \right) \sim \mathcal{N}(0, \sigma_u^2 (1 - \rho_i^2))$.¹

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 . The model in Section 3 of the paper allows exploiting also the information on the stock market returns to infer the latent variable. We next form a prediction of the total income shock $e_{i,1}$, $\hat{e}_{i,1}$, by computing the expected value of $e_{i,1}$ conditional on $\hat{v}_{i,1}$. The difference between the actual and the predicted total income shocks is the measurement error that is used to compute the posterior for $v_{i,1}$, which turns to be the prior for the next point in time. We iterate the system up to T and we reconstruct the

¹Here we let the aggregate shock and the stock market return be perfectly correlated, while Section 3 relaxes this assumption.

Figure 7. 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.



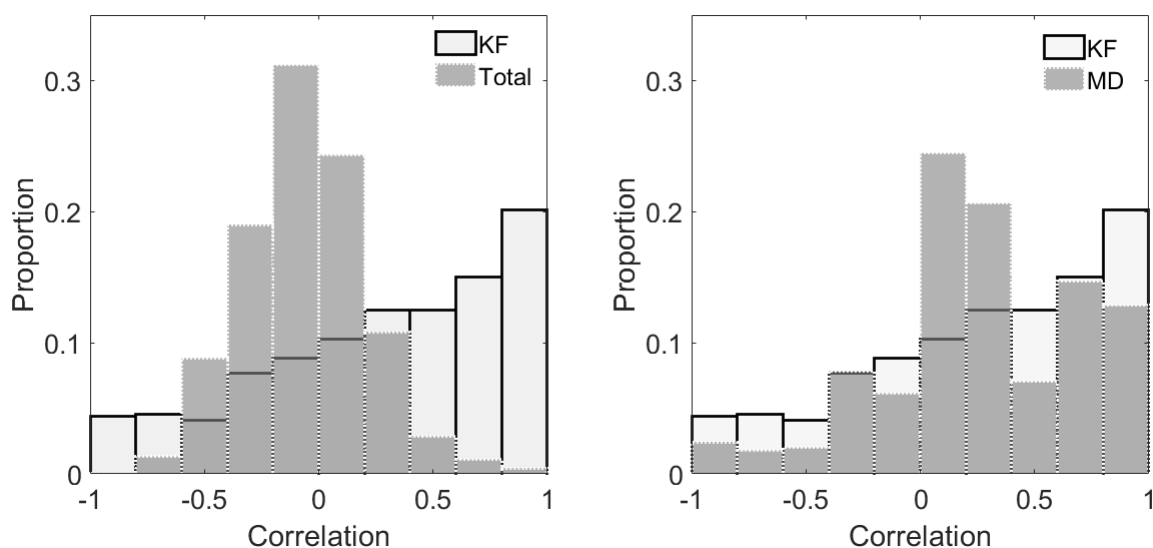
PI shocks by computing the first differences of the random walk. Consequently, for each individual, we obtain the dynamics of the permanent shocks to the labor income over the entire time series.

To represent our results, Figure 7 plots the dynamics of the stock market returns and the PI shocks reconstructed using the KF for the individual with minimum and maximum sample correlation between stock market returns and PI shocks, respectively.

Meanwhile, Figure 8 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

Figure 8. 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.



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.

1.2 Participation Frequency

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

We track revisions in the correlation between PI shocks and stock market returns by sequentially computing, 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. Then, we estimate a probit model in which the dependent variable is equal to 1 if the individual invests in stocks, either directly or through mutual funds, in wave t , and the main independent variable is the correlation between PI shocks and stock market returns up to t .

Table 15 shows that the revised correlations significantly predict the sequence of individual decisions to participate in the stock market in subsequent waves. It demonstrates that the lower the revised correlation between stock market returns and PI shocks, the higher the propensity of the individual to enter (or remain) in the equities market. Both the economic and statistical significance of the MD correlation between PI shocks and stock market returns increase when we also control for the revised correlation, since they are based on different types of information. The MD estimates on the one hand exploit the information embedded in the variance-covariance matrices of total income shocks, both the intertemporal one for each agent and the contemporaneous one across agents, 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 shocks over time. Thus, the latter information increase the relevance of MD correlation estimates, while both cluster and covariance information increase the economic relevance of the individual KF correlation.

Finally, we study the frequency of participation in the stock market for the whole sample using a Poisson regression, where the dependent variable is a discrete counting variable equal to the number of waves in which the individual invested in stocks. We use the same explanatory variables as in Table 4, including all of the unreported control variables. Table 16 confirms that individuals remain in the market longer if their labor income shocks are negatively correlated with stock market returns.

Importantly, the correlation between stock market returns and PI shocks has a negative and significant impact when coupled with the KF-based one, which is larger than that of the correlation computed using previous approaches.

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

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

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

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

Consider, for instance, an individual participating in the stock market over 10 years. Increasing the revised correlation between PI shocks and stock market returns from -0.6 to 0.5 reduces participation by 21 months $((-0.6 - 0.5) \times 0.162 \times 10)$ (38 months if we consider 0.239, the sum of the coefficients of both MD and KF correlation terms). Instead, if we similarly increase the correlation between PI shocks and stock market returns, computed according to Bonaparte, Korniotis, and Kumar (2014), the individual reduces participation in the equities market by 14 months $((-0.6 - 0.5) \times 0.11 \times 10)$. These results are free from unobserved heterogeneity concerns that may plague cross-sectional results. They support both our approach to the measurement of hedging needs and the role of correlation implied by life-cycle theory.

1.3 Income Shocks and Participation by Cohort and by Cluster

In this section, we challenge our Kalman Filter-based correlation estimates asking whether they deliver PI shocks that align with economic intuition. We scrutinize the distribution of both PI and total shocks (TS) across clusters that are similar in all characteristics but cohort, sex and risk aversion. We expect differences in participation to be more closely related to differences in PI shocks than TS. Moreover, we expect to see patterns that have been uncovered in the existing literature across clusters with different characteristics.

We start with the pattern of stock market participation by age for three cohorts that are relatively more numerous. Figure 9 plots average stock market participation by cohort against both the TS and the PI shocks estimated using the KF. It appears that the pattern of participation responds to the latter while having a limited relationship with the former, consistent with economic intuition.

We then compare two clusters with similar characteristics but risk aversion (Figure 10). We find that the time patterns for PI shocks are relatively similar for the two groups (see intermediate panel), while the TS are not (see top panel). In response to these similar PI shocks, the cluster with higher risk aversion participates less in the equities market (see bottom panel), in line with economic intuition.

We perform a similar analysis across two identical clusters (including risk aversion) but for sex (Figure 11). We find that females participate less in the equities market (see bottom panel), a known result. It also appears that the volatility of PI shocks is greater for females than for males (see middle panel).

2 The Sensitivity of Optimal Life-Cycle Portfolios to Correlation Estimates

This Appendix shows the sensitivity of optimal portfolio age-profiles to correlation estimates, based on the calibration of a life-cycle model.

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

$$U_{t_0} = E_{t_0} \left[\frac{C_{t_0}^{1-\gamma}}{1-\gamma} + \sum_{j=1} \beta^j \left(\prod_{k=-1}^{j-2} p_{t_0+k} \right) \left(p_{t_0+j-1} \frac{C_{t_0+j}^{1-\gamma}}{1-\gamma} + (1-p_{t_0+j-1}) b \frac{(X_{t_0+j}/b)^{1-\gamma}}{1-\gamma} \right) \right] \quad (\text{A.4})$$

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

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

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

$$\log Y_{it} = \log \lambda + f(t_{0+K}, \mathbf{Z}_{it_{0+K}}) + u_{it_{0+K}} \quad t_0 + K < t \leq T \quad (\text{A.5})$$

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

We allow savings to be invested in a short-term riskless asset, yielding each period a constant gross real return R^f , and in a risky asset characterized as "stocks" yielding a stochastic gross real returns R_t^s . We maintain that investment opportunities do not vary over time, thus modeling excess stock

returns over the riskless asset as:

$$R_t^s - R^f = \mu^s + \nu_t^s \tag{A.6}$$

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

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

$$X_{it+1} = (X_{it} - C_{it})R_{it}^P + Y_{it+1} \tag{A.7}$$

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

$$R_{it}^P = \alpha_{it}^s R_t^s + (1 - \alpha_{it}^s) R^f \tag{A.8}$$

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

2.1 Solving the Life-Cycle Problem

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

$$\max_{\{C_{it}\}_{t_0}^{T-1}, \{\alpha_{it}^s, \alpha_{it}^b\}_{t_0}^{T-1}} \left(\frac{C_{it_0}^{1-\gamma}}{1-\gamma} + E_{t_0} \left[\sum_{j=1}^T \beta^j \left(\prod_{k=0}^{j-2} p_{t_0+k} \right) \left(p_{t_0+j} \frac{C_{it_0+j}^{1-\gamma}}{1-\gamma} + (1-p_{t_0+j}) b \frac{(X_{it_0+j}/b)^{1-\gamma}}{1-\gamma} \right) \right] \right) \quad (\text{A.9})$$

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

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

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

$$V_{it}(X_{it}, u_{it}) = \max_{C_{it}, \alpha_{it}^s} \left(\frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta E_t \left[p_t V_{it+1}(X_{it+1}, u_{it+1}) + (1-p_t) b \frac{(X_{it+1}/b)^{1-\gamma}}{1-\gamma} \right] \right) \quad (\text{A.10})$$

At each time t the value function V_{it} describes the maximized value of the problem as a function of the two state variables, the level of cash on hand at the beginning of time t , X_{it} , and the level of the stochastic permanent component of income at the beginning of time t , u_{it} . To reduce the dimensionality of the original problem to one state variable, we exploit the homogeneity of degree $(1-\gamma)$ of the utility function and normalize the entire problem using the permanent component of income u_{it} . Accordingly, we can rewrite (A.10) as:

$$V_{it}(X_{it}) = \max_{C_{it}, \alpha_{it}^s} \left(\frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta E_t [p_t V_{it+1}(X_{it+1}) + (1-p_t) b \frac{(X_{it+1}/b)^{1-\gamma}}{1-\gamma}] \right) \quad (\text{A.11})$$

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

2.2 Calibration and Results

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

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

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

Figure 12 reports the correlations between the PI shocks and stock market returns as estimated for the 80 clusters of individuals defined in the main text. The estimated individual correlations range from -0.9 to 0.9 , with most values concentrated in the range between 0.1 and 0.72 and the median value calculated as equal to 0.31 .

Using standard numerical techniques, we solve the model considering the parameter values presented in Table 2.2 and the estimated correlations. For the representative investors for each cluster, we obtain the optimal stock share over their working life. Figure 13 plots the optimal conditional stock share over the investor's working life where risk aversion is 5. The traditional glide path adopted by Target Date Funds is optimal when stock market returns and permanent income shocks are uncorrelated. When the correlation is zero, the optimal share invested in stocks until the age of 30 is 100%; it gradually decreases to reach 40% at retirement. In the event of correlation equal to -0.5 , the optimal share invested in stocks until the age of 40 is 100%; it then decreases more slowly, reaching 60% at retirement. In the event of correlation equal to 0.5 , the optimal share invested in stocks is 47% at the beginning of the investor's working life; it decreases to reach 25% at retirement.

Figure 14 plots the optimal conditional stock share over an investor's working life when risk aversion is 8. The traditional glide path adopted by Target Date Funds is optimal when stock market return and permanent income shocks are uncorrelated. When the correlation is zero the optimal share invested in stocks until the age of 35 is 100%; it gradually decreases to reach 20% at retirement. In the event of correlation equal to -0.5 , the optimal share invested in stocks until the age of 40 is 100%; it then decreases more slowly, reaching 40% at retirement. In the event of correlation equal to 0.5 , the optimal stock share is zero until the age of 28; it increases to reach 11% at retirement.

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

3 Individual Correlations between the PI Shocks and the Stock Market Returns

Throughout, we have documented that the individual exposure to the latent aggregate shocks shapes their hedging motives. This section repeats the analysis focusing on individual exposure to the stock market returns when the latter are imperfectly correlated with the aggregate shocks. That is, we estimate the correlations between the PI shocks and the stock market returns (instead of the ones with the aggregate risk factor) and substitute the old measures with these correlations in the Probit and Tobit analyses.

Section 3.1 adds to both the model and the moment conditions while the subsequent ones replicate the empirical analysis.

3.1 Model and Moment Conditions

We assume that the stock market return at time t is a linear combination of the aggregate factor and an idiosyncratic noise:

$$r_t = \sigma_r W_{r,t}^p,$$

where σ_r denotes the standard deviation of the stock market returns and

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

where $W_{r,t}$ is a standard Normal random variable. Therefore, we can express the stock market return as the sum of a systematic component, $\xi_{r,t}$, and an idiosyncratic component, $\omega_{r,t}$:

$$\begin{aligned}\xi_{r,t} &= \sigma_r (\rho_r W_t) \sim \mathcal{N}(0, \sigma_r^2 \rho_r^2), \\ \omega_{r,t} &= \sigma_r \left(\sqrt{1 - \rho_r^2} W_{r,t} \right) \sim \mathcal{N}(0, \sigma_r^2 (1 - \rho_r^2)),\end{aligned}$$

so that

$$r_t \sim \mathcal{N}(0, \sigma_r^2 \rho_r^2 + \sigma_r^2 (1 - \rho_r^2)), \quad \text{i.e.} \quad r_t \sim \mathcal{N}(0, \sigma_r^2). \quad (\text{A.12})$$

In equation (A.12), ρ_r and σ_r are unknown, so the vector θ including the set of unknown parameters extends to:

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

Meanwhile, the set of moment conditions that we can use to estimate the model parameters expands too, by including the co-movements between the individual income shocks and the stock market returns. In fact, the covariance between the time variation in the income shocks for each individual and the stock market returns is given by:

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

Note that in the first line we have used both the orthogonality of the transitory shocks to the stock market returns and independence between the stock market return at t and the permanent income

shocks at any point in 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, so the co-movement between each individual's income shocks and the stock market returns is only due to their common but heterogeneous exposure to the aggregate shocks. Then, by using the covariances between the time variation in the income shocks for the N individuals and the stock market returns, we obtain N moment conditions.

The last moment condition involves the variance of the stock market returns, which is equal to:

$$\text{var}(r_t) = \sigma_r^2.$$

As a result, the set of cross-sectional restrictions, including these additional $N + 1$ conditions, is now of size equal to:

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

We then implement the MD estimator, as in Section 3, to draw inference on the model parameters included in θ .

3.2 Empirical Results

We estimate the model using the DHS data between 1993 and 2019 using the same clusters presented in Section 4. In Figure 15, we show that the cluster-based correlation estimates in the baseline case are in line with those obtained with this specification of the stock market returns. We also report summary statistics of the correlation parameter estimates in Table 17.

Next, we relate the correlation between income shocks and stock market returns to both the decision

to participate to the stock market and asset allocation. In Table 18 and Table 19, we show that also the results presented in Tables 4, 5, 6, and 7 hold. More precisely, while the mean cluster-based correlation coefficient obtained is slightly lower (0.189 in Table 17 and opposed to 0.257 Table 3) the estimated coefficients are larger (-0.442 in Table 18 instead of -0.267 in Table 4 and -0.129 in Table 19 instead of -0.085 in Table 5). The economic impact on portfolio choice of hedging either an aggregate risk factor or stock market returns is therefore similar. Results are hardly affected, as expected, since the stock market returns and the aggregate risk factor are positively correlated.

3.3 Kalman Filter

With imperfect correlation between the stock market returns and the aggregate shocks, we can express the aggregate shocks as a linear function of the stock market returns by using equation (A.12):

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

and embed W_t in the equation that describes the dynamics of the PI shocks:

$$u_{i,t} = \sigma_u (\rho_i W_t) + \omega_{i,t},$$

that becomes:

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

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 form a prediction about the permanent component of the labour income, denoted by the random walk $v_{i,t}$, by computing the expectation conditional on the past value of the random walk and the contemporaneous stock market

return:

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

This specification nests the baseline case when $\rho_r=1$. As in Section 1, we pin down the dynamics of the permanent shocks to the labor income over the entire time series for each individual. Then, we compute the sample correlation between the individual PI shocks and the stock market returns. We track revisions in the correlation between PI shocks and stock market returns by sequentially computing the sample correlation between stock market returns and PI shocks obtained using the KF up to a given wave $t < T$, where T is the total number of available waves in our sample. Then, we use these sample correlations in a Probit analysis of stock market participation over time and we report results in Table 20. Results are similar to the ones obtained when we use estimates of KF-based correlations between PI shocks and the aggregate risk factor, allowing for imperfect correlation between the latter and the stock market returns. That is, revised correlations retain explanatory power for participation revisions provided that MD correlations with stock returns are omitted. Finally, in Table 21, we replicate and support the out-of-sample Probit and Tobit analysis of Table 11.

References

- Bonaparte, J., G. Korniotis, and A. Kumar, 2014, Income hedging and portfolio decisions, Journal of Financial Economics 113, 300–324.
- Campbell, J.Y., and L. Viceira, 2002, Strategic asset allocation: Portfolio choice for long-term investors, Oxford University Press, Oxford, UK. .

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.

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.

Figure 9. Income shocks and participation by cohort

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

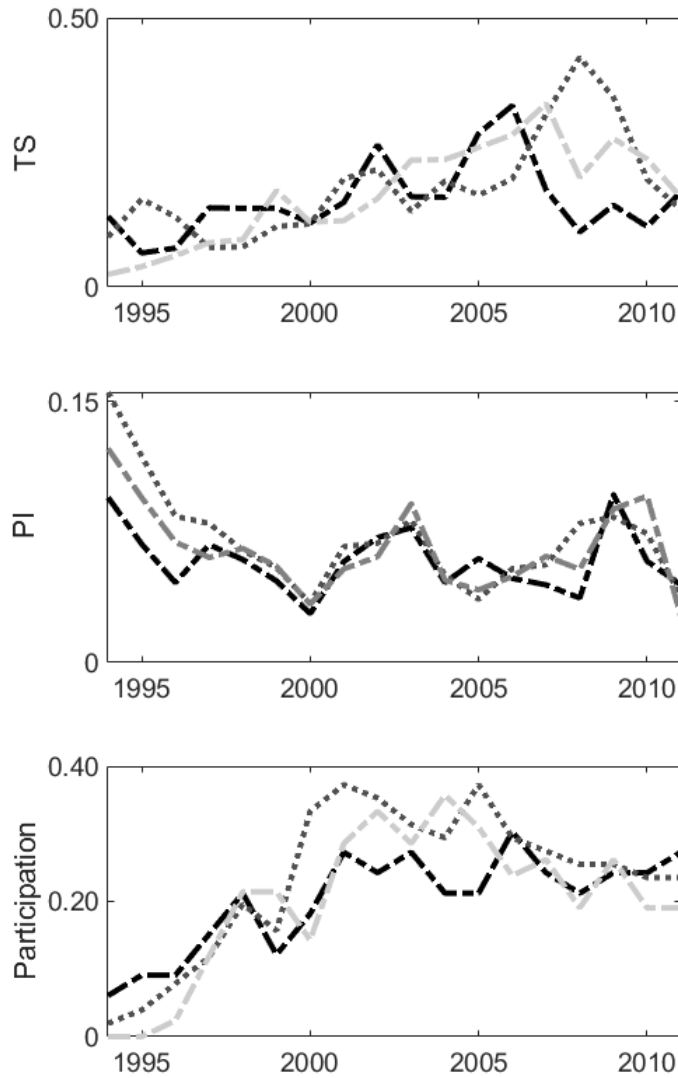


Figure 10. Income shocks and participation by cluster: Risk aversion

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

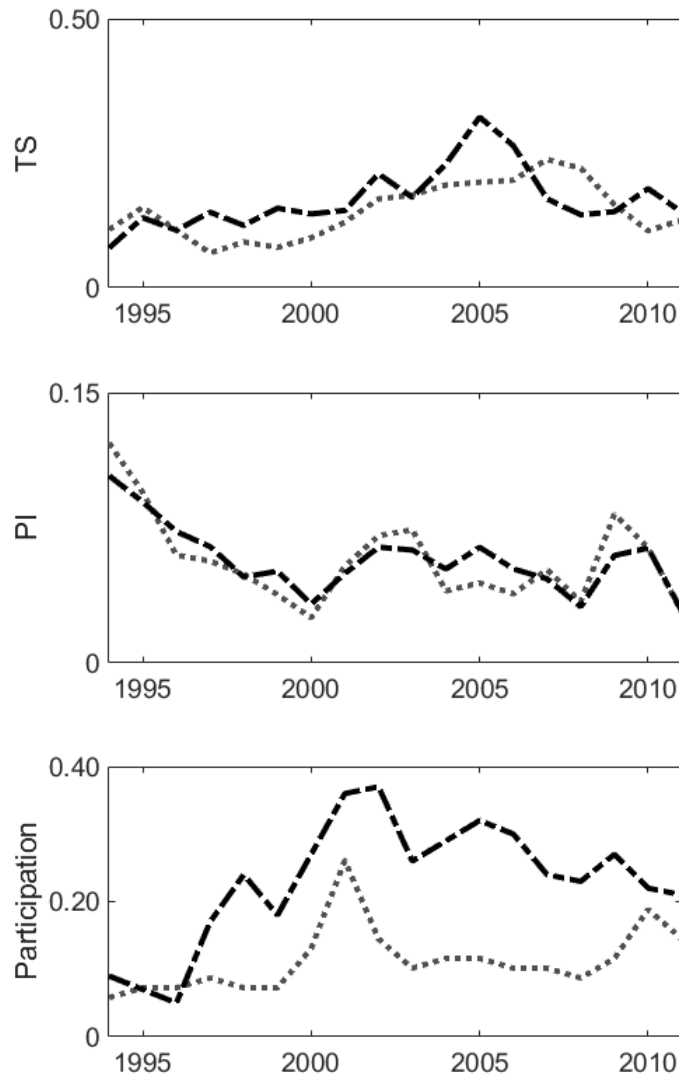


Figure 11. Income shocks and participation by cluster: Sex

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

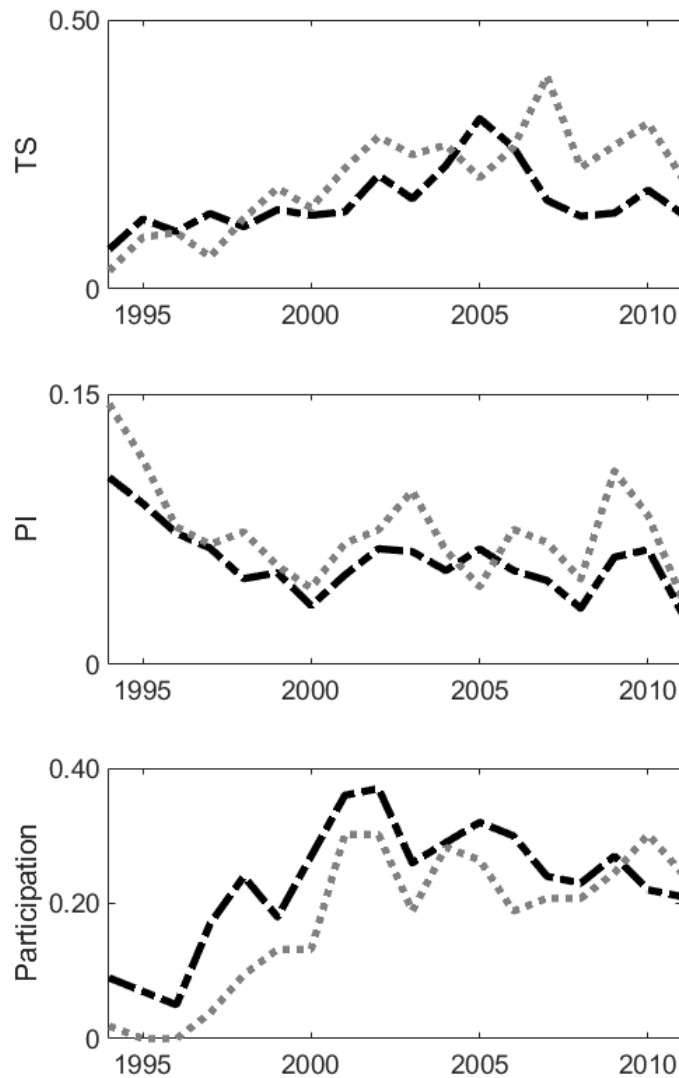


Figure 12. Distribution of cluster-based individual correlations

The figure reports the distribution of correlations between stock market returns and PI shocks as estimated after clustering the individuals in 80 homogeneous groups.

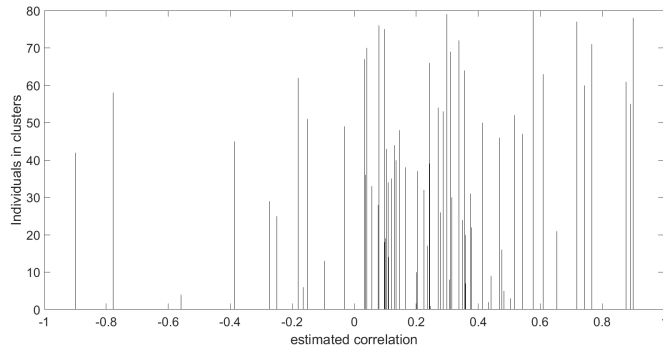


Figure 13. Optimal stock share: Risk aversion of 5 and correlation going from -0.9 to -0.1 (left panel) or from 0.1 to 0.9 (right panel)

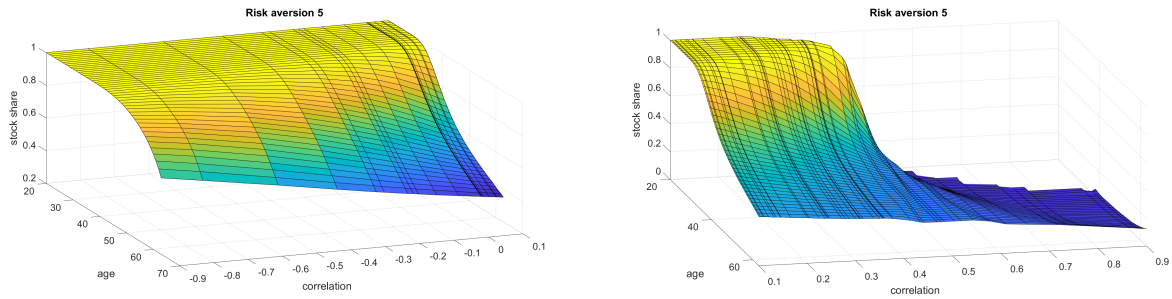


Figure 14. Optimal stock share: Risk aversion of 8 and correlation going from -0.9 to -0.1 (left panel) or from 0.1 to 0.9 (right panel)

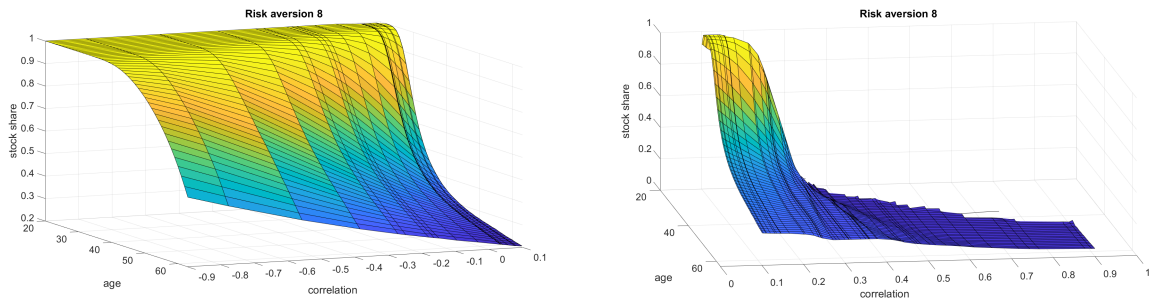


Figure 15. Cluster-based correlations: A comparison across model specifications

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 3 (y -axis). The data are from the DNB Household Survey and cover waves for the period 1993–2019.

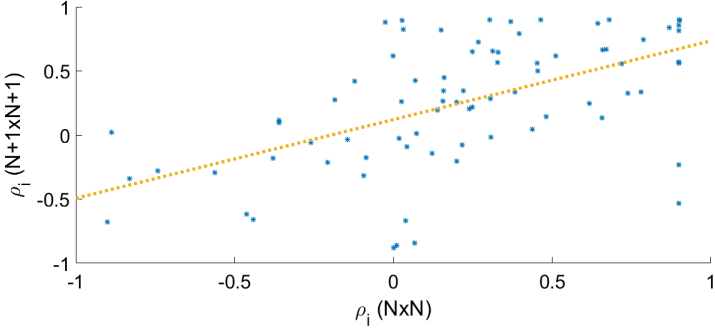


Table 15 Probit Estimates for Participation Revision

The table reports the Probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stock either directly or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation, the correlation between permanent income shocks and aggregate shocks, the correlation between stock market return and different specifications of labor income shocks as described in Table 4, and different measures of the relationship between labour income growth rate and stock market returns used in previous papers as described in Table 5. The correlation and the beta coefficients used here as independent variables are computed on an expanding time window going from the first wave up to the t -th wave, where t goes from 10 to the 19 (the last wave). We use the clustered correlation obtained with minimum distance estimation as ex-ante correlation parameter of the state-space model to reconstruct the PI shocks using Kalman filter at the individual level. Then, we compute the yearly change in the correlation between these PI shocks and the stock market returns, where the first change is computed with respect to the clustered correlation estimated using minimum distance (KFexp80). We include the same control variables as in Table 4, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

Independent Variable	OwnSTKMF					
	(1)	(2)	(3)	(4)	(5)	(6)
St. Dev.	-0.207*** (0.046)	-0.206*** (0.046)	-0.202** (0.043)	-0.203*** (0.057)	-0.187*** (0.060)	-0.187*** (0.042)
Corr(Tot,Rm)	-0.039 (0.032)					
Corr(Det,Rm)		0.064 (0.034)				
Corr(Stoc,Rm)		-0.067* (0.042)				
Corr(PermMD80)			-0.288*** (0.037)			
Corr(KFexp80,Rm)			-0.270** (0.131)			
Corr(tranBKK,Rm)				-0.031 (0.034)		
Corr(permBKK,Rm)				-0.047 (0.036)		
Beta (Guvenen et al. (2017))					0.068 (0.061)	
Beta (C&V(2002))						-0.080 (0.066)
Controls	YES	YES	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES	YES
N	9,345	9,299	10,423	7,586	10,423	10,423
Pseudo R ²	0.277	0.277	0.285	0.280	0.272	0.272

Table 16 Frequency of market participation: Poisson regression estimates

The table reports results from cross-sectional Poisson regressions. The dependent variable is the number of waves in which respondents reported investing in stock either directly or through mutual funds (FreqSTKMF). The main independent variables are the standard deviation, the correlation between permanent income shocks and aggregate shocks estimated by using the minimum distance methodology at the cluster-level (Corr(permMD80)), the correlation between stock market return and different specifications of labor income shocks as described in Table 4, and different measures of the relationship between labour income growth rate and stock market returns used in previous papers as described in Table 5. We use the clustered correlation obtained with minimum distance estimation as ex-ante correlation parameter of the state-space model to reconstruct the PI shocks using the Kalman filter at the individual level (KF80). We include the same control variables as in Table 4, but we suppress coefficients to save in space. We report in parentheses the robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

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

Table 17 Correlation Parameters from DHS. Summary Statistics on Full Sample.

The table reports 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 ($\text{Corr}(\text{PermMD80}, \text{Rm})$) when allowing for imperfect correlation between the aggregate shocks and the stock market returns, as well as the estimated correlations between stock market return and the permanent component obtained using the Kalman Filter both when computing the correlation at T ($\text{Corr}(\text{KF80}, \text{Rm})$) and when updating the correlation at each t ($\text{Corr}(\text{KFexp80}, \text{Rm})$). In the last column, we report the t-test for the statistical significance of the parameter and ***, **, * denote statistical significance at the 0.1%, 1%, and 5% significance levels. The data are from the DNB Household Survey and cover waves from 1993–2019.

Correlation	Mean	St. Dev.	p10	Median	p90	T-test
$\text{Corr}(\text{PermMD80}, \text{Rm})$	0.189	0.234	-0.121	0.195	0.472	35.211***
$\text{Corr}(\text{KF80}, \text{Rm})$	0.351	0.514	-0.442	0.479	0.925	29.706***
$\text{Corr}(\text{KFexp80}, \text{Rm})$	0.379	0.566	-0.543	0.574	0.956	23.776***

Table 18 Probit Estimates for Stock Market Participation

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

OwnSTKMF					
	(1)	(2)	(3)	(4)	(5)
St. Dev.	-0.141*** (0.017)	-0.144*** (0.038)	-0.499*** (0.118)	-0.144*** (0.038)	-0.144*** (0.038)
Corr(PermMD80,Rm)	-0.571*** (0.039)	-0.442*** (0.063)	-0.359*** (0.134)	-0.442*** (0.068)	-0.428*** (0.064)
Corr(tranBKK,Rm)			0.086 (0.114)		
Corr(permBKK,Rm)			-0.172 (0.105)		
Beta (Guvenen et al. (2017))				0.002 (0.092)	
Beta (C&V(2002))					-0.155** (0.087)
Controls	NO	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES
N	27,445	12,758	3,286	12,758	12,758
Pseudo R ²	0.025	0.275	0.322	0.276	0.276

Table 19 Tobit Estimates for Asset Allocation

The table reports the Tobit regression results for asset allocation decision. The dependent variable is the portfolio shares in stocks either directly or through mutual funds (propSTKMF). The main independent variables are the standard deviation of the labour income growth rate, the correlation between the permanent component of labor income shocks and the stock market returns estimated by using the MD methodology at the cluster-level (Corr(permMD80,Rm)) when allowing for imperfect correlation between the aggregate shocks and the stock market returns, as well as different measures of the relationship between labour income growth rate and stock market returns used in previous papers: the correlation between stock market returns and the transitory and the permanent shocks to labour income estimated as in Bonaparte et al. (2014) (tranBKK and permBKK, respectively), the beta coefficients of the regression of the labour income growth rate over the stock market returns as in Guvenen et al. (2017) and Campbell and Viceira (2002), estimated at the cluster-level. We include the same control variables as in Table 4, but we suppress coefficients to save in space. We report in parentheses the Tobit-robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

PropSTKMF					
	(1)	(2)	(3)	(4)	(5)
St. Dev.	-0.092*** (0.012)	-0.030* (0.014)	-0.182* (0.118)	-0.030* (0.014)	-0.030* (0.014)
Corr(PermMD80,Rm)	-0.389*** (0.023)	-0.129*** (0.024)	-0.119*** (0.042)	-0.124*** (0.027)	-0.123*** (0.025)
Corr(tranBKK,Rm)			0.077* (0.036)		
Corr(permBKK,Rm)			-0.084* (0.034)		
Beta (Guvenen et al. (2017))				0.017 (0.036)	
Beta (C&V(2002))					-0.065* (0.034)
Controls	NO	YES	YES	YES	YES
Year Dummy	YES	YES	YES	YES	YES
N	27,275	12,498	3,224	12,498	12,498
Pseudo R ²	0.020	0.291	0.368	0.291	0.291

Table 20 Probit Estimates for Participation Revision

The table reports the Probit regression results for stock market participation decision. The dependent variable is a dummy variable for respondents who reported to own stock either directly or through mutual funds (OwnSTKMF). The main independent variables are the standard deviation, the correlation between the permanent component of labor income shocks and the stock market returns estimated by using the MD methodology at the cluster-level ($\text{Corr}(\text{permMD80}, \text{Rm})$) when allowing for imperfect correlation between the aggregate shocks and the stock market returns. Then, we use this clustered correlation obtained with MD estimation as ex-ante correlation parameter of the state-space model to reconstruct the PI shocks using Kalman filter at the individual level and we compute the sample correlation between these PI shocks and the stock market return (KF80). In addition, we also compute the yearly change in the correlation between these PI shocks and the stock market returns, where the first change is computed with respect to the clustered correlation estimated using minimum distance (KFexp80). We include the same control variables as in Table 4, but we suppress coefficients to save in space. We report in parentheses the Probit-robust standard errors and ***, **, * over the regression coefficients denote statistical significance at the 0.1%, 1%, and 5% significance levels, respectively. The data are from the DNB Household Survey and cover waves from 1993–2019.

OwnSTKMF			
Independent Variable	(1)	(2)	(3)
St. Dev.	-0.146*** (0.038)	-0.182*** (0.042)	-0.191*** (0.043)
Corr(PermMD80,Rm)			-0.443*** (0.072)
Corr(KF80,Rm)	-0.056* (0.022)		
Corr(KFexp80,Rm)		-0.272** (0.117)	-0.142 (0.119)
Controls	YES	YES	YES
Year Dummy	YES	YES	YES
N	12,758	10,423	10,423
Pseudo R ²	0.272	0.279	0.285

Table 21 Out-Of-Sample Probit Estimates

The table reports the out-of-sample Probit regression results for stock market participation decision for the waves 2012-2019. The dependent variable is a dummy variable for respondents who reported to own stock either directly or through mutual funds (OwnSTKMF), own stock only (OwnSTK) or mutual funds only (OwnMF), between 2012 and 2019. The main independent variable is the correlation between the PI shocks and the stock market returns estimated at the cluster-level using the MD methodology (Corr(permMD80,Rm)) and data up to 2011, when allowing for imperfect correlation between the aggregate shocks and the stock market returns. This correlation is then assigned to each individual on the basis of the corresponding cluster to which the individual is allocated according to her personal characteristics observed in 2012. We also include other independent variables, and additional control variables as defined in Table 4. N is the number of observations. The data are from the DNB Household Survey and cover waves from 1993–2019.

	OwnSTKMF	OwnSTK	OwnMF
	(1)	(2)	(3)
St. Dev. (dy)	-0.244*** (0.072)	-0.162* (0.090)	-0.217** (0.075)
Corr(PermMD80,Rm)	-0.302*** (0.064)	-0.009 (0.082)	-0.325*** (0.066)
Controls	YES	YES	YES
Year Dummy	YES	YES	YES
N	4,336	4,336	4,336
Pseudo R ²	0.327	0.273	0.269