What Do the Portfolios of Individual Investors Reveal About the Cross-Section of Equity Returns?

Sebastien Betermier, Laurent E. Calvet, Samuli Knüpfer, and Jens Kvaerner*

February 24, 2022

Abstract

We extract a parsimonious set of equity factors from comprehensive administrative data on the stockholdings of all Norwegian individual investors in 1996-2017. A three-factor model, featuring the market portfolio and long-short portfolios of stocks sorted by investor age or wealth, explains both the common variation in portfolio holdings and the cross-section of stock returns. Portfolio tilts toward investor factors correlate with indebtedness, macroeconomic exposure, gender, education, and investment experience. Our results are consistent with hedging and sentiment jointly driving portfolio decisions and the cross-section of equity premia.

JEL Codes: G11, G12.

Keywords: Asset pricing, factor-based investing, household finance, portfolio allocation.

*Betermier: Desautels Faculty of Management, McGill University, 1001 Sherbrooke St West, Montreal, QC H3A 1G5, Canada; sebastien.betermier@mcgill.ca. Calvet: EDHEC Business School, 16 rue du Quatre-Septembre, 75002 Paris, France, and CEPR; laurent.calvet@edhec.edu. Knüpfer, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway; samuli.knupfer@bi.no. Kvaerner: Tilburg University, Warandelaan 2 5037 AB Tilburg, Netherlands; jkverner@gmail.com. We thank David Brown, David Chapman, Serhiy Kozak, Hughes Langlois, Jiacui Li, and Lars Lochstoer for helpful discussions, and we acknowledge useful comments from Fabio Braggion, Joost Driessen Paul Ehling, Evan Jo, Thore Johnsen, Patrick Konermann, Richard Priestley, Espen Sirnes, and seminar participants at Arizona State, BI Norwegian Business School, EDHEC, McGill University, Tilburg University, the CEPR Sixth European Workshop on Household Finance, the 2021 Northern Finance Association, the 2021 Red Rock Finance Conference, the 2021 Academic Research Colloquium for Financial Planning and Related Disciplines, the 2021 European Winter Meeting of the Econometric Society, the 15th International Conference on Computational and Financial Econometrics, and the 2022 Annual Meeting of the American Finance Association. We gratefully acknowledge financial support from Finansmarkedsfondet at Research Council of Norway under Project No. 309855.

Conflict-of-Interest Disclosure

Sebastien Betermier

I have nothing to disclose.

Laurent E. Calvet

I have nothing to disclose.

Samuli Knüpfer

I have nothing to disclose.

Jens Kvaerner

I have nothing to disclose.

Investor demand for stocks plays a central role in leading explanations of equity risk premia (Campbell, 2018; Merton, 1973; Lintner, 1966; Sharpe, 1964). In equilibrium, portfolio diversification, hedging needs, or sentiment drive the portfolio choices of investors and the compensation they earn for holding risky securities (Barberis et al. 2015; Campbell et al. 2018; Kozak, Nagel, and Santosh 2018). Perhaps surprisingly, however, empirical work on the cross-section of returns rarely uses direct observations on investor portfolios. Most empirically successful pricing factors rely instead on firm characteristics (Fama and French, 2015; Hou et al., 2015), and a large literature attempts to tie the firm-based factors to aggregate consumption.[1]

Breaking away from these traditional approaches, a new literature has recently started to exploit the holdings of large U.S. institutional investors to study their impact on asset market movements, volatility, and predictability (Koijen and Yogo, 2019, 2020; Koijen et al., 2020b). The use of large holdings data has permitted researchers to estimate demand systems in the style of the industrial organization literature, which has renewed interest in the role of investor demand in equity markets. Until now, however, this emerging body of research has not considered the rich holdings of individual investors, nor has it developed pricing factors tied to the socioeconomic characteristics of stock owners.

In this paper, we propose that the portfolios of individual investors represent a promising avenue for empirically understanding the cross-section of stock returns. This strategy follows from financial theory and is also motivated by a set of empirical findings from the household finance literature. The portfolios of individual investors are closely linked to demographics (Campbell, 2006; Gomes et al., 2021; Guiso and Sodini, 2013) and exhibit a strong factor structure (Balasubramaniam et al., 2021). These empirical regularities indicate that individual portfolios are revealing about investor preferences and beliefs, and may be informative about risk premia. Moreover, within the financial portfolio, direct stockholdings are likely to be even more informative than indirect holdings because they are not clouded by the frictions due to delegated asset management (He and Xiong, 2013). Betermier, Calvet, and Sodini (2017) accordingly show that the links between demographics and portfolio exposures to the value factor are most pronounced in direct stockholdings.

Despite the appeal of relating individual investor portfolios to asset returns, a major challenge is that available datasets of stock holdings often lack dimensions that are crucial for performing rigorous asset pricing tests, such as a long time series, a large and diverse pool of investors, and detailed investor characteristics. For example, the well-known Barber

[1]See Constantinides (2017), Ludvigson (2013), Mehra (2012) and the references therein.

and Odean (2000, 2001) dataset includes five years of transactions by retail investors trading through a particular discount broker. By comparison, studies of the cross-section of equity returns frequently use at least twenty years of data.[2]

Our paper overcomes this challenge by using comprehensive administrative data on all Norwegian individual investors who directly own stocks in 1996-2017. For each investor, the panel contains socioeconomic characteristics and disaggregated holdings at the stock level. This complete ownership record covers more than 400 stocks listed on the Oslo Stock Exchange (OSE) and is remarkable for its large cross-section of investor portfolios (about 365,000 investors a year) and long time series (22 years).

Leveraging this unique dataset, we investigate what the stock holdings of individual investors reveal about the factors that price the cross-section of equity returns. Our analysis aims to answer four questions. If one sorts stocks by the characteristics of the individual investors who own them, do these characteristics produce factors that price the cross-section of stock returns? Which investor characteristics should matter in theory and which ones matter in the data? How do investor factors compare with traditional factors constructed from firm characteristics? Last but not least, how do stockholders' socioeconomic characteristics, risks, and biases relate to portfolio tilts toward the investor pricing factors?

To answer these questions, we begin the analysis by deriving theoretical results that aid the development of the empirical methodology. We study the theoretical link between the cross-section of investor portfolios and the cross-section of stock returns. For a given factor structure of investor portfolios, market clearing implies that the returns on the portfolio factors generate pricing factors that explain the cross-section of equity returns. Crucially for our approach, we derive conditions under which pricing factors can be recovered from a sufficiently heterogeneous set of investor portfolios. These results only depend on the portfolio factor structure and market clearing, and do not require us to specify investor preferences, beliefs, or biases.

We next use financial theory to endogenize the portfolio factor structure. We demonstrate that two investor characteristics, age and wealth, are likely to drive the cross-section of investor portfolios and therefore the cross-section of equity returns. We derive this result

[2] As Merton (1980) explains, the high level of volatility in stock returns makes statistical inference on average returns challenging in small samples. As a back-of-the-envelope calculation, consider an asset whose abnormal return has a sample monthly average of 1% and a volatility of 4%. Given 5 years of monthly data, the probability of correctly rejecting the null hypothesis that the average abnormal return (alpha) is 0% is only equal to 47%. Given 20 years of monthly data, the probability goes up to 97%.

in two complementary settings: an overlapping generations version of the ICAPM (Merton, 1973) that combines time-varying investment opportunities and labor income risk, and a model with sentiment in the spirit of Fedyk et al. (2013) and Sandroni (2000). These models predict that the portfolios of mature and high-wealth investors should be closer to the mean-variance efficient frontier than the portfolios of young and low-wealth investors. These results hold irrespective of the details of the model, such as the number of variables driving common hedging needs or sentiment. Mature and wealthy investors should therefore earn higher CAPM alphas than young and less wealthy investors.

We assess the empirical validity of these predictions on the Norwegian data set. Consistent with the recent findings of Balasubramaniam et al. (2021), we uncover a strong factor structure in the portfolio holdings of individual investors. Commonalities across investor portfolios are clearly visible when we aggregate portfolio holdings according to investor age, wealth, and other socioeconomic characteristics. A principal component analysis (PCA) of these portfolios reveals that three principal components explain 85% of the cross-section of investor portfolios. Motivated by theory, we also define a three-factor model of investor portfolios that contains the market portfolio and long-short portfolios of investors sorted by either age or wealth. Our three-factor model explains 73% of the cross-section of investor portfolios. It therefore substantially improves on the 28% explained by the market portfolio alone and accounts for the bulk of the 85% explained by the top three principal components of investor portfolios.

We next turn to the cross-section of stock returns and assess if equity premia can be explained by the returns on age, wealth, and market portfolios. We consider two methods for the construction of these factors, which provide consistent results. One approach, which we follow in the Appendix, is to use the returns on the age and wealth portfolio factors previously defined. Another and perhaps more standard approach, which we follow in the main text, consists of forming long-short portfolios of stocks sorted by characteristics. For the first time, these characteristics are measured from each stock's individual investor base. Specifically, a stock's age characteristic is the average age of its individual owners in a particular year, weighted by the number of shares that they hold at the beginning of the year. Similarly, the wealth characteristic is the average net worth of the stock's individual owners, where net worth is defined as the value of financial and non-financial assets net of liabilities. The age and wealth characteristics display substantial heterogeneity across stocks and over time. We define an investor factor as a portfolio that is long stocks in the top 30% of the stock's investor characteristic and short stocks in the bottom 30%.

Consistent with our theoretical predictions, the age and wealth factors generate average returns that are strictly positive and economically significant. Their monthly CAPM alphas are 1.08% (t-value of 2.7) and 1.01% (t-value of 2.9), respectively, which correspond to yearly alphas of about 12%. An equal-weighted portfolio of the age and wealth factors has an even stronger t-value of 4.15, along with an alpha of 1.05% per month and an annual Sharpe ratio of 0.68 over the sample period. Moreover, 62% of this alpha is not explained by the most commonly used firm-based factors: market, size, value, investment, profitability, and momentum. These results confirm that the stocks held by more mature and wealthier investors deliver significantly higher abnormal returns than the stocks owned by other investors.

We investigate if investor characteristics other than age and wealth also contain relevant pricing information. We find that the age and wealth factors explain the return on all long-short portfolios constructed from other stockholders' characteristics, including gender, occupations, and education. The age and wealth factors thus appear to capture reasonably well all the pricing information contained in investor portfolio holdings.

Our three-factor investor-based model is a strong performer out of sample. We demonstrate this property by implementing out-of-sample bootstrap tests similar to Fama and French (2018). We randomly select in-sample periods, construct tangency portfolios from the factors' in-sample return moments, and evaluate performance in the remaining out-of-sample-periods. Our investor-based model produces an average out-of-sample Sharpe ratio of 0.66 in annual units. As a comparison, the Sharpe ratio of the Norwegian market is 0.32. Moreover, the 0.66 out-of-sample Sharpe ratio of our three-factor model exceeds the 0.50 average Sharpe ratio of the six-factor model containing the market, size, value, investment, profitability, and momentum factors. Using fixed factor weights, as advocated by DeMiguel et al. (2009), delivers similar results.

We next study investor exposures to the age and wealth factors over the life-cycle and across the wealth distribution. To avoid any mechanical correlations, we partition our sample of investors into two randomly chosen groups. We define the age and wealth factors using one group and measure the factor tilts of investors in the other group. The factor tilts of investors in the second group vary with age and wealth as one expects. The results hold even among investors in their first year of direct stock market participation, which indicates that the migration in portfolio tilts is not primarily driven by factor loadings of firms that evolve over time. Instead, investors progressively adjust their stockholdings and therefore their factor tilts over the life cycle.

To understand the drivers of this life cycle migration, we regress the age and wealth factor tilts on a set of investor characteristics. We find that the effects of age and wealth on the factor tilts are robust to the inclusion of controls. Investors with high income beta to GDP growth and high debt-to-income ratio also tilt away from these factors, which is consistent with hedging demands. Additionally, investors prone to sentiment, such as men or investors with little stock market experience, no business education, or no professional experience in finance, also tilt their portfolio away from the age and wealth factors. Echoing the recent survey results in Choi and Robertson (2020) and Giglio et al. (2021), our results suggest that hedging and sentiment jointly drive factor tilts. Our findings are also in line with Kozak, Nagel, and Santosh (2018), who show that hedging and sentiment channels can generate pricing factors that are observationally equivalent to each other.

We gain additional insights into the nature of investor factor tilts by analyzing the characteristics of firms that make up the age and wealth factors of stock returns. Relative to other investors, mature and wealthy investors tend to hold stocks with large market capitalizations, high book-to-market ratios, high profitability, low investment, and low CAPM betas. These tilts are similar to those of U.S. institutions reported in Koijen and Yogo (2019). We also document large differences in firm characteristics that prior literature typically associates with sentiment (Baker and Wurgler, 2006; Stambaugh and Yuan, 2017). Young and less wealthy investors are more likely to hold volatile stocks with high share turnover and low institutional ownership. These are the stocks about which investors disagree the most and in which arbitrage can be limited.

Our paper contributes to the extensive literature on the cross-section of equity premia (Cochrane, 2011; Harvey, Liu, and Zhu, 2015) by extracting rich pricing information from the stock holdings of individual investors. Constructing equity factors from investor portfolios allows researchers to tie equity pricing directly to investor risks, preferences, and biases. By contrast, firm characteristics may be more informative about firm production decisions that micro-found production-based asset pricing models. Thus, both types of factors can provide useful complementary information about the sources of equity premia, as we show in the data. In fact, both categories of factors are expected to theoretically price the cross-section of stock returns since asset prices are determined by both investor and firm characteristics in general equilibrium (see, e.g., Betermier, Calvet, and Jo, 2022).

Our results are particularly relevant for the growing research on the interaction of investor portfolio holdings and asset prices. The contributions of Koijen and Yogo (2019) and Koijen, Richmond, and Yogo (2020b) identify the types of institutional investors that have the strongest price impact in equity markets. Other studies use institutional holdings to examine the allocation of interest rate risk (Hoffmann et al., 2018), currency risk (Maggiori et al., 2020), and the transmission of monetary policy (Carpenter et al., 2015; Koijen et al., 2020a). Additional work in household finance shows that retail investors also impact stock prices even though they only own a limited fraction of aggregate equity (Blume and Keim, 2012; Barber et al., 2009; Kaniel et al., 2008; Kelley and Tetlock, 2013). In contrast to these contributions, the present paper does not study the price impact of trades but shows instead that the stock portfolios of individual investors contain useful information about their risk preferences and the cross-section of equity premia, consistent with a revealed preference approach (Barber et al., 2016; Berk and van Binsbergen, 2016; Guercio and Tkac, 2002).

Our paper also builds on recent advances in the household finance literature. Household portfolios are known to exhibit a high degree of heterogeneity, which has motivated a wealth of empirical and theoretical explanations. Building on the work of Betermier, Calvet, and Sodini (2017) who uncover linkages between a firm's value loading and the socioeconomic characteristics of its investors, we show how to construct pricing factors directly from the holdings of individual investors and evaluate the performance of these factors. Our analysis also relates to the contemporaneous paper by Balasubramaniam et al. (2021) which documents a factor structure in the stock portfolios of Indian investors. We document new facts about the factor structure using a broad set of investor characteristics and show the factor structure of holdings has important implications for equity pricing.[3]

Finally, we contribute to the literature at the intersection of household finance and macroeconomics documenting how heterogeneity in household portfolio returns impacts wealth inequality. Our result that wealthy investors earn higher average returns in equity markets is consistent with the findings in Bach, Calvet, and Sodini (2020) and Fagereng et al. (2020). We show that the return differential can be explained by heterogeneous exposures to a common pricing factor, which is informative about the sources of differences in performance across investors.

The rest of the paper unfolds as follows. Section I develops the theoretical framework linking the cross-section of investor portfolios to equity factors. Section II presents the data and empirical evidence on the factor structure in portfolio holdings. Section III constructs investor pricing factors and assesses their ability to price the cross-section of stock returns. Section IV studies the drivers of investor portfolio tilts toward the new factors. Section V

[3]Using data on different U.S. institutional types, Büchner (2020) also finds evidence of commonality in investor demand.

concludes. An online appendix provides proofs and additional empirical results.

I. Theoretical Linkages Between Investor Portfolios and Pricing Factors

This section investigates the strong theoretical links between investor portfolios and equity risk premia and provides guidance on the empirical methodology. In Section I.A, we develop a mapping between the cross-section of investor portfolios and the cross-section of stock returns under rather general conditions. Section I.B shows how to recover pricing factors from a heterogeneous set of investor portfolios. In Section I.C, we demonstrate that for hedging or behavioral reasons, investor age and wealth are likely to be key drivers of portfolio heterogeneity and equity premia.

A. *Linking Pricing Factors to Aggregate Portfolio Tilts*

We consider a financial market in which investors $i \in \{1, \cdots, I\}$ can trade a risk-free asset and stocks $j \in \{1, \cdots, J\}$. We focus on the equilibrium of security markets at a particular point in time and do not use a time subscript in this section for convenience. Let R_f denote the risk-free rate, $\boldsymbol{R^e}$ the J-dimensional column vector of excess stock returns, and $\mathbf{1}$ the J-dimensional column vector with components equal to unity. We also consider the vector of expected stock returns, $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{R^e}) + R_f\,\mathbf{1}$, and the variance-covariance matrix of stock returns, $\boldsymbol{\Sigma}$.

The market portfolio $\boldsymbol{m}$ is the portfolio of the J stocks weighted by market capitalization. The tangency portfolio

$$\boldsymbol{\tau} = \frac{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\,\mathbf{1})}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\,\mathbf{1})} \tag{1}$$

is the portfolio of stocks with the highest Sharpe ratio. The market portfolio and the tangency portfolio have expected returns μ_m and μ_τ and volatilities σ_m and σ_τ, respectively.

Each investor i invests the wealth E^i in stocks. The vector of weights in her equity portfolio is given by $\boldsymbol{\omega^i} \in \mathbb{R}^J$, where $\mathbf{1}'\boldsymbol{\omega^i} = 1$. The investor can also invest in the riskless asset. Building on the empirical findings of Balasubramaniam et al. (2021), we assume that the cross-section of investor portfolios has the factor structure:

$$\boldsymbol{\omega^i} = \boldsymbol{\tau} + \sum_{k=1}^{K} \eta_k^i \, \boldsymbol{d_k} + \boldsymbol{u^i}, \tag{2}$$

where $\boldsymbol{d_k}$ denotes the k^{th} *portfolio factor*, η_k^i is investor i's loading on the k^{th} factor, and $\boldsymbol{u^i}$ is an idiosyncratic tilt.

The factors $\boldsymbol{d_k}$ capture the common directions along which investor portfolios deviate from the tangency portfolio in the portfolio space $\mathbb{R}^J$. For this reason, we also refer to each $\boldsymbol{d_k}$ as a *deviation portfolio*. As we explain in Section I.C, deviation portfolios can originate from hedging or sentiment motives. The idiosyncratic portfolio $\boldsymbol{u^i}$ represents deviations from the tangency portfolio that are unrelated to deviation portfolios. These tilts may stem from preferences, beliefs, or forms of inertia that are specific to each investor. To guarantee the additivity condition $\boldsymbol{1'\omega^i} = 1$, we assume that $\boldsymbol{d_k}$ and $\boldsymbol{u^i}$ are zero-investment portfolios, so that their components add up to zero $\boldsymbol{1'd_k} = 0$ and $\boldsymbol{1'u^i} = 0$ for every k and i. In addition, the idiosyncratic tilts of investors cancel out in the aggregate: $\sum_{i=1}^{I} E^i \, \boldsymbol{u^i} = \boldsymbol{0}$.

Market clearing imposes that the aggregate portfolio of investors coincides with the market portfolio of stocks: $\sum_{i=1}^{I} E^i \boldsymbol{\omega^i} / \sum_{i=1}^{I} E^i = \boldsymbol{m}$. The aggregation of individual stock portfolios (2) implies that

$$\boldsymbol{m} = \boldsymbol{\tau} + \sum_{k=1}^{K} \eta_k^m \, \boldsymbol{d_k}, \tag{3}$$

where $\eta_k^m = \sum_{i=1}^{I} E^i \eta_k^i / \sum_{i=1}^{I} E_i$ is the aggregate loading on the k^{th} deviation portfolio. We assume without loss of generality that $\eta_k^m \geq 0$ for every k.[4]

The factor structure of individual portfolios in the portfolio space $\mathbb{R}^J$ has strong implications for the cross-section of stocks in the space of returns. Let $f_0 = \boldsymbol{m'R^e}$ denote the excess return on the market portfolio and let $f_k = \boldsymbol{d_k'} \boldsymbol{R^e}$ denote the return on the k^{th} deviation portfolio. Market clearing and equations (1) and (3) imply that the vector of factor returns

$$\boldsymbol{f} = (f_0, \ldots, f_K)'$$

prices the cross-section of stock returns, as we show in the Appendix.

[4]Otherwise we replace $\boldsymbol{d_k}$ by -$\boldsymbol{d_k}$ and η_k^i by -η_k^i for every investor i in equation (2).

Proposition 1. *The average excess return on every stock j satisfies*

$$\mu_j - r_f = \boldsymbol{\beta_j'}\,\mathbb{E}(\boldsymbol{f}) \tag{4}$$

where $\boldsymbol{\beta_j}$ is the vector of linear regression coefficients of stock j's return on the factors.

Equation (3) and Proposition 1 show a direct connection between priced factors and aggregate tilts. If the aggregate tilts η_k^m are all equal to zero, the market portfolio is mean-variance efficient and the CAPM holds. By contrast, if $\eta_k^m > 0$ for some $k \in \{1, \dots, K\}$, the k^{th} return factor f_k is also priced. This result is a direct consequence of market clearing and therefore holds whether the deviation portfolio $\boldsymbol{d_k}$ is risk-based or sentiment-based.[5]

Stocks with high exposures to deviation portfolios generate negative CAPM alphas, as we now explain. Let $b_{j,m}$ denote stock j's univariate beta to the market portfolio, and let $\alpha_j = \mu_j - r_f - b_{j,m}(\mu_m - r_f)$ denote its CAPM alpha. In the Appendix, we show that

$$\alpha_j = -\phi \sum_{k=1}^{K} \eta_k^m \, \sigma_k^2 \, (b_{j,k} - b_{j,m}\, b_{m,k}), \tag{5}$$

where $\phi = (\mu_\tau - r_f)/\sigma_\tau^2$ is a positive constant, σ_k is the volatility of the k^{th} deviation portfolio, $b_{j,k}$ is the univariate beta of stock j relative to the k^{th} deviation portfolio, and $b_{m,k}$ is the univariate beta of the market relative to the k^{th} deviation portfolio. The coefficient $(b_{j,k} - b_{j,m}\, b_{m,k})$ measures the stock's exposure to the k^{th} deviation portfolio, $b_{j,k}$, net of the stock's deviation exposure resulting from its loading on the market, $b_{j,m}\, b_{m,k}$. If this difference is positive, the stock earns *negative* alpha. The stock is in high demand so it trades at a premium relative to the CAPM.

In addition to having a negative alpha, a stock with high exposures to deviation portfolios tends to have a high market beta. In the Appendix, we show that a stock's market beta is a weighted average of its beta to the tangency portfolio, $b_{j,\tau}$, and its betas to deviation portfolios, $b_{j,k}$:

$$b_{j,m} = \frac{\sigma_\tau^2}{\sigma_m^2} b_{j,\tau} + \sum_{k=1}^{K} \eta_k^m \, \frac{\sigma_k^2}{\sigma_m^2} \, b_{j,k}. \tag{6}$$

Since a stock highly exposed to deviation portfolios is in high demand, it represents a large

[5] Fama and French (2007) obtain a similar result in a simple framework that relates asset prices to disagreement and tastes.

share of the market portfolio and therefore has a high beta. In Section III, we verify that these predictions on alpha and beta hold empirically.

B. Extracting Pricing Factors from Portfolio Data

Proposition 1 provides a roadmap for constructing pricing factors from a cross-section of investor portfolios. Consider a group $\mathcal{G}$ of investors and a set of investor weights $\{z_1^i\}_{i\in\mathcal{G}}$ where $\sum_{i\in\mathcal{G}} z_1^i = 0$. We construct a zero-investment portfolio of stocks as follows:[6]

$$\boldsymbol{g_1} = \sum_{i\in\mathcal{G}} z_1^i\, \boldsymbol{\omega^i}. \tag{7}$$

The portfolio $\boldsymbol{g_1}$ has several appealing properties. By (2), its loading on the tangency portfolio is zero. Moreover, assuming that it provides sufficient diversification so that $\sum_{i\in\mathcal{G}} z_1^i\, \boldsymbol{u^i} \approx 0$, $\boldsymbol{g_1}$ can be expressed as a linear combination of the deviation portfolios:

$$\boldsymbol{g_1} = \sum_{k=1}^{K} \left(\sum_{i\in\mathcal{G}} z_1^i\, \eta_k^i \right) \boldsymbol{d_k}. \tag{8}$$

If investor portfolios are sufficiently heterogeneous, we can construct K linearly independent portfolios $\boldsymbol{g_1}, \ldots, \boldsymbol{g_K}$ from different sets of investor weights. By (8), these portfolios fully span the deviation portfolios $\boldsymbol{d_1}, \ldots, \boldsymbol{d_K}$.[7] Consequently, the returns on the market portfolio $\boldsymbol{m}$ and the portfolios $\boldsymbol{g_1}, \ldots, \boldsymbol{g_K}$ price the cross-section of stocks. Hence, pricing factors can be obtained by constructing a well-chosen set of long-short portfolios derived from individual portfolios.

We make several observations about the implementation of this empirical strategy. To construct the pricing portfolios $\boldsymbol{g_1}, \ldots, \boldsymbol{g_K}$ from investor portfolio data, it is not necessary to include every investor. Nor is it necessary to use a representative subset of investors. As long as the holdings are sufficiently heterogeneous and provide sufficient diversification of idiosyncratic tilts, the proposed empirical strategy will deliver the pricing factors. This point suggests that the direct portfolio holdings of individual investors may contain valuable information about equity factors even when these investors only own a modest fraction of

[6] The property that $\boldsymbol{g_1}$ is a zero-investment portfolio follows from the fact that $\boldsymbol{1}'\boldsymbol{g_1} = \sum_{i\in G} z_1^i\, \boldsymbol{1}'\boldsymbol{\omega^i} = \sum_{i\in\mathcal{G}} z_1^i = 0$.

[7] The linear subspace generated by $\boldsymbol{g_1}, \ldots, \boldsymbol{g_K}$ coincides with the linear subspace generated by the deviation portfolios, or more compactly $\text{Span}[\boldsymbol{g_1}, \ldots, \boldsymbol{g_K}] = \text{Span}[\boldsymbol{d_1}, \ldots, \boldsymbol{d_K}]$.

the aggregate market capitalization.

Importantly, we note that the model does not predict that every portfolio $\boldsymbol{g_k}$ should carry a non-zero risk premium. As Proposition 1 explains, a deviation portfolio $\boldsymbol{d_k}$ is priced if and only if the aggregate tilt η_k^m is not zero. For example, assume that employees in a given occupational sector tend to hold similar stocks. Their deviation portfolios do not generate a positive risk premium if employees from other sectors have the opposite tilt. With this caveat, we will verify that some pricing portfolios are able to generate non-zero risk premia in the data.

In practice, the investor weights z_1^i can be chosen as a function of observable investor characteristics that are likely to be correlated to aggregate portfolio tilts. For example, if investor age drives deviations from the tangency portfolio, one would assign a positive weight to all investors above a given age threshold and a negative weight to all investors below this threshold. The resulting portfolio $\boldsymbol{g_1}$ is long the portfolios of mature investors and short the portfolios of young investors. We now show that two investor characteristics, age and wealth, are prime candidates for constructing investor-based equity factors.

C. Main Directions of Investor Portfolio Heterogeneity

We develop two complementary models of portfolio choice providing guidance on the socioeconomic characteristics that are likely to produce investor factors. We first derive a standard rational ICAPM model in the style of Merton (1973) and Breeden (1979) populated by investors with heterogeneous ages and income profiles. Second, we consider a model with sentiment in the spirit of Fedyk et al. (2013) and Sandroni (2000). The models endogenize the factor structure of portfolio holdings and connect them to investor characteristics.

Case 1: Hedging. We consider an overlapping generations economy populated by investors indexed by i. Time is discrete.[8] Every period, investors can invest in a short-term bond with risk-free rate R_f and in stocks with excess returns $R^e_{1,t+1}, ..., R^e_{J,t+1}$. The conditional distribution of the return vector $(R_f, R^e_{1,t+1}, ..., R^e_{J,t+1})$ at date t is driven by a state vector $\boldsymbol{y_t}$ that follows a first-order Markov process. In applications, the state vector $\boldsymbol{y_t}$ may for instance follow a vector autoregression. Consistent with the original ICAPM (Merton, 1973), the distribution of asset returns and the state vector $\boldsymbol{y}_t$ are exogenous to the model.

[8]The Appendix develops a continuous-time version of the model.

An investor i is born in period b^i and lives until period $b^i + T$. She receives an initial endowment $W^i_{b_i}$ and labor income $L^i_{b^i}$ in period $t = b^i$. In all subsequent periods, the investor receives the non-financial income L^i_t, which grows at the stochastic rate $g_{L,t+1} = L^i_{t+1}/L^i_t$. We assume for simplicity that the income growth rates $\{g_{L,t+1}\}$ are common to all investors, independent through time, and do not depend on past realizations of labor income. Aggregate income growth is correlated to stock returns and is therefore a source of hedging demand, along with time variation in investment opportunities captured by $\boldsymbol{y_t}$.

In every period t, the investor selects the risky share s^i_t, the portfolio of stocks $\boldsymbol{\omega^i_t}$ and the consumption level C^i_t that maximize expected utility $\mathbb{E}_{b^i}\left[\sum_{t=b^i}^{b^i+T} \delta^{t-1} u(C_t)\right]$ subject to the budget constraint

$$W^i_{t+1} = L^i_t\, g_{L,t+1} + (W^i_t - C^i_t)\left(1 + R_f + s^i_t \sum_{j=1}^{J} \omega^i_{j,t} R^e_{j,t+1}\right), \tag{9}$$

where $\sum_{j=1}^{J} \omega^i_{j,t} = 1$. The value function $J(t, W^i_t, L^i_t, \boldsymbol{y_t})$ satisfies the Bellman equation

$$J(t, W^i_t, L^i_t, \boldsymbol{y_t}) = \max_{\{s^i_t, \boldsymbol{\omega^i_t}, C^i_t\}} \left[u(C^i_t) + \delta\, \mathbb{E}_t J(t+1, W^i_{t+1}, L^i_{t+1}, \boldsymbol{y}_{t+1})\right] \tag{10}$$

subject to the budget constraint (9). The optimal portfolio of stocks, $\boldsymbol{\omega^i_t}$, is a function of age, wealth, and labor income:

$$\boldsymbol{\omega^i_t} = \boldsymbol{\tau_t} + \boldsymbol{d}(A^i_t, W^i_t, L^i_t, \boldsymbol{y_t}), \tag{11}$$

where $A^i_t = t - b^i$ denote the investor's age at date t. In the Appendix, we derive the relation between $\boldsymbol{\omega^i_t}$ and the value function. The deviation portfolio is zero for an investor in the last investment period ($A^i_t = T - 1$) with zero labor income ($L_{T-1} = 0$).

If the utility function is CRRA, $u(C) = C^{1-\gamma}/(1-\gamma)$, the deviation portfolio can be directly expressed in terms of the income-to-wealth ratio: $\boldsymbol{d}(A^i_t, L^i_t/W^i_t, \boldsymbol{y_t})$. We apply a Taylor expansion to the deviation portfolio around the last investment period ($A^i_t = T - 1$) and a labor income-to-wealth ratio equal to 0, and obtain the portfolio factor structure:

$$\boldsymbol{d}(A^i_t, W^i_t, L^i_t, \boldsymbol{y_t}) = (T - 1 - A^i_t)\, \boldsymbol{d_{1,t}} + \frac{L^i_t}{W^i_t} \boldsymbol{d_{2,t}}, \tag{12}$$

where $\boldsymbol{d_{1,t}}$ and $\boldsymbol{d_{2,t}}$ are deviation portfolios. The investor's time horizon and income-to-wealth ratio drive the magnitude of portfolio deviations from the tangency portfolio. The model predicts that the portfolios of mature and wealthy investors should be closer to the tangency portfolio and therefore earn higher CAPM alphas than the portfolios of younger and less wealthy investors.

This example illustrates that an ICAPM model with heterogeneous investors naturally generates a factor structure of investor portfolios. Furthermore, the dimensionality of the factor structure (11)-(12) is solely driven by the dimensionality of the investor characteristics that drive portfolio choice. Quite strikingly, the rank of the factor structure does not depend on the dimension of the state vector $\boldsymbol{y_t}$.

Case 2: Sentiment. Deviations of investor portfolios from mean-variance efficiency can also originate from sentiment. Investors may choose inefficient stock portfolios because they overreact to recent returns (Barberis et al., 2015). They may also adjust their portfolios to forms of public information that do not impact the composition of the tangency portfolio, or they may over- or under-estimate the impact of these data on the tangency portfolio.

While the literature on sentiment is extensive (see Hirshleifer (2015) for a survey), many studies emphasize that the strength of sentiment co-varies with two key variables: age and wealth. Age is generally associated with a reduction in the size of inefficiencies. Young investors tend to be prone to fads and invest in bubbly stocks (Greenwood and Nagel, 2009). As they age, they accumulate experience on the outcomes of past decisions, learn from past mistakes, and end up making more efficient decisions (Seru, Shumway, and Stoffman, 2010). The impact of age is also a natural consequence of Bayesian learning (Barberis, 2000; Ehling, Graniero, and Heyerdahl-Larsen, 2018; Skoulakis, 2008).

Wealth is also positively correlated with more efficient behavior (Vissing-Jorgensen, 2003). Sentiment drives portfolio allocation and therefore wealth accumulation. Over the longer run, investors with more rational expectations are likely to be wealthier than other households (Sandroni, 2000). This effect is especially strong in general equilibrium in the presence of multiple assets, as Fedyk, Heyerdahl-Larsen, and Walden (2013) show. There is also empirical evidence that wealthier investors hold financial portfolios with higher Sharpe ratios (Calvet, Campbell, and Sodini, 2007).

These considerations motivate the following reduced-form model: $\boldsymbol{\omega_t^i} = \boldsymbol{\tau_t} + \boldsymbol{d_t^i}$, where the deviation is given by

$$\boldsymbol{d_t^i} = \boldsymbol{d}(A_t^i, W_t^i, \boldsymbol{\xi_t})$$

and $\boldsymbol{\xi_t}$ is the common information set driving portfolios. If more mature investors with large amounts of wealth converge to the tangency portfolio, a simple linearization implies that

$$\boldsymbol{d_t^i} = (T - 1 - A_t^i)\,\boldsymbol{d}_{1,t} + \frac{1}{W_t^i}\,\boldsymbol{d}_{2,t} \tag{13}$$

in every period t.[9] Since investor age and wealth capture sentiment, constructing long-short portfolios according to these characteristics will allow us to recover the factor structure.

To sum up our theoretical discussion, investor factors can price stock returns and can be recovered from large and diverse datasets of investor holdings. The selected group of investors does not need to include every single stock market investor as long as the dispersion in holdings is informative about the drivers of portfolio tilts. Investor age and wealth are two prime candidates for constructing investor-based equity factors because they capture a combination of hedging and behavioral effects. In the next Section, we apply this factor extraction methodology to a high-quality dataset of Norwegian retail investors.

II. Data and Factor Structure of Portfolio Holdings

A. Data

We use several data sources on Norway's stock market. The Oslo Stock Exchange (OSE) is Norway's only regulated market for securities trading. We retrieve the complete record of stock ownership in the OSE over the 1996 to 2017 period from the Norwegian Central Securities Depository (VPS). For each security listed on the exchange, we observe the anonymized personal identification number of its owners and the number of shares that each owner holds at the end of each year. Individual investors are classified in the VPS database as investors with a non-professional investor account. On average, 365,000 individual investors directly hold OSE-listed stocks each year. A stock has a median number of 1,560 individual investors.

We obtain the demographic and financial characteristics of individual investors from Statistics Norway (SSB). The financial information is collected by the Norwegian Tax Administration and includes a complete breakdown of individuals' balance sheets. This information is collected annually for tax purposes, which means that banks and other third parties are legally required to provide this information to the Tax Administration. Using the

[9]Specifically, consider the function $\boldsymbol{d^*}(\tau, v, \boldsymbol{\xi_t}) \equiv \boldsymbol{d}(T - 1 - \tau, 1/v, \boldsymbol{\xi_t})$. The linearization of d^* around $(0, 0, \boldsymbol{\xi_t})$ implies (13), where $\boldsymbol{d}_{1,t} = \partial \boldsymbol{d^*}/\partial\tau(0, 0, \boldsymbol{\xi_t})$ and $\boldsymbol{d}_{2,t} = \partial \boldsymbol{d^*}/\partial v(0, 0, \boldsymbol{\xi_t})$.

personal identification numbers, we merge the SSB data with the stock ownership data to track the owners and their socioeconomic characteristics for each stock listed in Norway. We restrict the sample to investors who file a tax return and are at least 18 years old, the minimum age required to open a personal trading account, and have a minimum liquid financial wealth of 10,000 Norwegian kroner (NOK) at the end of the calendar year. For international comparison, 1 Norwegian krone traded at 0.122 U.S. dollar (USD) on December 29, 2017. We occasionally convert nominal amounts into USD at this fixed rate of conversion.

Monthly ticker prices, market capitalizations, and information about all corporate events are available from the OSE for our sample period. We complement this information with accounting data from the Norwegian School of Economics (NHH) for 1996-2011 and Thomson Reuters Worldscope (TRW) for 2012-2018.[10] We use stock data up to the end of 2018 because we wish to track the performance of portfolios held at the end of 2017 over the following year. The NHH data set has broader coverage than TRW at the beginning of the sample. TRW reports the fraction of free-floating shares (item NOSHFF) from 1997 onward. Free-float adjusted market values ensure that our sample is not dominated by a few large companies predominantly controlled by the Norwegian government.[11]

The stock universe considered in our analysis consists of OSE-listed stocks that satisfy the following requirements at the end of June of each year. Following the common practice in the literature, we require stocks to have at least 12 months of return history, non-missing common equity as of December 31 of the previous year, and a share price above 1 NOK in the month of portfolio formation. Our universe includes 484 unique stocks over the sample period with an average of 178 firms per year, which is typical for a European stock market. To ensure that our results are not driven by outliers, we winsorize all monthly returns at the 99.9% level.[12] The market portfolio is the value-weighted portfolio of all stocks in our universe.

[10]Link: https://www.nhh.no/en/library/databases/

[11]The government owns a substantial fraction of a few large companies: Equinor ASA (67%, energy), Norsk Hydro (34%, energy), Telenor (54%, telecommunications), DnB (34%, banking), Entra (22.4%, real estate), Yara (36%, chemicals) and Kongsberg gruppen (50%, technology). Data on government ownership is available at: https://www.regjeringen.no/no/tema/naringsliv/statlig-eierskap/selskaper---ny/id2604524/?expand=factbox2607470. The mandate of the Norwegian sovereign wealth fund precludes it from investing in domestic companies.

[12]As a result, all winsorized stock returns are less than 154% per month.

B. *Factor Structure of Investor Portfolios*

We document the factor structure exhibited by the portfolios of individual investors in Norway. To eliminate the idiosyncratic tilts present in the holdings of individuals, we construct well-diversified portfolios obtained from large investor groups $\mathcal{I}_1, \ldots, \mathcal{I}_G$. Equation (2) predicts that for each group $\mathcal{I}_g$, the weight of stock j in the group's aggregate portfolio, $\omega^g_{j,t} = \sum_{i \in \mathcal{I}_g} \omega^i_{j,t} E^i_t / \sum_{i \in \mathcal{I}_g} E^i_t$, is the linear combination of the portfolio factors:

$$\omega^g_{j,t} = \tau_{j,t} + \sum_{k=1}^{K} \eta^g_{k,t} \, d_{k,j,t}, \tag{14}$$

where $\eta^g_{k,t} = \sum_{i \in \mathcal{I}_g} \eta^i_{k,t} E^i_t / \sum_{i \in \mathcal{I}_g} E^i_t$ is the group's loading on the deviation portfolio $\boldsymbol{d_k}$.

We form 93 investor groups each year from a broad and diverse set of investor characteristics. We require that each group contains at least 10,000 investors in a year to ensure sufficient diversification. The details of these groups, which are provided in the Appendix, can be summarized as follows.

Motivated by the theoretical discussion in Section I.C, we form 10 groups of investors sorted by age and 12 groups of investors sorted by net worth each year. The first age group includes all investors below 30, the next eight groups are set in five-year increments, and the last age group includes all investors above 70. We define net worth as the sum of the investor's liquid financial wealth, real estate, vehicles, and business assets, net of liabilities.[13] The net worth groups consist of the first 9 deciles of the net worth distribution (groups 1-9), the 90^{th}-99^{th} percentiles (group 10), the 99^{th}-99.9^{th} percentiles (group 11), and the 99.9^{th}-100^{th} percentiles (group 12).[14]

For each year in the sample, we also form 12 groups of investors sorted by permanent real income,[15] two groups of investors sorted by gender, three groups based on educational

[13] Non-traded assets include private dwellings, holiday houses, boats, vehicles, forestland, farmland, and other real capital, machinery and equipment, house contents and movables, and real assets held abroad. Liquid financial wealth includes stocks, mutual funds, money market funds, and bank account balances. Financial assets are evaluated at market prices. Other assets are evaluated by using assessed tax values for the 1997-2009 period and estimated market values from 2010 onward.

[14] The net worth distribution is based on the entire Norwegian population that are 18-100 years in a given year and have at least 10,000 NOK of liquid financial wealth. It is therefore not limited to the sample of investors with an investment account. Similar filters have been advocated by for example Fagereng et al. (2017) in their analysis of portfolio choice in Norway. We relax the minimum requirement of 10,000 investors for the top wealth group in order to capture the investment preferences of high-wealth investors.

[15] The percentiles are defined in a similar way as those of the wealth groups. Permanent income at year t is calculated as the average real earnings over the 5-year period between years $t-6$ to $t-2$. The details of

attainment (high school or lower, bachelor's degree, master's degree or higher), nine groups based on the field of study (such as management or science), 19 groups based on occupation, and 26 groups based on the region of residence. Because the coverage of occupational information is of lower quality prior to 2002, we use 2002-2017 for this analysis.

We assess the number of common factors present in the 93 group portfolios by conducting a principal component analysis. For each year t, we denote by J_t the number of stocks in the year and by $\boldsymbol{\omega_t^g}$ the $J_t \times 1$ column vector of portfolio weights of group g. We also consider the $J_t \times G$ matrix of demeaned portfolios, $\boldsymbol{\Omega_t} = (\boldsymbol{\omega_t^1} - J_t^{-1}\,\mathbf{1}, \ldots, \boldsymbol{\omega_t^G} - J_t^{-1}\,\mathbf{1})$. We apply an eigenvalue decomposition of the covariance matrix $\boldsymbol{\Omega_t'}\,\boldsymbol{\Omega_t}$ and obtain:

$$\boldsymbol{\Omega_t} = \boldsymbol{F_t}\,\boldsymbol{\Lambda_t}, \tag{15}$$

where $\boldsymbol{F_t} = (f_{k,j,t})_{1 \leq j \leq J_t,\, 1 \leq k \leq G}$ is a $J_t \times G$ matrix of principal components and $\boldsymbol{\Lambda_t} = (\lambda_{k,t}^g)_{1 \leq k,g \leq G}$ is $G \times G$ matrix of loadings. By equation (15), the share of stock j in the portfolio of group g at time t satisfies:

$$\omega_{j,t}^g = \frac{1}{J_t} + \sum_{k=1}^{G} \lambda_{k,t}^g\, f_{k,j,t}. \tag{16}$$

This equation is an empirical analogue to (14), with the important difference that the empirical specification (16) does not contain the unobserved tangency portfolio and is therefore empirically feasible.[16]

Figure 1 reports the average cumulative proportion of the cross-sectional variance of aggregate portfolio holdings explained by the principal components (PCs), where the average is calculated across 2002-2017. The figure reveals a strong factor structure: a small number of PCs is sufficient to explain the cross-section of investor portfolios. The first PC explains 74% of the average cross-sectional variation, whereas adding two more PCs explains 85%.

the calculation are given in the Appendix.

[16] Our principal component analysis is based on the eigenvectors of the $G \times G$ variance-covariance matrix of portfolio weights, $\boldsymbol{\Omega}_t'\,\boldsymbol{\Omega}_t$, which, for each pair of groups, contains the covariance of the groups' portfolio weights across stocks. The resulting factors are linear combinations of group portfolios, consistent with the theoretical construction of Section I. We therefore depart from the PCA approach of Balasubramaniam et al. (2021), who consider the $J_t \times J_t$ variance-covariance matrix $\boldsymbol{\Omega}_t\,\boldsymbol{\Omega}_t'$ containing, for each pair of stocks, the covariance of the stocks' portfolio weights across investors. In our context, our methodology offers the key benefits of offering tighter links to theory and of requiring the diagonalization of a variance-covariance matrix of lower dimension. Connor and Korajczyk (1986, 1988) similarly discuss the benefits of using different PCA approaches in the context of pricing factors in return space.

[Insert Figure 1 here]

C. Linking the Factor Structure to the Market, Age, and Wealth Portfolios

We next examine the extent to which a three-factor model consisting of the market portfolio and the age and wealth portfolio tilts captures the common factors in investor portfolios.

Building on the factor construction approach given by (7), we construct two zero-weight portfolios of investors sorted by age and wealth. The first portfolio, denoted by $\boldsymbol{g}_{\mathbf{AGE},\boldsymbol{t}}$, takes a long position in the portfolios of investors who are 60 and older (age groups 8, 9, and 10, with equal weights assigned to each group) and a short position in the portfolios of the youngest investors age 30 and lower (age group 1). The second portfolio, denoted by $\boldsymbol{g}_{\mathbf{WEALTH},\boldsymbol{t}}$, takes a long position in the portfolios of 10% wealthiest investors (wealth groups 10, 11, and 12 equally-weighted) and a short position in the portfolios of the 30% least wealthy (wealth groups 1, 2, and 3 equally-weighted).

For each PC portfolio k, we run a pooled OLS regression of the weight of stock j in the k^{th} PC portfolio, $f_{k,j,t}$, on stock j's weight in the market, age, and wealth portfolios:

$$f_{k,j,t} = a^k + \lambda^k_{\text{MKT}}\, m_{j,t} + \lambda^k_{\text{AGE}}\, g_{\text{AGE},j,t} + \lambda^k_{\text{WEALTH}}\, g_{\text{WEALTH},j,t} + \epsilon^k_{j,t}. \tag{17}$$

This regression is estimated over the full sample period.

Table I reports the three-factor model's ability to explain the cross-section of portfolio holdings for several specifications of (17). In the top row, we report the average proportion of the cross-sectional variance of the 93 portfolio holdings explained by each PC from Figure 1. In the next set of rows, we report the *cumulative* proportion of the cross-sectional variance explained by (i) each PC, (ii) the projection of the PC on the market portfolio, which corresponds to imposing $\lambda^k_{\text{AGE}} = \lambda^k_{\text{WEALTH}} = 0$ in equation (17), and (iii) the projection of the PC on the market, age, and wealth portfolios as given by the unconstrained version of (17).[17]

[Insert Table I here]

The results in Table I confirm that the market portfolio alone does not accurately sum-

[17] For each specification of equation (17), we compute for each PC the product of (i) the proportion of the cross-sectional variance in group portfolios explained by the PC, and (ii) the R^2 coefficient of the pooled regression (17). We then aggregate these products across PCs to obtain the cumulative proportion of the cross-sectional variance of group portfolios explained by the specification.

marize the portfolio holdings of individual investors. The market portfolio on its own only explains 28% of the total variation in portfolio holdings through the top 10 PCs.

However, including the age and wealth portfolios into the pooled regression (17) strongly improves goodness of fit. Together with the market, these portfolios explain 71% of the total variation in group holdings through the top 3 PCs. The remaining variation explained by these factors is small – only 2% once we account for PCs 4 to 10. In total, the market, age, and wealth portfolios explain 73% of the total variation in portfolio holdings. In the Appendix, we obtain similar results when we estimate (17) for each of the 93 portfolios instead of the PC portfolios. The evidence altogether shows that the market portfolio, the age portfolio, and the wealth portfolio successfully capture common variation in portfolio holdings.

III. Investor Pricing Factors

Section III.A constructs an asset pricing model based on the market, an investor age factor, and an investor wealth factor. Section III.B investigates the return properties of investor factors. Section III.C evaluates the ability of our three-factor model to price traditional firm-based factors. Section III.D studies whether our model misses out on important pricing information contained in investor holdings. In Section III.E, we compare the out-of-sample performance of investor-based and firm-based pricing models by implementing bootstrap simulations in the style of Fama and French (2018).

A. Construction of Investor Factors

We construct the age and wealth factors by following the sorting methodology commonly used in the asset pricing literature. In a first step, we sort stocks by a characteristic, such as the average age or the average net worth of each stock's individual investor base. In a second step, the pricing factor is defined as the return on a long-short portfolio of sorted stocks.

In the Appendix, we alternatively follow the portfolio construction outlined in Section I.B. While the two approaches deliver consistent results, the method used in the main text offers several benefits. The preferred method has high statistical power because it focuses on stocks in top and bottom quantiles of the distribution of the sorting characteristic. Furthermore, it allows us to use identical construction methods for investor-based and firm-based factors, which improves the comparability of the results.

Age and Wealth Characteristics. The age characteristic of stock j in year t is the weighted average age of individual investors who own the stock:

$$\text{Age}_{j,t} = \frac{\sum_{i=1}^{I} N_{j,t}^{i} A_{t}^{i}}{\sum_{i=1}^{I} N_{j,t}^{i}}, \tag{18}$$

where A_t^i denotes the age of investor i and $N_{j,t}^i$ the number of shares of stock j held by the investor at the end of year t. Under this definition, the age of each investor i is weighted by her share of the firm's equity held by retail investors, $N_{j,t}^i / \sum_{i'} N_{j,t}^{i'}$.

Figure 2 illustrates the evolution of the age characteristic for two well-known companies listed on the OSE from 1997 to 2017: the global aluminium company Norsk Hydro (blue curve), and the wireless device company Nordic Semiconductor (blue curve). The age characteristic of Norsk Hydro increases from 62 years in 1997 to 67 years in 2017. By comparison, the age characteristic of Nordic Semiconductor is 55 in 1997, 48 in 2004, and 56 in 2017. More generally, the data reveal rich cross-sectional and time-series variation in the age characteristic of firms.

[Insert Figure 2 here]

The distribution of investor net worth is fat-tailed and positively skewed, so that a few high net worth investors can heavily influence wealth-weighted averages. To mitigate the impact of outliers, our measure of a stock's wealth characteristic is based on *brackets* of investors' net worth instead of net worth itself. We use the 12 wealth brackets defined in Section II.B and denote by $\text{WB}_t^i \in \{1, \dots, 12\}$ the bracket of investor i at date t. The stock's wealth characteristic,

$$\text{Wealth}_{j,t} = \frac{\sum_{i=1} N_{j,t}^{i} \text{WB}_{t}^{i}}{\sum_{i=1} N_{j,t}^{i}}, \tag{19}$$

is the weighted average of its individual investors' wealth bracket.[18]

[18]In a study of the low-risk anomaly, Bali et al. (2020) use detailed Swedish data and construct a measure of a stock's rich ownership as the proportion of the stock's shares outstanding that are directly held by individual investors in the top 10% of the wealth distribution. One important difference here is that our measure $\text{Wealth}_{j,t}$ is strictly based on the wealth of investors who directly hold the stock. We do not consider the ownership share of institutional and foreign investors in the calculation. Our wealth characteristic thus allows us to compare the demand for stocks by the high and low wealth investors, irrespective of the stocks' aggregate share that is directly held by individual investors.

Table II reports summary statistics on Norwegian investors in 2017. The average investor is 55 years old and holds a net worth of 6 million NOK (or 670,000 USD). The cross-sectional standard deviation of wealth is 48 million NOK (or 5 million USD), and the wealth bracket WB_t^i defined on a 1-to-12 scale has a standard deviation of 3.

The table also reports summary statistics on stocks. A stock's investor age, $\text{Age}_{j,t}$, has a cross-sectional standard deviation of 7.5 years. The firm's wealth characteristic, $\text{Wealth}_{j,t}$, has a standard deviation of 2 on the 1-12 scale. These estimates confirm that the ownership base is strongly heterogeneous across stocks.

[Insert Table II here]

Pricing Factors. The market factor, MKT_t, is the excess return on the market portfolio of Norwegian stocks, as defined in Section II.A.

We form investor pricing factors as follows. For each year t and each characteristic $C_{j,t} \in \{\text{Age}_{j,t}, \text{Wealth}_{j,t}\}$, we sort stocks by $C_{j,t}$ and group them into: (i) the low portfolio L containing stocks below the 30^{th} percentile, (ii) the middle portfolio M containing stocks between the 30^{th} and the 70^{th} percentiles, and (iii) the high portfolio H containing stocks above the 70^{th} percentile. Each portfolio is value-weighted by the stocks' free-float market value. We define the resulting pricing factor as the return on a zero-investment portfolio that is long H and short L. Under this method, the age factor, AGE_t, is the return on a mature-minus-young portfolio, while the wealth factor, $WEALTH_t$, is the return on a high wealth-minus-low wealth portfolio.

We use the exact same method to construct benchmark equity factors based on firm characteristics, which we henceforth call *firm factors*. Following Fama and French (1992, 1993, 2015), Hou et al. (2018), Carhart (1997), and Novy-Marx (2013), we sort stocks by market capitalization to form the size factor (SMB_t), by book-to-market ratio to form the value factor (HML_t), by profit margin to form the profitability factor (RMW_t), by investments to form the investment factor (CMA_t), and by the stocks' geometric return over the previous 12 months to form the momentum factor (MOM_t).[19] The long-short portfolios use the 30th and 70th percentiles as breakpoints. Following tradition, the size portfolio departs from this rule and uses the median of the size distribution as single breakpoint.[20]

[19]The most recent month is left out.

[20]The 30^{th} and 70^{th} percentiles ensure that the factors are well diversified. The details of the factor construction are provided in the Appendix.

These five firm factors are sensible benchmarks for our analysis because they are known to price with reasonable precision the cross-section of stock returns around the world (Fama and French, 2012; Griffin, Ji, and Martin, 2003). Moreover, firm factors are based on standard accounting and stock price information that is available for almost all stocks in our database.

B. Return Properties of Investor Factors

Table III, Panel A, reports the average excess returns on portfolios of stocks sorted by age or wealth over the 1997-2018 period. Average portfolio returns increase with the age and wealth characteristics. As a result, the average monthly return on investor factors is large and statistically significant: 0.98% (t-value = 2.37) for age and 0.92% (t-value = 2.59) for wealth. These monthly values correspond to average returns of 12.42% and 11.62%, respectively, in annual units. By comparison, the average monthly excess return on the market portfolio is 0.56% (t-value= 1.51) and the monthly return on firm factors ranges from -0.13% (t-value= -0.52) for the size factor to 0.85% (t-value= 2.34) for the profitability factor over the same sample period, as we report in the Appendix.

[Insert Table III here]

In Panel B of Table III, we show that the average returns on investor factors are not explained by their exposures to market portfolio risk. CAPM regressions of the age and wealth factors on the market over the sample period yield monthly intercepts that are significantly positive and equal to 1.08 (t-value = 2.65) for the age factor and 1.01% (t-value = 2.91) for the wealth factor. The age and wealth factors thus deliver significant and positive abnormal returns relative to the CAPM.

Panel C of Table III shows that investor factors, which exhibit positive alphas, have small but significantly negative market betas. Furthermore, market beta and average return are negatively related across portfolios of stocks sorted by the characteristics of their individual investors, as Panels A and C of Table III reveal.

These findings are in line with the theoretical analysis in Section I. Young and less wealthy investors tilt their portfolios toward stocks that provide hedging benefits or are attractive to sentiment-prone investors. In equilibrium, these attractive stocks generate negative alphas and represent a large share of the market portfolio, which in turn lead to high market betas.[21]

[21]Equations (5) and (6) summarize this logic. These equations predict that a positive tilt toward the

By contrast, more mature and affluent investors tilt away from these stocks, thereby holding portfolios with positive alphas and low betas.

In Figure 3, we plot the cumulative log growth of 1 NOK invested in 1997 in either the long and short legs of the age and wealth factor portfolios. We use the market portfolio as the benchmark. Economic recessions are shaded in blue. Panel A shows that the long leg of the age and wealth factors outperformed the market throughout the sample. By contrast, Panel B of Figure 3 shows that the short legs of the age and wealth factors performed well in the late 1990s but underperformed the market over the full sample. Underperformance is most pronounced after the 2008 crisis.

[Insert Figure 3 here]

Panel C of Figure 3 illustrates the cumulative performance of the age and wealth factors. Both factors have a high average return and a low volatility. The contemporaneous return correlation between the age and wealth factors is only about 0.2, which highlights the importance of including both factors in the pricing model.

Panel D of Figure 3 further illustrates the benefits of using both factors by reporting the performance of an equal-weighted portfolio of the age and wealth factors. This combined factor yields significantly higher performance than the market portfolio, while also displaying lower volatility than each factor taken separately. In the Appendix we report that the combined age-wealth factor an annual Sharpe ratio of 0.68, which is more than twice the Sharpe ratio of the market portfolio (0.32). For this reason, we refer to this factor as the *age-wealth investor factor* and denote it by AW_t in the following sections. The use of equal weights for constructing efficient factors builds on DeMiguel et al. (2009), who show that equal-weighted factor portfolios tend to perform best out of sample.

C. Spanning Regressions

We use spanning regressions to compare the pricing performance of our investor-based three-factor model to the pricing performance of benchmark firm-based models. Barillas and Shanken (2016) show that a candidate model's ability to price the full cross-section of stock

deviation portfolio $\boldsymbol{d_k}$ should yield a negative alpha *and* a high beta. In the context of our model, a positive exposure to the age and wealth factors can therefore be interpreted as a tilt away from $\boldsymbol{d_k}$.

returns better than a benchmark model is fully driven by the candidate model's ability to price the benchmark factors. Our testing procedure builds on this result.

Table IV, Panel A, reports OLS spanning regressions of the combined age-wealth factor, AW_t, defined in Section III.B, on different firm-based factor models. We consider the following sets of benchmark factors: 1) the market factor, 2) the market, size and value factors as in Fama and French (1993), 3) the market, size, value, and momentum factors as in Carhart (1997), 4) the market, size, value, profitability, and investment factors as in Fama and French (2015), and 5) the momentum and the Fama and French (2015) factors.

For every benchmark model, we reject the null hypothesis that pricing errors are zero at the 1% significance level. The monthly intercept of the combined age-wealth factor is as high as 1.05% when the market is the sole benchmark factor, with a t-value of 4.15. It decreases to 0.66% when all 6 firm factors are included but remains highly significant, with a t-value of 2.79. The combined age-wealth investor factor, AW_t, therefore captures pricing information that is not contained in the traditional firm factors.

[Insert Table IV here]

In Panel B of Table IV, we perform the reverse analysis and regress firm factors on the market and combined age-wealth factors. The alphas of the firm factors are smaller than those in Panel A, ranging from 0.343 to 0.528, and are all statistically insignificant. The combined age-wealth factor explains the traditional firm factors over the 1997-2018 period.

Overall, the results of the spanning regressions indicate that our three-factor investor-based asset pricing model spans the size, value, investment, profitability, and momentum factors. Our analysis therefore suggests that a model based on investor-based factors provides a complementary and parsimonious approach to equity pricing. This finding provides encouraging news to theoretical asset pricing models using investor characteristics as inputs.

D. Do Portfolio Holdings Contain Additional Pricing Information?

We next examine whether the age and wealth factors capture all the pricing information contained in individual investor data. To answer this question, we construct pricing factors from the additional investor socioeconomic characteristics described in Section II.B. That is, we consider gender, a retirement dummy, permanent income brackets, labor income-to-wealth brackets, occupation, and categorical variables for three levels of educational attainment. For each investor characteristic, C_t^i, we define the corresponding characteristic of stock j as the weighted average: $\mathrm{C}_{j,t}^* = \sum_i N_{j,t}^i \, \mathrm{C}_t^i / \sum_i N_{j,t}^i$. We then sort stocks by $\mathrm{C}_{j,t}^*$ and form the equity factor as a long-short portfolio of the sorted stocks.

In Table V, we regress each of these additional investor factors on the combined age-wealth factor, AW_t. Importantly, we find that none of the additional investor factors has a significant alpha against the combined age-wealth factor. This result suggests that age and wealth capture most of the pricing information contained in investor portfolio holdings. Furthermore, the additional investor factors have statistically significant loadings on AW_t. The female, education, permanent income, and retirement factors load positively on the combined age-wealth factor, whereas the labor-to-wealth factor loads negatively on the combined factor. Section IV further analyzes the sign of these loadings.

[Insert Table V here]

The results of Table V confirm our earlier findings. In the portfolio space, the age and wealth factors explain most of the cross-sectional heterogeneity in investor portfolios sorted by socioeconomic characteristics, as Section II.B documents. Correspondingly, Table V shows that in the return space, pricing factors based on alternative socioeconomic characteristics have significant loadings on the combined age-wealth factor but do not contain additional pricing information. This spanning property confirms the theoretical analysis of Section I, which predicts that pricing factors other the market are returns on portfolio factors.

E. Cross-Validation of Investor Factor Models

We next compare the out-of-sample performance of the investor-based and firm-based factor models considered in earlier sections. We do so by constructing maximum-Sharpe-ratio portfolios of the factors in sample and then by estimating their Sharpe ratios out of sample. An important pitfall is that a factor with an unusually high mean in sample is overweighted in the estimated tangency portfolio, which reduces out-of-sample portfolio performance. We correct this effect by implementing a bootstrap evaluation approach similar to Fama and French (2018).

We run bootstrap simulations from our sample of $T = 264$ months of factor return observations. We adopt the bootstrap aggregating ("bagging") method of Breiman (1996).

In each simulation, we draw with replacement a new dataset of size T, which we call the training sample. The hold-out sample consists of the months not included in the training sample.[22] We will refer to the training sample as in-sample data, and the hold-out sample as out-of-sample data.

We use the training sample to obtain the maximum-Sharpe-ratio factor from mean-variance optimization. To reduce estimation risk, we follow Kozak et al. (2020) and shrink the in-sample covariance matrix of factor returns, $\boldsymbol{\Sigma_f}$, as follows:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{f}} = \boldsymbol{\Sigma_f} + \gamma \boldsymbol{I}, \tag{20}$$

where γ is a shrinkage parameter and $\boldsymbol{I}$ is the identity matrix.[23] The estimated weights of the tangency portfolio are given by:

$$\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^{-1} \boldsymbol{\mu_f}, \tag{21}$$

where $\boldsymbol{\mu_f}$ is the vector of in-sample average returns on the factors. The shrinkage methodology (20) adds the fixed increment γ to the variance of each factor, which in (21) reduces portfolio weights the most for factors with the lowest volatility.

Consistent with earlier sections of the paper, we focus on a pricing model specified by the market factor, MKT_t, and K additional long-short factors. For each asset pricing model, we form the proxy tangency portfolio over the training period and then compute the portfolio's Sharpe ratio over the out-of-sample period. We repeat the simulation 100,000 times and calculate the average in-sample and out-of-sample Sharpe ratios for each factor model. Whereas in-sample Sharpe ratios are subject to an upward bias, out-of-sample Sharpe ratios are less affected by it because monthly returns are close to being serially uncorrelated, as Fama and French (2018) explain.

Column 1 of Table VI reports the average out-of-sample annualized Sharpe ratio for different factor models. Factor models are grouped according to the number of factors. On its own, the market portfolio has a Sharpe ratio of 0.32, consistent with typical estimates of market Sharpe ratios (Doeswijk et al., 2020).

[22] When T is large, the expected fraction of observations in the original data set that are contained in the training sample is $1 - \exp(-1) \approx 63.2\%$. The hold-out sample therefore contains about 97 months ($\exp(-1) \times 264$) on average.

[23] The details are based on Kozak, Nagel, and Santosh (2020). The parameter γ, which controls the level of shrinkage, is equal to $\gamma = \text{tr}(\boldsymbol{\Sigma_f})/\mathbb{E}(SR^2) \times T$, where SR denotes the Sharpe ratio. We select a value of 0.5 for the average Sharpe ratio and verify that other values of γ do not impact our conclusions.

[Insert Table VI here]

Among two-factor models combining the market portfolio and $K = 1$ additional factor, investor factors provide the highest Sharpe ratios. The age and market factors generates a Sharpe ratio of 0.51, while the wealth and market factors generate a Sharpe ratio of 0.54. No combination of a firm factor and the market performs better than investor-based factor models. The third best model consists of profitability and the market and has a Sharpe ratio of 0.49.

Among three-factor models combining the market portfolio and $K = 2$ additional factors, our three-factor investor-based model $(MKT_t, AGE_t, WEALTH_t)$ performs best, with a Sharpe ratio of 0.66 in annual units. The best two additional firm-based factors consist of profitability and momentum, which generate a Sharpe ratio of 0.55 in combination with the market.

Adding more firm factors to the market does not change the results. The portfolio containing the market and $K = 5$ firm factors generates an average Sharpe ratio of 0.50. This performance result is weaker than that generated by our preferred three-factor investor-based model.

Column 2 of Table VI shows that investor-based models are less subject to the upward in-sample bias than firm-based models. We report the ratio of out-of-sample to in-sample average Sharpe ratio for each model resulting from the optimization in (21). The ratio indicates the proportion of the in-sample average Sharpe ratio that is retained in the out-of-sample period. Investor factor models have the highest ratio. For example, the three-factor model consisting of the market, age, and wealth factors has a ratio of 73%, whereas the highest performing three-firm-factor model has a ratio of 68%. A natural explanation is that investor factors are less volatile than firm factors, so that the optimization method produces more accurate estimates of the tangency portfolio in sample and a higher Sharpe ratios out of sample with investor factors compared to firm factors. The Appendix verifies the validity of this logic.

As a robustness check, we consider an alternative method that assigns fixed weights to each of the model's pricing factors, as in DeMiguel et al. (2009) and Section III.B of the present paper. The weight on the market is set to unity and the weight on each long-short factor is set to $1/(2K)$. For instance, in the case of our three-factor investor-based model, the return of the maximum-Sharpe-ratio portfolio in year t is proxied by:

$$MKT_t + \frac{1}{4}AGE_t + \frac{1}{4}WEALTH_t. \tag{22}$$

A stock's weight in the proxy tangency portfolio (22) is the sum of the stock's weight in the market portfolio, MKT_t, and half the stock's weight in the combined age-wealth investor factor, AW_t. Column 3 of Table VI reports the corresponding results. The excellent out-of-sample performance of investor-based models is robust to this alternative estimation approach.

The results of this section show that our investor-based three-factor model is a strong contender for pricing the cross-section of stocks. The model matches or outperforms some of the best firm-based factor models available. This empirical result is in some sense remarkable because investor-based factor models are new to the literature. Our results provide empirical support for the strong links between portfolio choice and risk premia predicted by financial theory, and also reveal the rich pricing information available in large panels of individual holdings.

IV. The Cross-Section of Investor Factor Tilts

The strong pricing performance of the age and wealth factors raises the question of their economic origins. Are investor deviations from the tangency portfolio driven by hedging needs, sentiment, or a combination of both motives? In this section, we address this question by studying how the portfolio tilts of individual investors relate to their socioeconomic characteristics. Section IV.A documents how investors adjust their portfolio tilts toward investor factors as they migrate along the wealth distribution over the life-cycle. Section IV.B shows that socioeconomic characteristics other than age and wealth also drive investor portfolio tilts. In Section IV.C, we build a bridge between investor-based and firm-based factors by documenting the characteristics of the firms that make up investor factor portfolios, which is informative about the economic drivers of these factors.

A. *How Do Investor Portfolio Tilts Vary with Age and Wealth?*

We document how investors adjust their portfolio tilts toward the age and wealth factors as they age and experience variation in net worth. To break any mechanical link between the factors and socioeconomic characteristics, we partition investors into two randomly chosen groups. The first group contains two-thirds of the investor population and is used to reestimate the age and wealth factors.[24] The second group is used to study the links between portfolio tilts and characteristics.

We calculate the portfolio tilts of an investor as follows. Consider factor f with long leg H and short leg L at time t. The proportion of investor i's stock portfolio invested in equities contained in the long leg is:

$$\omega^i_{H,t} = \sum_{j=1}^{J} \omega^i_{j,t} \mathbb{1}_{j,H,t}, \tag{23}$$

where $\omega^i_{j,t}$ is the weight of stock j in investor i's stock portfolio and $\mathbb{1}_{j,H,t}$ is an indicator variable equal to unity if stock j belongs to the long leg H at time t. A similar definition provides the portfolio share invested in short leg stocks, $\omega^i_{L,t}$.

We define the investor's *portfolio tilt toward factor* f by

$$\omega^i_{f,t} = \omega^i_{H,t} - \omega^i_{L,t}. \tag{24}$$

The tilt is bounded between -1 and 1. It is equal to -1 if the investor only selects stocks in the short leg, 0 if the investor allocates equal amounts of capital to the long and short legs, and +1 if the investor only selects stocks in the long leg. This definition provides a convenient and direct measure of an investor's tilt toward a factor based only on portfolio holdings at a given date t.

In Panel A of Figure 4, we plot the average portfolio tilt toward the age factor for the 10 groups of investors sorted by age. The groups are described in Section II.B. Means are equally-weighted and estimated over the full 1997-2018 sample. The age tilt is about 0.1 for investors under 30 and progressively increases to 0.4 for older investors. The panel shows a substantial and remarkably linear migration in the age factor tilt over the life-cycle.

[Insert Figure 4 here]

[24] In the Appendix, we verify that investor factors constructed from a subset of the investor population contain similar pricing information as the full-sample factors, albeit with lower accuracy.

The "age ladder" illustrated in Panel A of Figure 4 relates to the findings in Betermier, Calvet, and Sodini (2017), who report a progressive life-cycle migration toward the value factor among Swedish households. This earlier paper shows that the linearity between the value tilt and age originates from life-cycle variation in age and other characteristics rather than from combinations of time and cohort fixed effects. The reason is that, in order to generate such a linear structure, cohort and year fixed effects would have to offset each other exactly, a zero probability event in a unrestricted panel model. The same logic applies to the age factor tilt in the present paper.

In Panel B of Figure 4, we plot the average age factor tilt of investors who are new to direct stock market investing (dotted line).[25] The portfolio tilts chosen by new entrants closely mimic the tilts of seasoned investors of the same age. This result confirms that the age ladder is unlikely to be due to portfolio inertia or time variation in firm characteristics. Instead, investors progressively adjust their age tilts over the life cycle.

In Panel C of Figure 4, we obtain similar results for the average tilt toward the wealth factor for 12 groups of investors sorted by net worth. The groups are described in Section II.B. Investors progressively migrate toward the wealth factor as they climb the wealth ladder. This migration is again economically significant. The wealth factor tilt is as low as -0.15 for investors in the bottom 10% (first bracket) and reaches 0.1 for investors in the top 0.1% (12^{th} bracket). The difference is most pronounced among the wealthiest investors.

Panel D of Figure 4 shows that new stock market entrants choose wealth factor tilts similar to those of equally wealthy pre-existing investors. Altogether, these results confirm that the factor tilts of investors vary with age and wealth as one would expect, even among investors in their first year of stock market investing.

B. Which Other Investor Characteristics Drive Portfolio Tilts?

We next examine if individual characteristics other than age and wealth can predict tilts toward investor factors.

Additional Socioeconomic Characteristics. We consider two sets of variables that have been shown to explain portfolio decisions in household finance research.

[25] Each point estimate contains at least 1,000 investors. Groups that do not satisfy this requirement are dropped.

The first set of characteristics captures risk exposures (see e.g., Cocco, Gomes, and Maenhout, 2005; Gomes and Michaelides, 2005; Heaton and Lucas, 1997, 2000; Viceira, 2001). We measure indebtedness by the debt-to-income ratio, which captures the investor's ability to withstand economic shocks (Campbell, 2006; Iacovello, 2008). We also compute the sensitivity of investor non-financial income to macroeconomic risk as in Guvenen et al. (2017). Specifically, we form 220 groups of investors sorted by employment sector, retirement status, and labor income percentile. For each group g, we run a panel regression of the annual income growth of investor i in year t, denoted by $\Delta y_{i,t}$, on real GDP growth in the same year:

$$\Delta y_{i,t} = a_g + \beta_g^{GDP} \Delta GDP_t + \varepsilon_{i,t}. \tag{25}$$

The regression yields a slope coefficient β_g^{GDP} for each group. We assign β_g^{GDP} to all individuals in the group and use it as a proxy for their exposure to macroeconomic risk. Estimation details are provided in the Appendix.

The second set of characteristics proxy for behavioral traits that may affect an investor's portfolio tilts toward the age and wealth factors. The impact of stock market experience on portfolio choice has been documented in empirical studies and field experiments (List, 2003; Seru, Shumway, and Stoffman, 2010). We correspondingly measure experience by the number of years an investor has held stocks. We also include a set of dummy variables corresponding to graduate education, business education, finance sector occupation, and gender. Previous research has shown that biases such as overconfidence are more prevalent among men than women (Barber and Odean, 2001) and more pronounced among investors with low education than among more investors with high education (Calvet, Campbell, and Sodini, 2009).

Panel Regressions. We run panel regressions of the age and wealth factor tilts on investor characteristics:

$$\omega_{f,t}^i = \eta_t + \boldsymbol{\gamma' X_t^i} + \epsilon_t^i, \tag{26}$$

where η_t is a time fixed effect, $\boldsymbol{X_t^i}$ is a vector of characteristics, and ϵ_t^i is the residual error term. The vector $\boldsymbol{X_t^i}$ includes the investor's debt-to-income ratio, macroeconomic risk exposure, gender, stock market experience, education variables, a finance occupation dummy, and indicator variables for age and wealth brackets. We use the 10 age groups and the 12 wealth brackets defined in Section II.B and we select median brackets as benchmarks. Standard are clustered by year and investor.

In Table VII, we report the results of the panel regression (26) when the dependent variable is the age factor tilt. Age remains statistically significant for most groups after controlling for the additional characteristics. Young investors, as represented by the first five age groups, tilt away from the age factor, whereas mature investors have positive tilts.

[Insert Table VII here]

Both measures of risk exposure are negatively related to the age factor tilt. The effect of income beta is particularly strong. A 0.5 difference in income beta, which approximately corresponds to the difference between working in public administration and working in the tourism industry for an individual with median income, is associated with a 0.045 reduction in the age factor tilt.

The results are also consistent with sentiment driving portfolio tilts. Graduate education, business education, finance sector occupation, and stock market experience are all associated with a higher age factor tilt. In terms of economic magnitude, ten years of additional experience explain a 0.14 increase in the age factor tilt. Female investors also have a greater age factor tilt than male investors. Gender has approximately the same effect on the age factor tilt as ten years of stock market experience.

[Insert Table VIII here]

Table VIII presents remarkably similar results for the wealth factor tilt. As with age, the explanatory power of the wealth dummy variables is robust to the inclusion of other characteristics. Less affluent investors have a negative wealth tilt, whereas more affluent investors have a significantly positive tilt. Characteristics based on risk exposure and sentiment also explain variation in the wealth factor tilt. A 0.5 increase in the income beta is associated with a 0.025 reduction in the wealth factor tilt. Ten years of stock market experience explain a 0.07 increase in the portfolio tilt. Being female and having a graduate degree or business education also predict a higher wealth factor tilt.

Taken together, these results suggest that hedging motives and sentiment jointly drive the cross-sectional variation in investor portfolio tilts. On the one hand, investors with low risk exposures have a stronger tilt toward the age and wealth factors than investors with high risk exposures, and investors progressively migrate toward the two factors as they become more mature and wealthier, consistent with the theoretical framework in Section I.C.[26] On the

[26] In the Appendix, we verify that the wealth factor correlates with a factor constructed from investors' wealth-to-income ratio, as the ICAPM predicts.

other hand, the positive relations between factor tilts and measures of financial sophistication suggest the presence of a parallel behavioral channel. Younger and less wealthy investors who are more prone to sentiment tend to tilt away from the age and wealth factors. This complementary explanation is consistent with empirical evidence on correlated sentiment trades in the portfolios of retail investors (Barber et al., 2009; Kumar and Lee, 2006).[27]

C. Firm Characteristics of Investor-Based Factors

To provide a bridge between firm-based and investor-based factors, we next analyze the characteristics of firms that make up the age and wealth factor portfolios. We consider the following firm characteristics: size, book-to-market ratio, profitability, investment, return volatility, the proportion of equity held by institutional investors, and share turnover, defined as the number of shares traded in a year divided by the number of free-float shares outstanding at the beginning of the year. For each factor, we consider four portfolios, corresponding to the stocks in the bottom 30% of the investor characteristic (short leg L), the middle 30%-70% bracket (M), the top 30% (long leg H), and the long-short portfolio H-L defining the factor portfolio.

Table IX reports the median characteristic of each portfolio, where the median is measured on the pooled cross-section. The table highlights clear differences in the properties of stocks in the long and short legs of investor factors. Stocks held by young and less wealthy investors have significantly higher volatility, higher share turnover, and lower institutional ownership than stocks held by mature and wealthy investors. These results are consistent with prior work arguing that these types of stocks are more difficult to arbitrage and therefore more sensitive to changes in sentiment (Stambaugh and Yuan, 2017). These properties of the long and short legs of investor factors, combined with the evidence on individual portfolio tilts reported in Section IV.B, provide further evidence that sentiment contributes to the strong pricing properties of the age and wealth factors.

[Insert Table IX here]

[27] The results are also consistent with Korniotis and Kumar (2011), who find that older U.S. retail investors are better diversified, trade less frequently, invest in lower-fee funds, and exhibit weaker behavioral biases than younger investors. One difference in their study is that they find older U.S. retail investors generally performed worse than younger investors between 1991 and 1996. One possible explanation for this result is the specific period used in their analysis. Our evidence about the high performance of older investors is based on 22 years of monthly return data.

Table IX also reveals that stocks held by mature and affluent investors tend to have a higher market capitalization, higher profitability, lower investment, and lower CAPM betas than stocks held by the young and the less wealthy. These links are important for several reasons. First, they support Koijen and Yogo (2019)'s modeling assumption that investor portfolio holdings are related to firm characteristics. Second, they reveal that mature and wealthy investors tend to invest in the same stocks as institutional investors, which Koijen and Yogo (2019) study in the U.S. context. This finding suggests that the observed dispersion in the direct stock holdings of individual investors contains valuable information about portfolio tilts outside the retail sector.

V. Conclusion

This paper constructs a parsimonious set of equity factors from the cross-section of individual investor portfolio holdings. We show theoretically that portfolios of stocks sorted by the age or wealth of their individual investors should produce powerful pricing factors. Using the complete stockholdings of Norwegian individual investors, we verify empirically that a three-factor model consisting of a mature-minus-young factor, a high wealth-minus-low wealth factor, and the market factor explains the bulk of common variation in portfolio holdings and prices the cross-section of stock returns. We also uncover a rich set of links between investor characteristics and portfolio tilts toward the age and wealth factors.

The analysis of investor factors opens new opportunities for equity pricing research. The tight connection between investor factors and the cross-section of portfolio holdings makes it possible to connect equity risk premia to the drivers of investor demand. Our finding that hedging motives and sentiment operate in tandem suggests that there might be interdependencies between both channels, as Kozak, Nagel, and Santosh (2018) explain.

Another interesting question is whether investor-based factors price other asset classes. This question seems important because limitations on firm accounting data may limit the statistical ability of traditional firm factors to price alternative asset classes such as private equity. Information on the characteristics of investors who own these assets provides an alternative avenue for pricing them. We leave these questions for future research.

REFERENCES

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini, 2020, Rich pickings? Risk, return, and skill in household wealth, *American Economic Review* 110, 2703–2747.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Balasubramaniam, Vimal, John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish, 2021, Who owns what? A factor model for direct stockholding, *Journal of Finance* forthcoming.

Bali, Turan G., A. Doruk Gunyadin, Thomas Jansson, and Yigitcan Karabulut, 2020, Do the rich gamble in the stock market? Low risk anomalies and wealthy households, Working paper, Georgetown University.

Barber, Brad, and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.

Barber, Brad M., Xing Huang, and Terrance Odean, 2016, Which factors matter to investors? Evidence from mutual fund flows, *Review of Financial Studies* 29, 2600–2642.

Barber, Brad M., and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Systematic noise, *Journal of Financial Markets* 12, 547–569.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-capm: An extrapolative capital asset pricing model, *Journal of Financial Economics* 115, 1–24.

Barillas, Francisco, and Jay Shanken, 2016, Which alpha?, *Review of Financial Studies* 30, 1316–1338.

Berk, Jonathan B., and Jules H. van Binsbergen, 2016, Assessing asset pricing models using revealed preference, *Journal of Financial Economics* 119, 1–23.

Betermier, Sebastien, Laurent E. Calvet, and Evan Jo, 2022, A supply and demand approach to capital markets, Working paper, EDHEC Business School and McGill University.

Betermier, Sebastien, Laurent E. Calvet, and Paolo Sodini, 2017, Who are the value and growth investors?, *Journal of Finance* 72, 5–46.

Blume, Marshall E., and Donald B. Keim, 2012, Institutional investors and stock market liquidity: Trends and relationships, Working paper, The Wharton School.

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Breiman, Leo, 1996, Bagging predictors, *Machine Learning* 24, 123–140.

Büchner, Matthias, 2020, What drives asset holdings? Commonality in investor demand, Working paper.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini, 2007, Down or out: Assessing the welfare costs of household investment mistakes, *Journal of Political Economy* 115, 707–747.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini, 2009, Fight or flight? Portfolio rebalancing by individual investors, *Quarterly Journal of Economics* 124, 301–348.

Campbell, John, Stefano Giglio, Christopher Polk, and Robert Turley, 2018, An intertemporal CAPM with stochastic volatility, *Journal of Financial Economics* 128, 207–233.

Campbell, John Y., 2006, Household finance, *Journal of Finance* 61, 1553–1604.

Campbell, John Y., 2018, *Financial Decisions and Markets: A Course in Asset Pricing* (Princeton University Press).

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Carpenter, Seth, Selva Demiralp, Jane Ihrig, and Elizabeth Klee, 2015, Analyzing Federal Reserve asset purchases: From whom does the Fed buy?, *Journal of Banking & Finance* 52, 230–244.

Choi, James, and Adriana Robertson, 2020, What matters to individual investors? Evidence from the horse's mouth, *Journal of Finance* 75, 1965–2020.

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

Cochrane, John H., 2011, Discount rates, *Journal of Finance* 66, 1047–1108.

Connor, Gregory, and Robert Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 15, 373–394.

Connor, Gregory, and Robert Korajczyk, 1988, Risk and return in equilibrium apt: Application of new test methodology, *Journal of Financial Economics* 21, 255–290.

Constantinides, Georges M., 2017, Asset pricing: Models and empirical evidence, *Journal of Political Economy* 125, 1782–1788.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

Doeswijk, Ronald, Trevin Lam, and Laurens Swinkels, 2020, Historical returns of the market portfolio, *Review of Asset Pricing Studies* 10, 521–567.

Ehling, Paul, Alessandro Graniero, and Christian Heyerdahl-Larsen, 2018, Asset prices and portfolio choice with learning from experience, *Review of Economic Studies* 85, 1752–1780.

Fagereng, Andreas, Charles Gottlieb, and Luigi Guiso, 2017, Asset market participation and portfolio choice over the life-cycle, *Journal of Finance* 72, 705–750.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri, 2020, Heterogeneity and persistence in returns to wealth, *Econometrica* 88, 115–170.

Fama, Eugene F, and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–466.

Fama, Eugene F, and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 43, 3–56.

Fama, Eugene F., and Kenneth R. French, 2007, Disagreements, tastes, and asset prices, *Journal of Financial Economics* 83, 667–689.

Fama, Eugene F., and Kenneth R. French, 2012, Size, value, and momentum in international stock returns, *Journal of Financial Economics* 105, 457–472.

Fama, Eugene F, and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Fedyk, Yurii, Christian Heyerdahl-Larsen, and Johan Walden, 2013, Market selection and welfare in a multi-asset economy, *Review of Finance* 17, 1179–1237.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five facts about beliefs and portfolios, *Journal of Political Economy* 111, 1481–1522.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai, 2021, Household finance, *Journal of Economic Literature* 59, 919–1000.

Gomes, Francisco, and Alexander Michaelides, 2005, Optimal life-cycle asset allocation: Understanding the empirical evidence, *Journal of Finance* 60, 869–904.

Greenwood, Robin, and Stefan Nagel, 2009, Inexperienced investors and bubbles, *Journal of Financial Economics* 93, 239–258.

Griffin, John M., Xiuqing Ji, and J. Spencer Martin, 2003, Momentum investing and business cycle risk: Evidence from pole to pole, *Journal of Finance* 58, 2515–2547.

Guercio, Diane Del, and Paula A. Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis* 37, 523–557.

Guiso, Luigi, and Paolo Sodini, 2013, Household finance: An emerging field, in Georges M. Constantinides, Milton Harris, and Rene M. Stulz, eds., *Handbook of the Economics of Finance*, volume 2.

Guvenen, Fatih, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo, 2017, Worker betas: Five facts about systematic earnings risk, *American Economic Review* 107, 398–403.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2015, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

He, Zhiguo, and Wei Xiong, 2013, Delegated asset management, investment mandates, and capital immobility, *Journal of Financial Economics* 107, 239–258.

Heaton, John, and Deborah Lucas, 1997, Market frictions, savings behavior, and portfolio choice, *Macroeconomic Dynamics* 1, 76–101.

Heaton, John, and Deborah J. Lucas, 2000, Portfolio choice and asset prices: The importance of entrepeneurial risk, *Journal of Finance* 55, 1163—1198.

Hirshleifer, David, 2015, Behavioral finance, *Annual Review of Financial Economics* 7, 133–159.

Hoffmann, Peter, Sam Langfield, Federico Pierobon, and Guillaume Vuillemey, 2018, Who bears interest rate risk?, *Review of Financial Studies* 32.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2018, Which factors?, *Review of Finance* 23, 1–35.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Iacovello, Matteo, 2008, Household debt and income inequality, 1963-2003, *Journal of Money, Credit, and Banking* 40, 929–965.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual investor trading and stock returns, *Journal of Finance* 63, 273–310.

Kelley, Eric K., and Paul C. Tetlock, 2013, How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229–1265.

Koijen, Ralph, Francois Koulischer, Benoit Nguyen, and Motohiro Yogo, 2020a, Inspecting the mechanism of quantitative easing in the euro area, *Journal of Financial Economics* forthcoming.

Koijen, Ralph S. J, Robert J. Richmond, and Motohiro Yogo, 2020b, Which investors matter for equity valuations and expected returns?, Working paper.

Koijen, Ralph S. J, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Koijen, Ralph S. J, and Motohiro Yogo, 2020, Exchange rates and asset prices in a global demand system, Working paper.

Korniotis, George M, and Alok Kumar, 2011, Do older investors make better investment decisions?, *The Review of Economics and Statistics* 93, 244–265.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance* 73, 1183–1223.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Kumar, Alok, and Charles Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451–2486.

Lintner, John, 1966, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

List, John A., 2003, Does market experience eliminate market anomalies?, *Quarterly Journal of Economics* 118, 41–71.

Ludvigson, Sydney C., 2013, Advances in consumption-based asset pricing: Empirical tests, in George M. Constantinides, Milton Harris, and Rene M. Stulz, eds., *Handbook of the Economics of Finance* (Elsevier).

Maggiori, Matteo, Brent Neiman, and Jesse Schreger, 2020, International currencies and capital allocation, *Journal of Political Economy* 128, 239–258.

Mehra, Rajnish, 2012, Consumption-based asset pricing models, *Annual Review of Financial Economics* 4, 385–409.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Sandroni, Alvaro, 2000, Do markets favor agents able to make accurate predictions?, *Econometrica* 68, 1303–1341.

Seru, Amit, Tyler Shumway, and Noah Stoffman, 2010, Learning by trading, *Review of Financial Studies* 23, 705–739.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Skoulakis, Georgios, 2008, Dynamic portfolio choice with Bayesian learning, Working paper.

Stambaugh, Robert, and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

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

Vissing-Jorgensen, Anne, 2003, Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions, *NBER Macroeconomics Annual* 18, 139–194.

Table I
Principal Component Analysis of Investor Portfolio Holdings

This table reports the principal components analysis of 93 aggregate portfolios of individual investors grouped by socioeconomic and geographic indicators. The first row reports the proportion of the cross-sectional variance of portfolio holdings explained by each of the top 10 principal components (PCs). The proportions are calculated each year in 2002-2017 and then averaged across years. The second row reports the cumulative proportion of the cross-sectional variance explained by each PC. The third and fourth sets of rows report the cumulative proportion of the cross-sectional variance explained by a projection of the PC on the market portfolio (third row) or on the market, age, and wealth portfolios (fourth row). Specifically, the variance proportions are calculated by summing up the variance of each PC multiplied by the regression model's R^2 coefficient. The age portfolio is long the portfolio of investors above 60 and short the portfolio of investors below 30. The wealth portfolio is long the portfolio of investors in the top 10% of the net worth distribution and short the portfolio of investors in the bottom 30%.

	Top Principal Components									
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Variance explained	0.74	0.07	0.04	0.03	0.02	0.02	0.01	0.01	0.01	0.01
Cumulative Variance Explained By:										
PC	0.74	0.81	0.85	0.88	0.90	0.92	0.93	0.95	0.95	0.96
Market	0.27	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Market, age, wealth	0.67	0.71	0.71	0.72	0.73	0.73	0.73	0.73	0.73	0.73

Table II
Summary Statistics on Investor Characteristics

This table reports summary statistics on (i) the age and wealth characteristics of Norwegian individuals investors and (ii) the age and wealth characteristics of Norwegian stocks. The analysis is based on investors holding stocks directly in 2017. For each characteristic, we report the standard deviation, mean, and the 10^{th}, 30^{th}, 50^{th}, 70^{th}, 90^{th}, and 99^{th} percentiles. Investor wealth is defined as the value of liquid and illiquid assets (liquid financial wealth, real estate, vehicles, business assets) net of liabilities. Wealth is expressed both in million NOK and on a 1 to 12 scale, where the first 9 categories represent the first 9 deciles of the wealth distribution, the 10^{th} category the 90-99^{th} percentiles, the 11^{th} category the 99-99.9^{th} bracket, and the 12^{th} category the top 0.1%. A stock's age characteristic is the average age of the stock's individual investor shareholders, weighted by the number of shares that they own at the beginning of the year. A stock's wealth characteristic is similarly defined.

			Percentiles					
Characteristic	SD	Mean	10^{th}	30^{th}	50^{th}	70^{th}	90^{th}	99^{th}
Investor Level								
Age (years)	16.5	55.1	32.0	46.0	56.0	65.0	76.0	89.0
Wealth (million NOK)	48.0	6.2	-0.0	1.5	3.1	5.2	11.1	50.8
Wealth bucket (on 1-12 scale)	3.0	7.0	2.0	5.0	8.0	9.0	10.0	11.0
Stock Level								
Age (years)	7.5	57.0	46.8	53.8	57.4	61.6	66.2	70.9
Wealth bucket (on 1-12 scale)	1.9	8.4	5.8	7.8	8.7	9.4	10.4	11.6

Table III
Return Performance of Age and Wealth Factors

This table reports the average historical performance on the stock portfolios held by Norwegian individual investors sorted by age or wealth in 1997-2008. The analysis is based on monthly return data on the universe of Norwegian stocks defined in section II.A. Panel A reports monthly value-weighted average excess returns. Panel B reports the intercept and Panel C reports the slope coefficient of times-series OLS regressions of monthly excess portfolio returns on the market factor. For each characteristic, the L portfolio correspond to investors in the bottom 30% of the characteristic's distribution, the M portfolio to investors in the mid 40%, and the H portfolio to investors in the top 30%. The investor factor H-L is long the H portfolio and short the L portfolio.

Panel A: Monthly Returns								
	Average Return				t(Average Return)			
	L	M	H	H-L	L	M	H	H-L
Age	0.09	0.91	1.07	0.98	0.17	2.12	2.96	2.37
Wealth	0.08	1.04	1.00	0.92	0.17	2.73	2.38	2.59
Panel B: Monthly CAPM Alphas								
	Alpha				t(Alpha)			
	L	M	H	H-L	L	M	H	H-L
Age	-0.82	0.02	0.26	1.08	-2.36	0.14	2.27	2.65
Wealth	-0.85	0.19	0.17	1.01	-2.83	1.93	0.75	2.91
Panel C: Monthly CAPM Betas								
	Beta				t(Beta)			
	L	M	H	H-L	L	M	H	H-L
Age	1.13	1.07	0.94	-0.19	19.41	36.69	49.50	-2.81
Wealth	1.17	1.01	0.99	-0.18	23.20	59.77	26.74	-3.12

Table IV
Spanning Regressions

Panel A reports OLS spanning regressions of the combined age-wealth factor, AW_t, on firm factors in 1997-2018. Panel B reports regressions of firm factors on AW_t. The combined age-wealth factor, AW_t, is an equally-weighted average of the age and wealth investor factors. The market factor, MKT_t, is the value-weighted portfolio of Norwegian stocks. We sort stocks by market capitalization to obtain the size factor, SMB_t, by industry adjusted book-to-market ratio to obtain the value factor, HML_t, by profit margin to obtain the profitability factor, RMW_t, and by total asset growth to obtain the investment factor, CMA_t. To construct the momentum factor, MOM_t, we consider the geometric return over the previous 12 months where the most recent month is left out. Standard errors are in parentheses and statistical significance is indicated by ***, **, and * for the 0.01, 0.05, and 0.10 levels.

Panel A: Regressions of Combined Age-Wealth Factor on Firm Factors

	Dependent Variable: Combined Age-Wealth Factor				
	(1)	(2)	(3)	(4)	(5)
Constant	1.047***	1.066***	0.852***	0.759***	0.661***
	(0.288)	(0.257)	(0.243)	(0.245)	(0.237)
MKT	−0.187***	−0.305***	−0.255***	−0.199***	−0.183***
	(0.048)	(0.045)	(0.043)	(0.046)	(0.044)
SMB		−0.479***	−0.452***	−0.311***	−0.327***
		(0.066)	(0.062)	(0.068)	(0.065)
HML		−0.186***	−0.159***	−0.164***	−0.149***
		(0.051)	(0.048)	(0.048)	(0.046)
MOM			0.200***		0.151***
			(0.033)		(0.033)
RMW				0.278***	0.217***
				(0.047)	(0.047)
CMA				0.061	0.053
				(0.046)	(0.044)
Observations	264	264	264	264	264
Adjusted R^2	0.050	0.248	0.339	0.346	0.393

Panel B: Regressions of Firm Factors on Combined Age-Wealth Factor

	Dependent Variable:				
	SMB	HML	MOM	RMW	CMA
	(1)	(2)	(3)	(4)	(5)
Constant	0.343	0.424	0.456	0.341	0.528
	(0.225)	(0.313)	(0.435)	(0.320)	(0.333)
MKT	−0.267***	−0.164***	−0.103	−0.093*	−0.194***
	(0.038)	(0.053)	(0.073)	(0.054)	(0.056)
Age-Wealth	−0.348***	−0.252***	0.570***	0.591***	0.123*
	(0.047)	(0.065)	(0.091)	(0.067)	(0.070)
Observations	264	264	264	264	264
Adjusted R^2	0.239	0.064	0.149	0.257	0.061

Table V
Additional Investor Factors

This table reports the intercept, slope, and R^2 coefficient of time-series OLS regressions of additional investor factors on the combined age-wealth factor, AW_t. Investor factors are constructed by sorting stocks on a particular characteristic and constructing a portfolio that takes a long (short) position in the top (bottom) 30% of stocks sorted by the characteristic. Investor characteristics include a male dummy, a categorical variable equal to 1, 2, or 3 based on the investor's educational attainment, labor income-to-wealth brackets, permanent income brackets, a retirement dummy, and occupational sector.

	Regressions of Additional Factor on Age-Wealth Factor AW_t				
	α	t(α)	b	t(b)	R^2
Additional Factor Defined by Socioeconomic Characteristic:					
Male dummy	0.05	0.15	-0.67	-8.90	0.23
Education level	-0.04	-0.11	0.16	2.26	0.02
Labor-to-wealth	-0.49	-1.49	-0.18	2.69	0.03
Permanent income	-0.51	-1.44	0.19	2.56	0.02
Retirement dummy	0.06	0.22	0.89	15.01	0.46
Additional Factor Defined by Occupational Sector:					
Resource industries	0.20	0.53	0.09	1.22	0.01
Petroleum	0.06	0.17	-0.11	-1.54	0.01
Consumer manufacturing	0.12	0.33	-0.03	-0.37	0.00
Material manufacturing	-0.25	-0.74	0.08	1.11	0.00
Technological manufacturing	0.02	0.05	-0.33	-4.52	0.08
Public administration	0.15	0.51	0.16	2.82	0.03
Construction	0.24	0.73	-0.48	-7.11	0.17
Trade	-0.06	-0.16	-0.27	-3.55	0.05
Transportation and logistics	-0.26	-0.69	-0.26	-3.50	0.05
Tourism	-0.04	-0.12	-0.06	-0.91	0.00
Media and ICT	0.33	0.98	-0.37	-5.47	0.11
Finance	0.15	0.41	-0.02	-0.27	0.00
Knowledge-based business services	-0.26	-0.76	-0.22	-3.24	0.04
Technological services	-0.18	-0.49	0.04	0.61	0.00
Business support services	0.13	0.31	-0.12	-1.47	0.01
Education	-0.05	-0.13	0.13	1.70	0.01
Health and social services	0.17	0.55	0.16	2.68	0.03
Non-profit and household services	-0.08	-0.24	0.15	2.21	0.02
Real estate activities	0.64	1.60	-0.23	-2.88	0.03

Table VI
Out-of-Sample Performance of Factor Models

This table reports out-of-sample (OS) Sharpe ratios of mean-variance efficient portfolios of pricing factors. The analysis is based on 100,000 simulations as in Fama and French (2018). In each simulation, we randomly draw with replacement a new data set of size T from our dataset of $T = 264$ months, which defines the in-sample sample, and evaluate performance in the subset of observations that have not been drawn, which we call the out-of-sample sample. For columns 1 and 2, we obtain the efficient portfolio by mean-variance optimization of in-sample factor returns and a shrinkage estimator of the covariance matrix. In column 1, we report the average out-of-sample Sharpe ratio across all simulations. In column 2, we scale the average Sharpe ratio by the average in-sample (IS) Sharpe ratio. In column 3, the efficient portfolio is proxied by setting a weight of 1 on the market factor and a weight of $1/(2K)$ on every other factor, where K is the number of factors other than the market. The Firm-4 model includes MKT, SMB, HML, and MOM, Firm-5 includes MKT, SMB, HML, RMW, and CMA, and Firm-6 includes MKT, SMB, HML, RMW, CMA and MOM.

	Optimized Weights		Fixed Weights
	OS Sharpe Ratio (1)	OS-IS Ratio (2)	OS Sharpe Ratio (3)
MKT, AGE	0.51	0.74	0.58
MKT, WEALTH	0.54	0.75	0.57
MKT, SMB	0.13	0.48	0.32
MKT, HML	0.17	0.44	0.34
MKT, MOM	0.44	0.69	0.55
MKT, CMA	0.34	0.61	0.15
MKT, RMW	0.49	0.72	0.56
MKT, AGE, WEALTH	0.66	0.73	0.61
MKT, SMB, HML	0.08	0.24	0.35
MKT, SMB, MOM	0.39	0.59	0.46
MKT, SMB, CMA	0.29	0.49	0.24
MKT, SMB, RMW	0.48	0.63	0.45
MKT, HML, MOM	0.38	0.55	0.48
MKT, HML, CMA	0.26	0.42	0.25
MKT, HML, RMW	0.43	0.59	0.48
MKT, CMA, RMW	0.52	0.65	0.36
MKT, CMA, MOM	0.48	0.62	0.37
MKT, RMW, MOM	0.55	0.68	0.59
Firm-4	0.34	0.48	0.44
Firm-5	0.44	0.50	0.36
Firm-6	0.50	0.52	0.41
Firm-6, AGE, WEALTH	0.65	0.58	0.48

Table VII
Panel Regressions of the Age Factor Tilt on Investor Characteristics

This table reports panel regressions of the age factor tilt on investor characteristics and age dummy variables. The estimation is run on a panel of Norwegian individual investors in 1997-2018. The age factor tilt is calculated annually from the direct stockholdings of investors. Income beta is the slope coefficient from a panel regression of an investor's annual income growth on real GDP growth, where the estimation is conducted on a group of investors in the same employment sector and labor income bracket. The debt-to-income ratio is the ratio of an investor's total debt to labor income. Stock market experience is defined as the number of years of stock market participation. The male dummy, the Master's degree dummy, the business education dummy, and the finance occupation dummy are indicator variables respectively equal to unity if the investor is male, has obtained a Master's degree, has studied business or economics, or works in a finance-related sector. The age dummy variables correspond to 10 groups of investors in five year increments. The median age group (50-55 years) is used as the reference point and the corresponding dummy is removed from the estimation. We include year fixed effects and twelve wealth-bracket fixed effects. Statistical significance is indicated by ***, **, and * for the 0.01, 0.05, and 0.10 levels. Standard errors are clustered at the calendar year and investor levels.

	Dependent Variable: Age Factor Tilt				
	(1)	(2)	(3)	(4)	(5)
Risk Exposures					
Income beta	-0.093***	-0.089***	-0.064***	-0.122***	-0.088***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Debt-to-income ratio			-0.010***		-0.010***
			(0.003)		(0.002)
Experience, Education, and Gender					
Stock market experience	0.014***	0.014***	0.015***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Male dummy			-0.135***		-0.134***
			(0.007)		(0.006)
Master's degree dummy				0.021**	0.027***
				(0.009)	(0.009)
Business education dummy				0.024***	0.028***
				(0.007)	(0.007)
Finance occupation dummy				0.116**	0.088
				(0.055)	(0.054)

(Continued)

Table VII - *Continued*

	Dependent Variable: Age Factor Tilt				
	(1)	(2)	(3)	(4)	(5)
Age Group Dummies					
$<$ 30	-0.041***	-0.033**	-0.021*	-0.032**	-0.021*
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
30-34	-0.072***	-0.062***	-0.054***	-0.065***	-0.059***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)
35-39	-0.073***	-0.065***	-0.059***	-0.067***	-0.062***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)
40-44	-0.056***	-0.052***	-0.048***	-0.052***	-0.050***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
45-49	-0.028***	-0.026***	-0.024***	-0.026***	-0.024***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
55-59	0.026***	0.024***	0.023***	0.025***	0.024***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
60-64	0.023***	0.021***	0.025***	0.018***	0.022***
	(0.007)	(0.007)	(0.007)	(0.004)	(0.005)
65-69	0.011	0.011	0.023*	-0.004	0.013
	(0.012)	(0.012)	(0.013)	(0.009)	(0.009)
$\geq$ 70	0.049***	0.050***	0.065***	0.022*	0.042***
	(0.014)	(0.014)	(0.015)	(0.012)	(0.012)
Year FE	Yes	Yes	Yes	Yes	Yes
Wealth bracket FE	No	Yes	Yes	Yes	Yes
Number of observations	943,457	943,457	943,457	880,319	880,319
Adjusted R^2	0.067	0.068	0.078	0.068	0.078

Table VIII
Panel Regressions of the Wealth Factor Tilt on Investor Characteristics

This table reports panel regressions of the wealth factor tilt on investor characteristics and wealth dummy variables. The estimation is run on a panel of Norwegian individual investors in 1997-2018. The wealth factor tilt is calculated annually from the direct stockholdings of investors. Income beta is the slope coefficient from a panel regression of an investor's annual income growth on real GDP growth, where the estimation is conducted within a group of investors in the same employment sector and labor income bracket. The debt-to-income ratio is the ratio of an investor's total debt to labor income. Stock market experience is defined as the number of years of stock market participation. The male dummy, the Master's degree dummy, the business education dummy, and the finance occupation dummy are indicator variables respectively equal to unity if the investor is male, has obtained a Master's degree, has studied business or economics, or works in a finance-related sector. The wealth dummy variables correspond to the first 9 deciles, the 90^{th}-99^{th} percentiles, the 99^{th}-99.9^{th} percentiles, and the top 0.1% of the wealth distribution. The median wealth group (50^{th}-60^{th} percentiles) is used as the reference point and the corresponding dummy is removed from the estimation. We include year fixed effects and ten age-group fixed effects. Statistical significance is indicated by ***, **, and * for the 0.01, 0.05, and 0.10 levels. Standard errors are clustered at the calendar year and investor levels.

	Dependent Variable: Wealth Factor Tilt				
	(1)	(2)	(3)	(4)	(5)
Risk Exposures					
Income beta	-0.047***	-0.044***	-0.036***	-0.070***	-0.059***
	(0.008)	(0.009)	(0.008)	(0.010)	(0.011)
Debt-to-income ratio		0.001		0.001	
		(0.002)		(0.002)	
Experience, Education, and Gender					
Stock market experience	0.007***	0.007***	0.007***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Male dummy			-0.046***		-0.043***
			(0.012)		(0.014)
Master's degree dummy				-0.002	0.0003
				(0.010)	(0.011)
Finance education dummy				0.032***	0.033***
				(0.007)	(0.008)
Finance occupation dummy				0.077**	0.067*
				(0.031)	(0.033)

(Continued)

Table VIII - *Continued*

	Dependent Variable: Wealth Factor Tilt				
	(1)	(2)	(3)	(4)	(5)
Wealth Percentile Dummies					
Bottom 10%	-0.048***	-0.046***	-0.042***	-0.048***	-0.043***
	(0.006)	(0.005)	(0.010)	(0.005)	(0.010)
10-20	-0.026***	-0.025***	-0.020***	-0.026***	-0.022***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
20-30	-0.018***	-0.018***	-0.014***	-0.019***	-0.016***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)
30-40	-0.010***	-0.011***	-0.008***	-0.010***	-0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
40-50	-0.004	-0.004	-0.004	-0.003	-0.002
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
60-70	-0.003	-0.003	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
70-80	-0.004	-0.004	-0.002	-0.003	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
80-90	0.004	0.003	0.007**	0.004	0.007**
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
90-99	0.030***	0.029***	0.035***	0.028***	0.034***
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
99-99.9	0.084***	0.083***	0.092***	0.078***	0.086***
	(0.011)	(0.011)	(0.009)	(0.013)	(0.010)
Top 1%	0.167***	0.166***	0.175***	0.163***	0.170***
Year FE	Yes	Yes	Yes	Yes	Yes
Age group FE	No	Yes	Yes	Yes	Yes
Number of observations	943,457	943,457	943,457	880,319	880,319
Adjusted R^2	0.047	0.047	0.049	0.048	0.049

Table IX
Firm Characteristics of Investor Factors

This table reports the median characteristics of the stock portfolios held by Norwegian individual investors sorted by age (columns 1 to 4) or wealth (columns 5 to 8). The analysis is based on the panel of monthly returns on Norwegian stocks. For each characteristic, the L portfolio correspond to investors in the bottom 30% of the characteristic's distribution, the M portfolio to investors in the mid 40%, and the H portfolio to investors in the top 30%. The H-L portfolio is long the H portfolio and short the L portfolio. Years in sample refer to the number of years the stock is in our panel. The share of institutional ownership is measured in percent. Volatility is the square root of the realized variance of daily returns measured over the previous 12 months. Monthly turnover is defined as the average daily trading volume multiplied by 30 and divided by the free-float-adjusted market valuation. CAPM beta is estimated by a 60-month rolling-window regression of excess returns on the market factor. Size is the market value of equity reported in million NOK. BE/ME is the ratio of the book value of equity to the market value of equity. Profitability is the ratio of gross profit (the difference between total revenue and cost of goods sold) to total assets. Investment growth is measured by the growth rate in total assets.

	Age-Sorted Portfolios				Wealth-Sorted Portfolios			
	L	M	H	H-L	L	M	H	H-L
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years in sample	8.00	10.00	13.00	5.00	7.00	9.00	16.00	9.00
Institutional ownership share (%)	3.10	6.40	6.40	3.36	5.10	6.40	4.40	-0.67
Turnover (% per month)	7.23	1.65	0.56	-6.67	5.26	2.17	0.38	-4.88
Volatility	0.25	0.13	0.09	-0.16	0.24	0.14	0.08	-0.16
CAPM beta	0.88	0.83	0.67	-0.22	0.94	0.84	0.66	-0.28
Size (million NOK)	384	1342	1485	1102	508	1103	2118	1610
BE/ME	0.72	0.70	0.68	-0.05	0.55	0.65	0.89	0.34
Profitability (%)	0.03	0.07	0.09	0.06	0.05	0.07	0.08	0.03
Investment growth (%)	0.04	0.07	0.09	0.05	0.09	0.07	0.07	-0.02

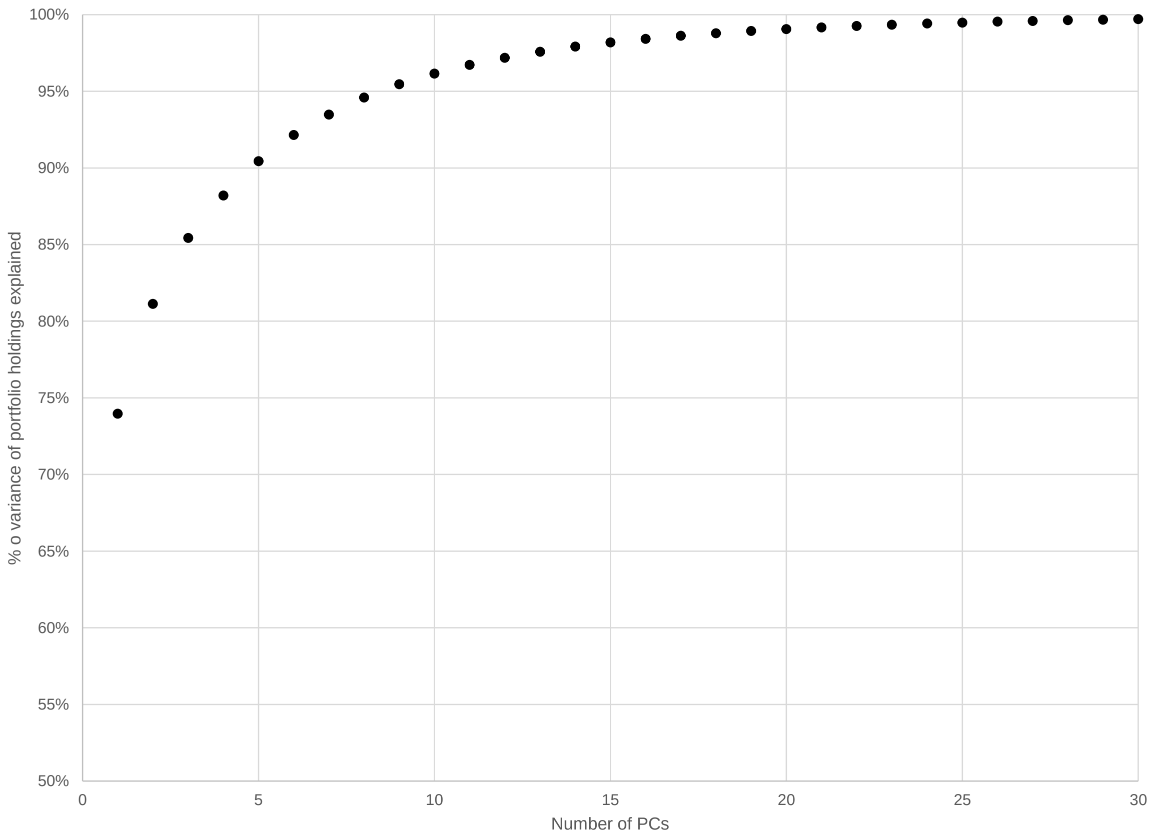


Figure 1 Factor Structure of Investor Portfolio. This figure plots the cumulative proportion of the cross-sectional variance of aggregate portfolio holdings explained by principal components (PCs). The cumulative proportion is calculated each year in 2002-2017 and then averaged across years. The principal component analysis is based on 93 portfolios containing the holdings of individual investors grouped by age, wealth, and other socioeconomic characteristics.

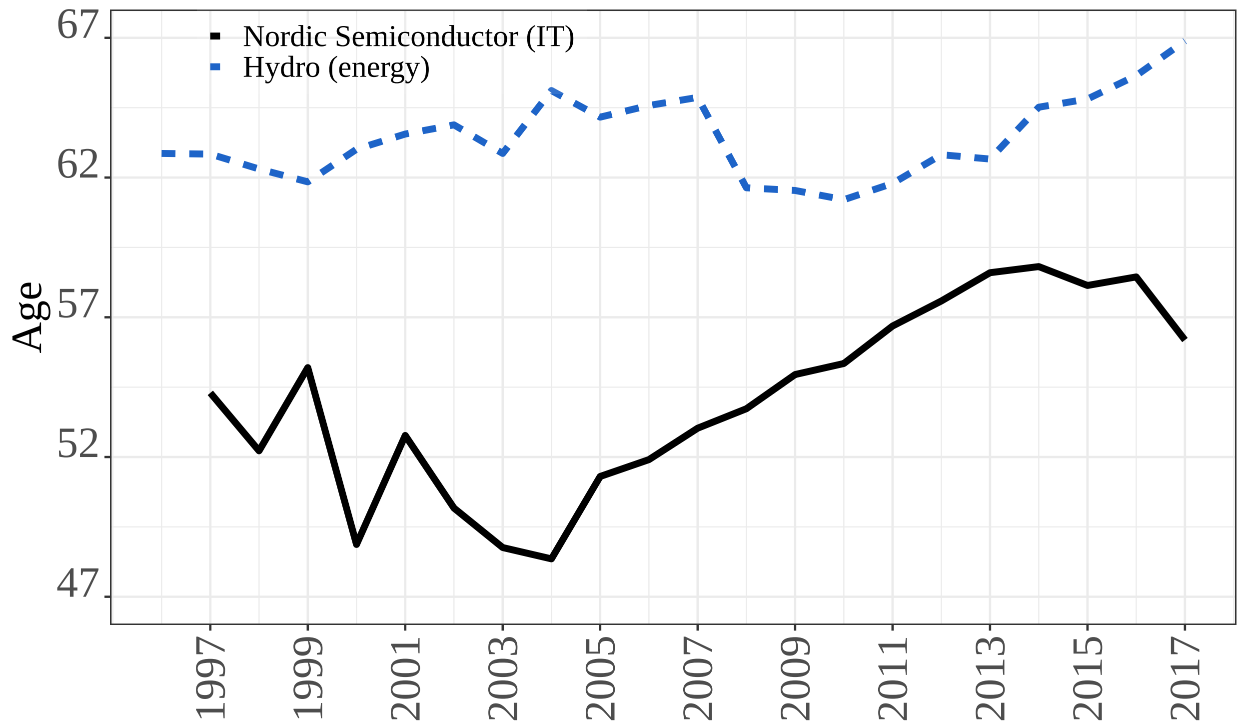
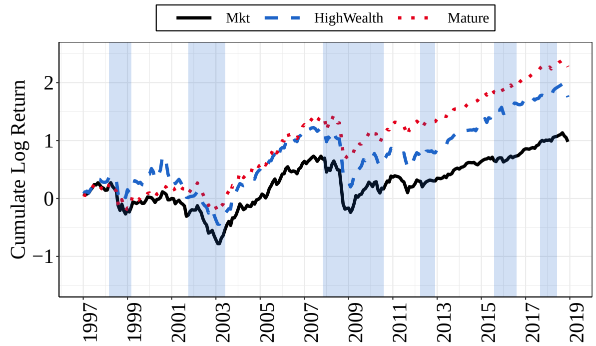
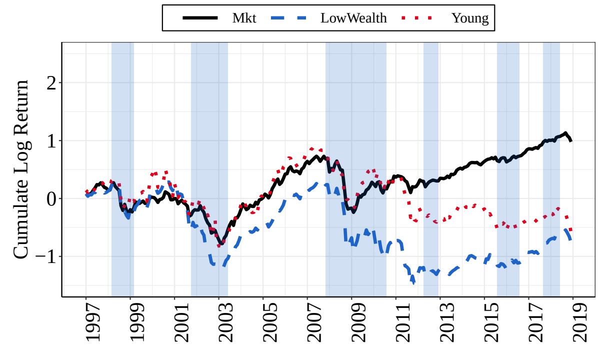


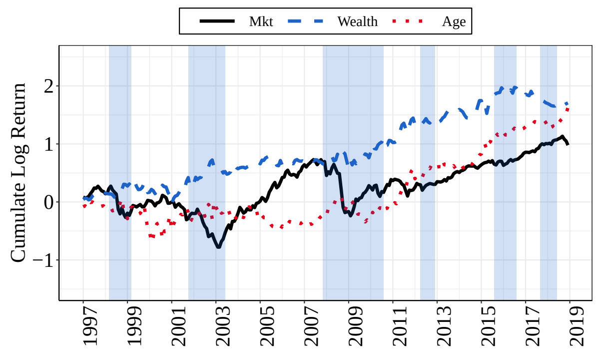
Figure 2 Investor age characteristic of two stocks. This figure plots the age characteristic of Norsk Hydro and Nordic Semiconductor in 1997-2018. Hydro is a fully integrated aluminium company. Nordic Semiconductor is a semiconductor company specializing in wireless technology. For each stock, the age characteristic is calculated as the average age of individual investors who directly own the stock, weighted by the number of shares that each investor holds at the beginning of the year.



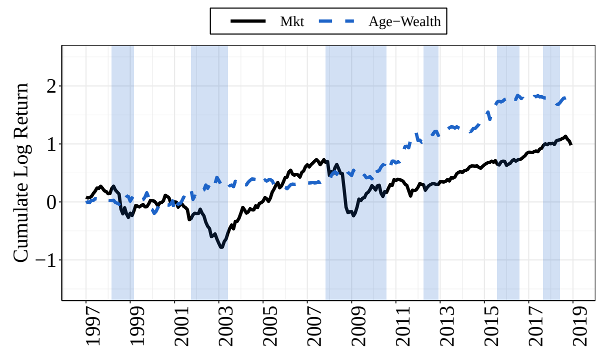
(a) High wealth and mature portfolios



(b) Low wealth and young portfolios

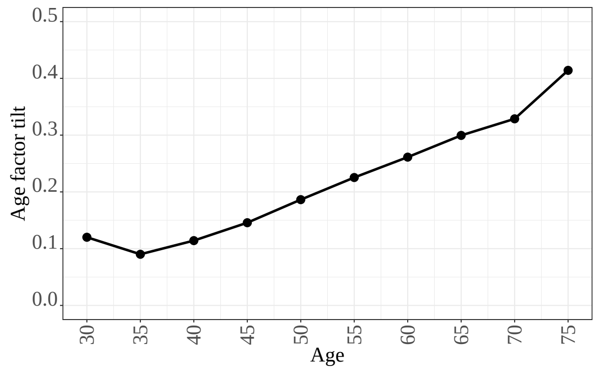


(c) Age and wealth factors

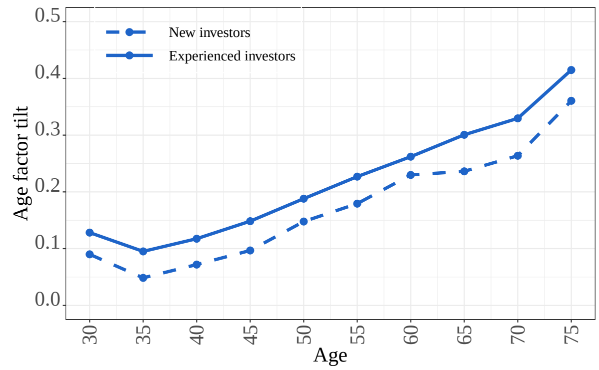


(d) Combined age-wealth factor

Figure 3 Cumulative return on investor factors. This figure plots the log cumulative return on portfolios of Norwegian stocks sorted by investor age and wealth characteristics in 1997-2018. Panel A plots historical returns on the mature portfolio and on the high-wealth portfolio. Panel B plots historical returns on the young portfolio and on the low-wealth portfolio. Panel C plots the age factor (mature-minus-young) and wealth factor (high wealth-minus-low wealth) portfolios, and Panel D plots the combined age-wealth factor, AW_t. In each panel, the black line represents the performance of the market portfolio and the blue shades indicate economic recessions. The portfolios are constructed as follows. We sort stocks by the age characteristic. We define the young portfolio as the value-weighted portfolio of stocks in the bottom 30%, and the mature portfolio as the value-weighted portfolio of stocks in the top 30% of the age distribution. We similarly define the high-wealth and low-wealth portfolios by sorting stocks according to the net worth of their investors.

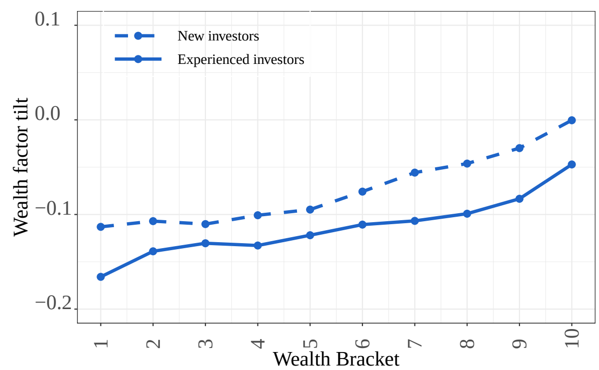


(a) Age tilt: All participating investors



(b) Age tilt: New and existing participants

(c) Wealth tilt: All participating investors



(d) Wealth tilt: New and existing participants

Figure 4 Factor tilts across investor groups. This figure plots the average tilts of the stock portfolios held by investors in different age and wealth groups. The analysis is based on the panel of Norwegian individual investors who hold stocks directly during the 1997-2018 period. Panel A plots the average age tilt across 10 age groups. Panel B plots the average age tilt of new participants (dotted) and preexisting participants (solid) each year. Panel C plots the average wealth tilt of individual investors across 12 different wealth groups. Panel D plots the average wealth tilt of new participants (dotted) and preexisting participants (solid) each year. Averages are equally-weighted. New participants are investors with less than one year of experience with direct stock investing, while preexisting participants have at least one year of experience.

Internet Appendix for "What Do the Portfolios of Individual Investors Reveal About the Cross-Section of Equity Returns?"

Sebastien Betermier, Laurent E. Calvet, Samuli Knüpfer, and Jens Kvaerner*

February 2022

This Internet Appendix provides the proofs of the main theoretical results, describes the details of data construction, discusses the empirical methodology, and presents additional empirical evidence.

*Betermier: Desautels Faculty of Management, McGill University, 1001 Sherbrooke St West, Montreal, QC H3A 1G5, Canada; sebastien.betermier@mcgill.ca. Calvet: EDHEC Business School, 16 rue du Quatre-Septembre, 75002 Paris, France, and CEPR; laurent.calvet@edhec.edu. Knüpfer, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway; samuli.knupfer@bi.no. Kvaerner: Tilburg University, Warandelaan 2 5037 AB Tilburg Netherlands; jkverner@gmail.com.

Contents

I. Appendix to Section I: Theoretical Linkages Between Investor Portfolios and Pricing Factors

I.A. *Proof of Proposition 1*

In Section I.A of the main text, we show that the market portfolio satisfies

$$\boldsymbol{m} = \boldsymbol{\tau} + \sum_{k=1}^{K} \eta_k^m \, \boldsymbol{d_k}, \tag{IA-1}$$

where $\eta_k^m = \sum_{i=1}^{I} E^i \eta_k^i / \sum_{i=1}^{I} E_i$ is the aggregate loading on the k^{th} deviation portfolio.

We now derive a multi-beta asset pricing equation, which shows that the model has $K+1$ pricing factors: the market $\boldsymbol{m}$ and the K deviation portfolios $\boldsymbol{d_k}$. Let μ_m denote the expected return on the market portfolio and let $\boldsymbol{\mu_d}$ denote the vector of expected returns on the deviation portfolios.

By (IA-1) and the definition of the tangency portfolio, the vector of excess returns satisfies

$$\boldsymbol{\mu} - R_f \mathbf{1} = \phi \, \boldsymbol{\Sigma} \boldsymbol{\tau}. \tag{IA-2}$$

The normalizing constant ϕ is the performance measure

$$\phi = \mathbf{1}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_f \mathbf{1}) = \frac{\mu_\tau - R_f}{\sigma_\tau^2}, \tag{IA-3}$$

where $\mu_\tau = \boldsymbol{\mu}' \boldsymbol{\tau}$ and $\sigma_\tau^2 = \boldsymbol{\tau}' \boldsymbol{\Sigma} \boldsymbol{\tau}$ denote, respectively, the drift and variance of the tangency portfolio.[1]

[1] The result can be derived as follows. The tangency portfolio has excess drift

$$\mu_\tau - R_f = (\boldsymbol{\mu} - R_f \mathbf{1})' \boldsymbol{\tau} = \frac{(\boldsymbol{\mu} - R_f \mathbf{1})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_f \mathbf{1})}{\mathbf{1}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_f \mathbf{1})}.$$

We use (IA-1) to substitute out the tangency portfolio,

$$\boldsymbol{\mu} - R_f \mathbf{1} = \phi \boldsymbol{\Sigma} \left(\boldsymbol{m} - \sum_{k=1}^{K} \eta_k^m \boldsymbol{d_k} \right). \tag{IA-4}$$

Let $\boldsymbol{\Sigma_{jm}} = \boldsymbol{\Sigma m}$ denote the $J \times 1$ covariance vector between the J assets and the market portfolio, $\boldsymbol{\Sigma_{jd}} = [\boldsymbol{\Sigma}\, \boldsymbol{d_1}, \cdots, \boldsymbol{\Sigma}\, \boldsymbol{d_K}]$ the $J \times K$ covariance matrix between the assets and the long-short portfolios, and $\boldsymbol{\Sigma_{j,md}} = [\boldsymbol{\Sigma_{jm}}, \boldsymbol{\Sigma_{jd}}]$. We can then write equation (IA-4) in vector form,

$$\boldsymbol{\mu} - R_f \mathbf{1} = \phi\, \boldsymbol{\Sigma_{j,md}} \begin{pmatrix} 1 \\ -\boldsymbol{\eta^m} \end{pmatrix}, \tag{IA-5}$$

where $\boldsymbol{\eta^m} = (\eta_1^m, \ldots, \eta_K^m)'$.

From (IA-4), we can express the risk premium on the market $\boldsymbol{m}$ and the portfolios $\boldsymbol{d_k}$ as

$$\begin{pmatrix} \mu_m - R_f \\ \boldsymbol{\mu_d} - R_f \mathbf{1} \end{pmatrix} = \phi\, \boldsymbol{\Sigma_{md,md}} \begin{pmatrix} 1 \\ -\boldsymbol{\eta^m} \end{pmatrix}, \tag{IA-6}$$

where $\boldsymbol{\Sigma_{md,md}}$ is the $(K+1) \times (K+1)$ covariance matrix of the market and deviation portfolio returns. We infer that

$$\phi \begin{pmatrix} 1 \\ -\boldsymbol{\eta^m} \end{pmatrix} = \boldsymbol{\Sigma^{-1}_{md,md}} \begin{pmatrix} \mu_m - R_f \\ \boldsymbol{\mu_d} - R_f \mathbf{1} \end{pmatrix}. \tag{IA-7}$$

The instantaneous variance of the tangency portfolio, $\sigma_\tau^2 = \boldsymbol{\tau}' \boldsymbol{\Sigma} \boldsymbol{\tau}$, therefore satisfies

$$\sigma_\tau^2 = \frac{(\boldsymbol{\mu} - R_f \mathbf{1})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_f \mathbf{1})}{[\mathbf{1}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_f \mathbf{1})]^2} = \frac{\mu_\tau - R_f}{\mathbf{1}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - R_f \mathbf{1})},$$

which implies that (IA-3) holds.

We plug (IA-7) into (IA-5) and obtain

$$\boldsymbol{\mu} - R_f \mathbf{1} = \boldsymbol{\beta}'_{\boldsymbol{j,md}} \begin{pmatrix} \mu_m - R_f \\ \boldsymbol{\mu_d} - R_f \mathbf{1} \end{pmatrix}, \tag{IA-8}$$

where $\boldsymbol{\beta_{j,md}} = \boldsymbol{\Sigma}^{-1}_{\boldsymbol{md,md}} \boldsymbol{\Sigma}'_{\boldsymbol{j,md}}$. We let $\boldsymbol{\beta_{j,md}} = [\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_J}]$ and conclude that the Proposition holds.

I.B. CAPM Alpha and Beta

We can re-write (IA-4) as follows:

$$\boldsymbol{\mu} - R_f \, \mathbf{1} = \phi \left(\sigma_m^2 \, \boldsymbol{b_m} - \sum_{k=1}^{K} \eta_k^m \, \sigma_k^2 \, \boldsymbol{b_k} \right), \tag{IA-9}$$

where $\boldsymbol{b_m} = (b_{j,m})$ is the $J \times 1$ vector of the stocks' univariate betas to the market factor, and $\boldsymbol{b_k} = (b_{j,k})$ is the $J \times 1$ vector of the stocks' univariate betas to the k^{th} deviation portfolio.

The risk premium of the market portfolio is equal to

$$\mu_m - R_f = \phi \left(\sigma_m^2 - \sum_{k=1}^{K} \eta_k^m \, \sigma_k^2 \, b_{m,k} \right), \tag{IA-10}$$

where $b_{m,k}$ is the market portfolio's univariate beta with respect to the k^{th} deviation portfolio.

A stock's CAPM-alpha is defined as $\alpha_j = \mu_j - R_f - b_{j,m}(\mu_m - R_f)$, where $b_{m,j}$ is the stock's univariate beta to the market portfolio. Combining (IA-9) and (IA-10) into the stock's alpha definition yields

$$\boldsymbol{\alpha} = -\phi \sum_{k=1}^{K} \eta_k^m \, \sigma_k^2 \, (\boldsymbol{b_k} - b_{m,k} \, \boldsymbol{b_m}). \tag{IA-11}$$

The alpha coefficient of stock j is therefore $\alpha_j = -\phi \sum_{k=1}^{K} \eta_k^m \, \sigma_k^2 \, (b_{j,k} - b_{j,m} \, b_{m,k})$, as we state in the main text.

Likewise, equation (IA-1) implies that the vector of market betas, $\boldsymbol{b_m} = \boldsymbol{\Sigma}\,\boldsymbol{m}/\sigma_m^2$, can be decomposed as follows:

$$\boldsymbol{b_m} = \frac{\sigma_\tau^2}{\sigma_m^2}\boldsymbol{b_\tau} + \sum_{k=1}^{K} \eta_k^m \, \frac{\sigma_k^2}{\sigma_m^2} \, \boldsymbol{b_k}. \tag{IA-12}$$

Stock j's beta to the market,

$$b_{j,m} = \frac{\sigma_\tau^2}{\sigma_m^2} b_{j,\tau} + \sum_{k=1}^{K} \eta_k^m \, \frac{\sigma_k^2}{\sigma_m^2} \, b_{j,k}, \tag{IA-13}$$

is the linear combination of its beta to the tangency portfolio, $b_{j,\tau}$, and its betas to the deviation portfolios, $b_{j,k}$. The linear coefficients are driven by the ratio of the tangency portfolio variance to the market portfolio variance, σ_τ^2/σ_m^2, the ratio of factor variances to the market variance, σ_k^2/σ_m^2, and the aggregate factor tilts, η_k^m.

I.C. Individual Portfolio Choice

In this Section, we compute the consumption-portfolio decision of an agent satisfying the specification outlined in Section I.C of the main text. We focus on the decision of a single agent and simplify notation by dropping the agent index i. We also asume without loss of generality that the agent is born at date 0.

The consumption-portfolio decision problem is defined as follows. The agent lives and consumes in periods $t = 0, \ldots, T$. At the beginning of every period t, she receives stochastic labor income L_t. The cash on hand available to her, W_t, is the sum of labor income, L_t, and the value at date t of previous investments in financial assets.

The agent uses cash on hand to consume C_t and invests the remainder, $W_t - C_t$, in financial assets. She can trade a riskless asset with net rate of return R_f and the stocks $j \in \{1, \ldots, J\}$. The stocks have excess returns $R^e_{j,t+1}$ between dates t, which we stack into the column vector $\boldsymbol{R^e_{t+1}}$. Let s_t denote the proportion of the agent's wealth invested in risky assets and $\boldsymbol{\omega_t} \in \mathbb{R}^J$, $\mathbf{1}'\boldsymbol{\omega_t} = 1$, denote the weights of stocks in the agent's risky portfolio.

With this notation, the cash on hand available to the agent is

$$W_{t+1} = L_{t+1} + (W_t - C_t)\left(1 + R_f + s_t \sum_{j=1}^{J} \omega_{j,t}\, R^e_{j,t+1}\right) \qquad \text{(IA-14)}$$

at the beginning of period $t+1$.

The agent selects the consumption-risky share-portfolio plan $\{(s_t, \boldsymbol{\omega_t}, C_t)\}$ that maximizes the lifetime utility

$$\mathbb{E}_0\left[\sum_{t=0}^{T} \beta^t\, u(C_t)\right], \qquad \text{(IA-15)}$$

subject to the budget constraint (IA-14).

We make the following assumptions on the stochastic processes driving the economy. Labor income is given by:

$$L_t = L_{t-1}\, g_{L,t}, \qquad \text{(IA-16)}$$

where the growth rates $g_{L,t}$ are independent and identically distributed. We also assume that there exists a state vector $\boldsymbol{y_t}$ that drives returns. The growth rate $g_{L,t}$ and the state vector $\boldsymbol{y_t}$ are known to the agent at the beginning of period t. Conditional on $\boldsymbol{y_t}$, the distribution of $(R^e_{1,t+1}, \ldots, R^e_{J,t+1})'$ has mean $\boldsymbol{\mu_t}$ and variance-covariance matrix $\boldsymbol{\Sigma_t}$. Let $R_{j,t+1} = R_f + R^e_{j,t+1}$ denote the return on every stock j. We assume that $(g_{L,t+1}, R_{1,t+1}, \ldots, R_{J,t+1})'$ is multivariate lognormal, which will be useful in Section I.D of this Internet Appendix.

Let $J(t, W_t, L_t, \boldsymbol{y_t})$ denote the value function of the agent at t. The value function satisfies the Bellman equation

$$J(t, W_t, L_t, \boldsymbol{y_t}) = \max_{\{s_t, \boldsymbol{\omega_t}, C_t\}} \left[u(C_t) + \delta\, \mathbb{E}_t J(t+1, W_{t+1}, L_{t+1}, \boldsymbol{y_{t+1}})\right]$$

At the terminal date T, the value function is $J(T, W_T, L_T, \boldsymbol{y_T}) = u(W_T)$.

The optimal portfolio is a linear combination of the tangency portfolio, a portfolio providing a hedge against labor income, and a portfolio providing a hedge against time-varying investment opportunities, as the following proposition shows.

Proposition IA.1. *The vector of stock weights in the agent's total wealth, $s_t \boldsymbol{\omega_t}$, is approximately given by*

$$s_t \boldsymbol{\omega_t} = -\frac{J_W}{(W_t - C_t) J_{WW}} \boldsymbol{\Sigma_t^{-1}} (\boldsymbol{\mu_t} - R_f \mathbf{1}) - \frac{L_t}{W_t - C_t} \left(1 + \frac{J_{WL}}{J_{WW}}\right) \boldsymbol{\Sigma_t^{-1}} \boldsymbol{b_t} - \frac{\boldsymbol{\Sigma_t^{-1}} \boldsymbol{D_t} \boldsymbol{J_{Wy}}}{(W_t - C_t) J_{WW}},$$

where the derivatives of the value function are evaluated at $(\mathbb{E}_t W_{t+1}, \mathbb{E}_t L_{t+1}, \boldsymbol{\mathbb{E}_t y_{t+1}})$,

$$\begin{aligned} \boldsymbol{b_t} &= \mathbb{E}_t[(g_{L,t+1} - \mathbb{E}_t g_{L,t+1}) \boldsymbol{R^e_{t+1}}], \\ \boldsymbol{D_t} &= \mathbb{E}_t[\boldsymbol{R^e_{t+1}} (\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}})'], \end{aligned}$$

for every $t < T$.

Proof. We note that wealth,

$$W_{t+1} = L_t \, g_{L,t+1} + (W_t - C_t) \left(1 + R_f + s_t \sum_{j=1}^{J} \omega_{j,t} \, R^e_{j,t+1}\right), \qquad \text{(IA-17)}$$

satisfies the following moment identities:

$$\mathbb{E}_t(W_{t+1}) = L_t \mu_L + (W_t - C_t) \left[1 + R_f + s_t \boldsymbol{\omega_t'} (\boldsymbol{\mu_t} - R_f \mathbf{1})\right], \qquad \text{(IA-18)}$$

$$\mathrm{Var}_t(W_{t+1}) = L_t^2 \, \sigma_L^2 + (W_t - C_t)^2 \, s_t^2 \, \boldsymbol{\omega_t' \Sigma \omega_t} + 2(W_t - C_t) \, L_t \, s_t \, \boldsymbol{\omega_t'} \, \boldsymbol{b_t}, \qquad \text{(IA-19)}$$

$$\mathrm{Cov}_t(W_{t+1}, L_{t+1}) = L_t^2 \, \sigma_L^2 + L_t \, (W_t - C_t) \, s_t \, \boldsymbol{\omega_t' b_t}, \qquad \text{(IA-20)}$$

$$\mathbb{E}_t[(W_{t+1} - \mathbb{E}_t W_{t+1}) \, (\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}})'] = L_t \, \boldsymbol{f_t} + (W_t - C_t) \, s_t \, \boldsymbol{\omega_t' D_t}, \qquad \text{(IA-21)}$$

where $\mu_L = \mathbb{E}_t(g_{L,t+1})$, $\sigma_L^2 = \mathrm{Var}_t(g_{L,t+1})$, and $\boldsymbol{f_t} = \mathbb{E}_t[g_{L,t+1} \, (\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}})']$.

We consider a quadratic expansion of $J(t+1, W_{t+1}, L_{t+1}, \boldsymbol{y_{t+1}})$ around the conditional mean $(\mathbb{E}_t W_{t+1}, \mathbb{E}_t L_{t+1}, \boldsymbol{\mathbb{E}_t y_{t+1}})$. We obtain

$$\begin{aligned} J(t+1, W_{t+1}, L_{t+1}, \boldsymbol{y_{t+1}}) \;\approx\; & J(t+1, \mathbb{E}_t W_{t+1}, \mathbb{E}_t L_{t+1}, \boldsymbol{\mathbb{E}_t y_{t+1}}) + \frac{1}{2}(W_{t+1} - \mathbb{E}_t W_{t+1})^2 J_{WW} \\ & + J_{WL} \left(W_{t+1} - \mathbb{E}_t W_{t+1}\right) \left(L_{t+1} - \mathbb{E}_t L_{t+1}\right) \\ & + \boldsymbol{J'_{Wy}} \left(\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}}\right) \left(W_{t+1} - \mathbb{E}_t W_{t+1}\right) + Q_{t+1} + \varepsilon_{t+1}, \end{aligned}$$

where $\varepsilon_{t+1} = J_W \left(W_{t+1} - \mathbb{E}_t W_{t+1}\right) + J_L \left(L_{t+1} - \mathbb{E}_t L_{t+1}\right) + \boldsymbol{J'_y \left(y_{t+1} - \mathbb{E}_t y_{t+1}\right)}$, and

$$\begin{aligned} Q_{t+1} \;=\; & \frac{1}{2}(L_{t+1} - \mathbb{E}_t L_{t+1})^2 J_{LL} + \frac{1}{2}(\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}})' \boldsymbol{J_{yy}} \left(\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}}\right) \\ & + \boldsymbol{J'_{Ly}} \left(\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}}\right) \left(L_{t+1} - \mathbb{E}_t L_{t+1}\right). \end{aligned}$$

We note that $\mathbb{E}_t \varepsilon_{t+1} = 0$.

The conditional expectation at t of the value function $J(t+1, W_{t+1}, L_{t+1}, \boldsymbol{y_{t+1}})$ satisfies

$$\begin{aligned} \mathbb{E}_t J(t+1, W_{t+1}, L_{t+1}, \boldsymbol{y_{t+1}}) \;\approx\; & J(t+1, \mathbb{E}_t W_{t+1}, \mathbb{E}_t L_{t+1}, \boldsymbol{\mathbb{E}_t y_{t+1}}) + \frac{1}{2} Var_t(W_{t+1})\, J_{WW} \\ & + J_{WL}\, \mathrm{Cov}_t(W_{t+1}, L_{t+1}) \\ & + \boldsymbol{J'_{Wy}}\, \mathbb{E}_t[(\boldsymbol{y_{t+1}} - \boldsymbol{\mathbb{E}_t y_{t+1}}) \left(W_{t+1} - \mathbb{E}_t W_{t+1}\right)] + \mathbb{E}_t Q_{t+1}. \end{aligned}$$

We infer that

$$\begin{aligned} \mathbb{E}_t J(t+1, W_{t+1}, L_{t+1}, \boldsymbol{y_{t+1}}) \;\approx\; & J(t+1, \mathbb{E}_t W_{t+1}, \mathbb{E}_t L_{t+1}, \boldsymbol{\mathbb{E}_t y_{t+1}}) \\ & + \frac{1}{2}[L_t^2\, \sigma_L^2 + (W_t - C_t)^2\, s_t^2\, \boldsymbol{\omega_t' \Sigma_t \omega_t} + 2(W_t - C_t) L_t\, s_t\, \boldsymbol{\omega_t' b_t}] J_{WW} \\ & + J_{WL} \left[L_t^2\, \sigma_L^2 + L_t \left(W_t - C_t\right) s_t\, \boldsymbol{\omega_t' b_t}\right] \\ & + \boldsymbol{J'_{Wy}} \left[L_t\, \boldsymbol{f_t} + \left(W_t - C_t\right) s_t\, \boldsymbol{D_t' \omega_t}\right] + \mathbb{E}_t Q_{t+1}. \end{aligned}$$

The first-order condition with respect to $\boldsymbol{\omega_t}$ is therefore[2]

$$(\boldsymbol{\mu_t} - R_f\mathbf{1})J_W + [(W_t - C_t)s_t\,\boldsymbol{\Sigma_t}\,\boldsymbol{\omega_t} + L_t\,\boldsymbol{b_t}]J_{WW} + J_{WL}\,L_t\,\boldsymbol{b_t} + \boldsymbol{D_t}\,\boldsymbol{J_{Wy}} = 0,$$

or equivalently

$$(\boldsymbol{\mu_t} - R_f\mathbf{1})J_W + (W_t - C_t)\,J_{WW}\,s_t\,\boldsymbol{\Sigma_t}\,\boldsymbol{\omega_t} + L_t\,(J_{WW} + J_{WL})\,\boldsymbol{b_t} + \boldsymbol{D_t}\,\boldsymbol{J_{Wy}} = 0.$$

We conclude that the proposition holds. ∎

The expression for the optimal portfolio considerably simplifies in the last trading period $t = T-1$, which is a cornerstone of the analysis. Indeed, since the value function is $J(T, W_T, L_T, \boldsymbol{y_T}) = u(W_T)$ at the terminal date, its partial derivatives satisfy $J_{WL} = 0$, $\boldsymbol{J_{Wy}} = 0$ at $(T, \mathbb{E}_{T-1}W_T, \mathbb{E}_{T-1}L_T, \boldsymbol{\mathbb{E}_{T-1}y_T})$. The optimal portfolio at date $T-1$ reduces to

$$s_{T-1}\,\boldsymbol{\omega_{T-1}} = -\frac{J_W}{(W_{T-1} - C_{T-1})J_{WW}}\boldsymbol{\Sigma_{T-1}^{-1}}(\boldsymbol{\mu_{T-1}} - R_f\mathbf{1}) - \frac{L_{T-1}}{W_{T-1} - C_{T-1}}\boldsymbol{\Sigma_{T-1}^{-1}}\,\boldsymbol{b_{T-1}}.$$

Since the agent stops trading at date $T-1$, the hedging demand against adverse variation in future investment opportunities drops out from the optimal portfolio. Furthermore, if $L_{T-1} = 0$, the agent does not need to hedge against labor income shocks and the optimal portfolio of stocks in the financial portfolio becomes proportional to the tangency portfolio: $s_{T-1}\,\boldsymbol{\omega_{T-1}} = -(W_{T-1} - C_{T-1})^{-1}\,J_{WW}^{-1}\,J_W\,\boldsymbol{\Sigma_{T-1}^{-1}}(\boldsymbol{\mu_{T-1}} - R_f\mathbf{1})$. The optimal portfolio of stocks in the agent's stock portfolio, $\boldsymbol{\omega_{T-1}}$ is then equal to the tangency porfolio:

$$\boldsymbol{\omega_{T-1}} = \boldsymbol{\tau_{T-1}}.$$

A mature investor with a large financial wealth-to-labor income ratio therefore holds the mean-variance efficient portfolio $\boldsymbol{\tau_{T-1}}$ and has a maximal Sharpe ratio.

[2] We neglect the higher-order terms involving the derivatives of $\mathbb{E}_t(Q_{t+1})$ with respect to $\boldsymbol{\omega_t}$.

If the agent has constant relative risk aversion (CRRA), $u(c) = C^{1-\gamma}/(1-\gamma)$, the optimization problem is homogenous in (W_t, L_t). Let A_t denote the age of the agent at date t. The optimal portfolio of stocks can then be written as

$$\boldsymbol{\omega_t} = \boldsymbol{\omega}\left(A_t, \frac{L_t}{W_t}, \boldsymbol{y_t}\right). \tag{IA-22}$$

We know in addition that $\boldsymbol{\omega}(T-1, 0, \boldsymbol{y_{T-1}}) = \boldsymbol{\tau_{T-1}}$ in the agent's last period of trading.

I.D. Linearization and Factor Structure

I.D.1. Cross-Section of Investors

We now consider a cross-section of investors $i = 1, \ldots, I$ at date t. As is explained in the main text, each investor i is born at date b^i and dies at date $b^i + T$. Investors have the same CRRA utility, the same lifespan T and are exposed to the same labor income growth $g_{L,t}$. These restrictions could be lifted in future work, but the current setup is parsimonious and sufficient to account for the main empirical regularities reported in the main text.

The stock portfolio of each investor i in period t is given by:

$$\boldsymbol{\omega_t^i} = \boldsymbol{\omega}\left(A_t^i, \frac{L_t^i}{W_t^i}, \boldsymbol{y_t}\right),$$

as equation (IA-22) implies. We linearize $\boldsymbol{\omega}(\,\cdot\,,\,\cdot\,, \boldsymbol{y_t})$ around $A_t^i = T-1$ and $L_t^i/W_t^i = 0$:

$$\boldsymbol{\omega_t^i} = \boldsymbol{\tau_t} + (T-1-A_t^i)\,\boldsymbol{d_{1,t}} + \frac{L_t^i}{W_t^i}\,\boldsymbol{d_{2,t}}.$$

We now explain how this linearization can be achieved.

I.D.2. Linearization of the Optimal Portfolio

We focus on the linearization of the optimal portfolio function. Without loss of generality, we focus on a single agent and drop the index i for notational convenience.

The linearization is based on two principles. First, the agent trades frequently between the start and the end of her trading life, that is between dates 0 and $T-1$ for an agent born at date 0. The behavior of the agent between 0 and $T-1$ can therefore be approximated by a continuous-time model when the trading frequency is sufficiently high. Second, the agent stops trading in the last period of her life, which has a fixed length equal to one unit of time. This second condition is technically useful because it implies that the agent remains sensitive to income and wealth at date $T-1$.[3] The corresponding utility function can therefore be written as

$$\sum_{l=0}^{h\,(T-1)} \beta^{l/h} u(C_s) + \beta^T u(C_T),$$

where h denotes the number of trading periods per unit of time.

The auxiliary continuous-time economy is defined as follows. We consider an agent with lifetime utility:

$$\mathbb{E}_0 \left[\int_0^{T-1} e^{-\rho\,(l-t)}\, u(c_s)\, ds + e^{-\rho\,(T-t)}\, u(c_T) \right]. \tag{IA-23}$$

The Bernoulli utility $u(\cdot)$ is same as in the discrete time model. The time discount rate is $\rho = -\ln(\beta)$, where β is the psychological discount factor of the discrete-time model.

The agent trades the riskless asset and the N stocks in continuous time between dates 0 and $T-1$. We assume for simplicity that the instantaneous riskless rate is $r = \ln(1 + R_f)$.

[3]If the length of the last period of life were allowed to converge to zero, the stock portfolio would converges to the tangency portfolio and the derivative of the optimal stock portfolio with respect to the labor income-to-wealth ratio would be zero in the limiting economy. A second order expansion of $\boldsymbol{\omega}$ would therefore be required to capture the link between the stock portfolio of wealth. Our assumption that the last investment period remains finite therefore allows us instead to rely on a more parsimonious, first-order Taylor expansion of $\boldsymbol{\omega_{T-1}}$.

The joint dynamics of the stock returns, labor income flow, and state vector $\boldsymbol{y_t}$ are given by:

$$\begin{aligned} \boldsymbol{dR_t} &= \boldsymbol{\mu(y_t)}\,dt + \boldsymbol{\sigma(y_t)}\,\boldsymbol{dz_t}, & \text{(IA-24)} \\ dL_t &= L_t\,(\mu_L\,dt + \boldsymbol{\sigma_L}\,\boldsymbol{dz_t}), & \text{(IA-25)} \\ \boldsymbol{dy_t} &= \boldsymbol{\mu_y(y_t)}\,dt + \boldsymbol{\sigma_y(y_t)}\,\boldsymbol{dz_t} & \text{(IA-26)} \end{aligned}$$

We assume that the dynamics of these variables in discrete time, as explained in Section I.C of this Internet Appendix, converge to their continuous-time equivalent, as given by (IA-24) to (IA-26), when the number of trading periods h goes to infinity. Since the convergence of discrete-time processes to Itô diffusions is the subject of a vast literature, we do not need to make specific parametric assumptions and only require that the convergence property holds.

We note that

$$\begin{aligned} \mathbb{E}_{\boldsymbol{t}}(\boldsymbol{dR_t}\,dL_t) &= \boldsymbol{b_t}\,dt, \\ \mathbb{E}_{\boldsymbol{t}}(\boldsymbol{dR_t}\,\boldsymbol{dy_t'}) &= \boldsymbol{D_t}\,dt, \\ \mathbb{E}_{\boldsymbol{t}}(\boldsymbol{dR_t}\,\boldsymbol{dR_t'}) &= \boldsymbol{\Sigma_t}\,dt, \end{aligned}$$

where $\boldsymbol{b_t} = \boldsymbol{\sigma(y_t)}\,\boldsymbol{\sigma_L'}$, $\boldsymbol{D_t} = \boldsymbol{\sigma(y_t)}\,\boldsymbol{\sigma_y(y_t)'}$, and $\boldsymbol{\Sigma_t} = \boldsymbol{\sigma(y_t)}\boldsymbol{\sigma(y_t)'}$.

Let W_t denote the wealth of the agent at t, c_t her consumption rate, s_t her risky share, and by $\boldsymbol{\omega_t}$ the weights of stocks in her stock portfolio. The budget constraint is:

$$dW_t = \left[L_t + W_t\,r + W_t\,s_t\,\boldsymbol{\omega_t'}(\boldsymbol{\mu} - r\mathbf{1}) - c_t\right]\,dt + W_t\,s_t\,\boldsymbol{\omega_t'}\boldsymbol{\sigma}\boldsymbol{dz_t}. \tag{IA-27}$$

The agent chooses $\{(s_t, \boldsymbol{\omega_t}, c_t)\}$ that maximize the lifetime utility (IA-23) under the budget constraint (IA-27). We denote by $V(t, L_t, W_t, \boldsymbol{y_t}; t)$ the corresponding value function.

Since the agent has CRRA utility, $u(C) = C^{1-\gamma}/(1-\gamma)$, the homogeneity of the problem implies that the optimal portfolio of the continuous-time problem can be written as:

$$\boldsymbol{\omega_t} = \boldsymbol{\omega}\left(A_t, L_t/W_t, \boldsymbol{y_t}\right). \tag{IA-28}$$

Furthermore, we know that $\boldsymbol{\omega_t}$ coincides with the portfolio of the discrete-time economy if $A_t = T-1$. In particular, $\boldsymbol{\omega}(T-1, 0, \boldsymbol{y_t}) = \boldsymbol{\tau}_t$. We linearize the function $\boldsymbol{\omega}(A_t, L_t/W_t, \boldsymbol{y_t})$ around the point $(T-1, 0, \boldsymbol{y_t})$ and obtain:

$$\boldsymbol{\omega}\left(t, L_t/W_t, \boldsymbol{y_t}\right) = \boldsymbol{\tau_t} + (T-1-t)\,\boldsymbol{d_{1,t}} + \frac{L_t}{W_t}\,\boldsymbol{d_{2,t}},$$

where

$$\boldsymbol{d_{1,t}} = -\frac{\partial \boldsymbol{\omega}}{\partial A}(T-1, 0, \boldsymbol{y_t}) \quad \text{and} \quad \boldsymbol{d_{2,t}} = \frac{\partial \boldsymbol{\omega}}{\partial L/W}(T-1, 0, \boldsymbol{y_t}).$$

The cross-section of portfolios can therefore be approximated by a three-factor model, in which the portfolio of each investor is the linear combination of the tangency portfolio, a portfolio linked to age, and a portfolio linked to wealth. Furthermore, Proposition IA.1 has a direct counterpart in continuous time.

Proposition IA.2. *The optimal portfolio is given by*

$$s_t\,\boldsymbol{\omega_t} = -\frac{V_W}{W_t\,V_{WW}}\boldsymbol{\Sigma_t^{-1}}(\boldsymbol{\mu_t} - r\mathbf{1}) - \frac{L_t}{W_t}\frac{V_{LW}}{V_{WW}}\boldsymbol{\Sigma_t^{-1}b_t} - \frac{\boldsymbol{\Sigma_t^{-1}\,D_t\,V_{Wy}}}{V_{WW}},$$

at every $t \le T-1$.

Proof. The optimal policy function satisfies the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned} 0 = \quad & \max_{s_t, \boldsymbol{\omega_t}, c_t} \quad [u(c_t)dt - \rho V\,dt + V_t\,dt + V_L\;\mathbb{E}_t(dL_t) + V_W\mathbb{E}_t(dW_t) \\ & + \boldsymbol{V_y'\mathbb{E}_t(dy_t)} + \tfrac{1}{2}V_{WW}\,(dW_t)^2 + \tfrac{1}{2}V_{LL}(dL_t)^2 + \tfrac{1}{2}\boldsymbol{dy_t'V_{yy}dy_t} \\ & + \boldsymbol{V_{Wy}'dy_t}\,dW_t + V_{LW}\,dL_t\,dW_t + \boldsymbol{V_{Ly}'dy_t}\,dL_t\big] \end{aligned}$$

or equivalently

$$
\begin{aligned}
0 = \quad \max_{s_t,\boldsymbol{\omega_t},c_t} \quad & [u(c_t) - \rho V + V_t + V_L \; L_t \; \mu_L + V_W \, [L_t + W_t \, r + W_t \, s_t \, \boldsymbol{\omega_t}'(\boldsymbol{\mu} - r) - c_t] \\
& + \boldsymbol{V_y'\mu_y(y_t)} + \tfrac{1}{2} V_{WW} \, W_t^2 \, s_t^2 \, \boldsymbol{\omega_t'\Sigma_t\omega_t} + \tfrac{1}{2} V_{LL} \; L_t^2 \, \|\boldsymbol{\sigma_L}\|^2 \\
& + \tfrac{1}{2} \mathrm{tr}(\boldsymbol{\sigma_y'} \, \boldsymbol{V_{yy'}} \, \boldsymbol{\sigma_y}) + W_t \, s_t \, \boldsymbol{\omega_t'\sigma\sigma_y'} \, \boldsymbol{V_{Wy}} + L_t \, W_t \, s_t \, \boldsymbol{\omega_t'\sigma\sigma_L'} V_{LW} \; + L_t \boldsymbol{V_{Ly}'\sigma_y\sigma_L'}].
\end{aligned}
$$

We write the first-order condition with respect to $\boldsymbol{\omega_t}$:

$$
V_W \, W_t \, (\boldsymbol{\mu} - r \, \mathbf{1}) + V_{WW} \, W_t^2 \, s_t \, \boldsymbol{\Sigma_t\omega_t} + W_t \boldsymbol{D_t} \, \boldsymbol{V_{Wy}} + L_t W_t \, \boldsymbol{b_t} \, V_{LW} = 0
$$

and conclude that the proposition holds. ■

II. Appendix to Section II: Data and Factor Structure of Portfolio Holdings

II.A. *Industry and Geographical Clusters*

In Section II.B of the main text, we create aggregate portfolio holdings for 93 groups of investors sorted by socioeconomic characteristics. Out of these 93 groups, 19 are based on investor occupation and 26 are based on the region of employment. We now explain how we form these groups.

Occupational Sectors. We assign a SIC industry code to each individual. In the following description, we explain the industry classification scheme used by Statistics Norway and describe the level of aggregation we use.

Our industry classification is based on the Norwegian Standard Industrial Classification (SIC2007), which has been used in Norway's official statistics since 2008. The first four digits of SIC2007 codes are identical to the EU's industrial classification, NACE Rev. 2.[4] As in other countries, SIC2007 codes are modified to incorporate a fifth digit that reflects local industrial conditions. Before 2008, the industry classification was based on NACE Rev. 1.1 (the 2002 update of NACE). We convert old classifications to new ones by using a linking table available on Statistics Norway's website, which ensures that our codes are consistent over time.

Our aggregation scheme involves two main steps. First, we follow the recommended standards in the National Accounts' second revision (A64 Rev. 2). However, in some cases Statistics Norway split some industries into two or more sub-industries. In these cases, we

[4]The description ias available at https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF

follow Statistics Norway. Overall, our procedure generates 70 industries, compared with 66 in A64 Rev. 2. We then aggregate the resulting industries into 20 main industries on the basis of our knowledge about the Norwegian economy and existing aggregations. An overview over this mapping process is provided in Table IA.1. One of these industries, "Household Services," includes fewer than 10,000 investors and is therefore excluded from the factor structure analysis.

Local Labor Markets. Following Butikofer et al. (2018) we aggregate municipalities (the lowest administrative level in Norway) into local labor markets based on commuting patterns. Details about the aggregation are in Bhuller (2009). The resulting 46 local labor markets cover the entire country, including urban and rural areas. A local labor market consists on average of nine municipalities and has an average population of 68,000 individuals. We restrict the sample to markets that have at least 10,000 investors in all years in our sample, which yields 26 local labor markets.

II.B. Age and Wealth Portfolio Regressions

In Section II.C of the main text, we run a pooled regression of the weight of stock j in the k^{th} principal component portfolio, $f^k_{j,t}$, on the stock's weight in the market, age, and wealth portfolio factors:

$$f^k_{j,t} = a^k + \lambda^k_{\text{MKT}}\, m_{j,t} + \lambda^k_{\text{AGE}}\, g_{\text{AGE},j,t} + \lambda^k_{\text{WEALTH}}\, g_{\text{WEALTH},j,t} + \epsilon^k_{j,t}. \tag{IA-29}$$

We now repeat the estimation when the left-hand side variables are the stock weights of the portfolios of the 93 investor groups. Specifically, we run a pooled OLS regression of the weight of stock j in group h's portfolio, $\omega^h_{j,t}$, on the weights of the market, age, and wealth portfolios:

$$\omega^h_{j,t} = a^h + \lambda^h_{\text{MKT}}\, m_{j,t} + \lambda^h_{\text{AGE}}\, g_{\text{AGE},j,t} + \lambda^h_{\text{WEALTH}}\, g_{\text{WEALTH},j,t} + \epsilon^h_{j,t}, \tag{IA-30}$$

for every period t, stock j, and group h.

Figure IA.1 reports the distribution of the R^2 coefficient and intercept across the 93 portfolios. These statistics are estimated both for a restricted version of (IA-30) that sets $\lambda^h_{\text{AGE}} = \lambda^h_{\text{WEALTH}} = 0$ (left plot on each panel) and for the unrestricted specification (right plot on each panel). In Panel A, the box plots illustrate the 25%, 50%, and 75% percentiles of the R^2 coefficient together with their minimum and maximum values. In Panel B, the box plots report the same statistics for the regression intercept.

Including the age and wealth factors into the estimation of (IA-30) significantly improves the fit for all portfolios. The interquartile range of R^2 jumps from 20%-40% in the restricted regression, which only considers the market portfolio, to 65%-86% in the full regression, which also includes the age and wealth factors. Additionally, the interquartile range of the intercept decreases from 30%-43% in the one-factor case to 9%-24% in the three-factor case. These findings confirm that the market, age, and wealth factors are able to explain the bulk of the common variation in investor portfolio holdings.

III. Appendix to Section III: Investor Pricing Factors

III.A. Firm Factors

Section III.A of the main text explains how we define Norwegian factors from firm characteristics. We now provide more details about the construction methodology and present statistics about the pricing performance of these factors.

The market portfolio is the value-weighted portfolio of all the stocks in the universe of Norwegian stocks defined in Section II.A. For the risk-free rate, we follow Oedegaard (2020) and use the monthly interbank rate, NIBOR, which we download from his website.[5]

The construction of firm factors is based on the following general rules. The momentum factor is rebalanced monthly. Other firm factors are rebalanced yearly. In the calculation of all factors, stocks are weighted by their free-float adjusted market weights (NOSHFF).[6] We now turn to the details of construction of each factor.

The size factor, SMB_t, is based on each firm's market value of equity in year t, which we obtain by multiplying the end-of-June closing price by the number of shares outstanding. Our share price variable corrects for stock splits, dividends, and other corporate events. If the closing price is unavailable, we proxy it by the bid or ask price. The long leg of the size factor contains stocks that are larger than the median, while the short leg contains stocks smaller than the median.

The value, profitability, and investment factors are defined from sorted portfolios as follows. We rank stocks by the selected characteristic (book-to-market ratio, gross profit,

[5] http://finance.bi.no/~bernt/financial_data/ose_asset_pricing_data/index.html

[6] In cases where the free-float number of share is not available, we use the total number of shares.

investment growth) at the end of year $t-1$. The Low (L) portfolio contains stocks below the 30$^{\text{th}}$ percentile, the Middle (M) portfolio contains stocks between the 30$^{\text{th}}$ and the 70$^{\text{th}}$ percentiles, and the High (H) portfolio contains stocks above the 70$^{\text{th}}$ percentile. The factor is the portfolio that is long the High (H) portfolio and short the Low (L) portfolio. Our methodology is similar to Stambaugh and Yuan (2017), with the exception that they use the 20$^{\text{th}}$ and 80$^{\text{th}}$ percentiles as breakpoints. Our motivation for using the 30$^{\text{th}}$ and 70$^{\text{th}}$ percentiles stems from the relatively small cross-section of stocks listed on the Oslo Stock Exchange. We choose percentiles that allow us to include more stocks in the High and Low portfolios, thereby ensuring that our factors are well diversified.

The value factor, HML_t, is based on firms' book-to-market ratio. We calculate the book value of equity using accounting data from NHH for the 1997 to 2011 period and from Datastream for 2012 onward. For the 1997-2011 sub-period, the book value of equity is the sum of the book value of stockholders' equity (*Sum egenkapital*) and the balance sheet deferred taxes and investment credit (*Utsatt skatt*), net of capital raised through preference shares (*Preferansekapital*). The book-to-market ratio for year t is the ratio of the book value of equity at the end of December of year t-1 to the market value of equity at the end of year t-1. For the period starting in 2012, we rely on Datastream and use common equity (Datastream code WC03501) on December 31 divided by the closing market value of the last trading day in December. If this information is not available, we use the inverse of the price-to-book value ratio (Datastream code WC09304).

Cohen and Polk (1998) and Cohen et al. (2003) show that, if book-to-market ratios are decomposed into an across-industry component and a within-industry component, then only the within-industry component is priced. Building on their finding, we adjust for industry effects by subtracting the industry's mean book-to-market ratio from each firm's book-to-market ratio. Industries are defined by 11 distinct 2-digit GICS codes. To reduce noise, we only subtract the industry average from firm book-to-market ratios if the industry average is based on five or more companies in a given year. We use the demeaned book-to-market ratio to form the Low, Medium, and High portfolios of the value factor. The industry adjustment leads to a small increase in the Sharpe Ratio of the resulting value factor.

The profitability factor, RMW_t, is defined as in Novy-Marx (2013). For the 1997-2011 sub-period, gross profit, GP_t, is the difference between total revenue (*Sum inntekter*) and cost of goods sold (*Vareforbruk*). Let $TA_{t-1,Dec}$ denote the value of total assets (*Sum eiendeler*). The profitability characteristic available in June of year t is $PROF_{t,June} = GP_{t-1,Dec}/TA_{t-1,Dec}$. For the 2012-2018 subsample, we define gross profit $GP_{t-1,Dec}$ as revenues (Datastream code WC01001) minus cost of goods sold (Datastream code WC01051).

The investment factor, CMA_t, is defined as in Fama and French (2015). We define a firm's investment growth from year $t-2$ to year $t-1$ as the growth rate of total assets $INV_{t,June} = (TA_{t-1,Dec} - TA_{t-2,Dec})/TA_{t-2,Dec}$.

The momentum factor, MOM_t, is constructed as in Carhart (1997). A stock's momentum characteristic is its geometric return over the previous 12 months, excluding the most recent month from consideration: $MOM_t = \prod_{\ell=2}^{12}(1 + r_{t-\ell}) - 1$. We sort stocks according to their momentum. The Low (L) portfolio contains stocks up to the 30$^{\text{th}}$ percentile and the High (H) portfolio contains stocks above the 70$^{\text{th}}$. The momentum factor is long the High portfolio and short the Low portfolio.

Table IA.2 reports pricing statistics for the firm factors in Norway between 1997 and 2018. The Sharpe ratio of the market factor, MKT_t, is 0.32. The profitability, investment, and momentum factors produce statistically significant CAPM-alphas during the sample period.

III.B. Alternative Age and Wealth Factors

In Section III.A of the main text, we construct investor pricing factors by sorting stocks according to the average age or average wealth of their individual investor base. Columns 1 and 2 of Table IA.3 reports summary statistics on the Sharpe ratio and CAPM-alpha of the combined age-wealth factor, AW_t resulting from this construction. In this Section, we investigate the robustness of these results to alternative specifications of the pricing factors.

Factors Constructed from Portfolio Holdings. An alternative construction of the factors, developed in Section II.C of the main text, relies on the underlying age and wealth *portfolio* factors, $\boldsymbol{g}_{\mathbf{AGE},t}$ and $\boldsymbol{g}_{\mathbf{WEALTH},t}$. The age portfolio is long the portfolio of investors above 60 and short the portfolio of investors below 30. The wealth portfolio is long the portfolio of investors in the top 10% of the net worth distribution and short the portfolio of investors in the bottom 30%. We now use the returns on the portfolios $\boldsymbol{g}_{\mathbf{AGE},t}$ and $\boldsymbol{g}_{\mathbf{WEALTH},t}$ as alternative pricing factors.

In columns 3 and 4 of Table IA.3, we report summary statistics on the resulting age-wealth factor, AW_t, that is is the equal-weighted average of the alternative age and wealth factors. Consistent with the baseline results, the alternative age-wealth factor has high and significantly positive Sharpe ratio and CAPM alpha. We note that the Sharpe ratio and alpha of the alternative age-wealth factor are smaller than the baseline results. The explanation is that the long and short legs of the alternative age and wealth factors do not differ from each other as much as their baseline equivalents; the long and short legs are aggregate portfolios of investor holdings and thus include many similar stocks, albeit in different proportions. By contrast, the baseline method generates no overlap between stocks in the long leg and stocks in the short leg, which produces a higher return spread.

Subset of Stocks. Some Norwegian stocks in Norway have high government ownership. To verify that these stocks do not drive our main results, we remove stocks with the highest government ownership share. These stocks include Equinor ASA, Norsk Hydro, Telenor, DnB, Entra, Yara, and Kongsberg gruppen. We then re-create the combined age-wealth factor from the subset of remaining stocks using the two-step method described in Section III of the main text. Columns 5 and 6 of Table IA.3 report the pricing results. The combined age-wealth factor has significantly positive Sharpe ratio and CAPM alpha.

Subset of Investors. We verify that the pricing performance of the age-wealth factor are not driven by a small set of investors. Using the two-step method, we re-create the age and wealth factors from the random subset of two-thirds of investors that is described in Section IV.A of the main text. Columns 7 and 8 of Table IA.3 report the pricing statistics of the combined age-wealth factor constructed from the sub-sample of investors. The results are equivalent to those obtained with the full sample.

III.C. Out-of-Sample Performance of Investor Factors

As Section III.E of the main text shows, the upward in-sample bias of the maximum Sharpe ratio (SR) generated by a factor model is less pronounced when factors other than the market are investor-based than when they are firm-based. This result is established by using the ratio of the average out-of-sample SR to the average in-sample SR of the tangency portfolio as a measure of performance. In this Section, we investigate the drivers of the performance differential between the two types of factors.

We begin by calculating additional statistics for each factor model. For each of the model's 100,000 simulations, we calculate the in-sample Sharpe ratio and average return of the estimated tangency portfolio predicted by the factor model. We then calculate four statistics: the average and standard deviation of the in-sample Sharpe ratio of the tangency portfolio across the simulations, and the average and standard deviation of the in-sample average return of the tangency portfolio across the simulations.

Table IA.4 reports four univariate regression results of a model's ratio of out-of-sample to in-sample average Sharpe ratio on the statistics described above. The regressions are estimated over the 22 factor models described in Section III.E of the main text. Column 1 shows that models with high average in-sample Sharpe ratios tend to perform better out-of sample. Column 2 also shows that models with stable in-sample Sharpe ratios also tend to perform better out-of-sample. Perhaps more importantly, column 4 shows that the precision of a model's in-sample average return has the greatest explanatory power on its out-of-sample performance. We see that a low standard deviation of the tangency portfolio's in-sample average return predicts a strong out-of-sample performance. The adjusted R^2 coefficient exceeds 50% for column 4.

The results of this analysis suggest that investor-based models have more stable in-sample moments than firm-based models, which helps to explain why they generally perform better in our out-of-sample tests.

IV. Appendix to Section IV: The Cross-Section of Investor Factor Tilts

IV.A. Calculation of Labor Income Risk

We calculate income risk by adopting the methodology of Guvenen et al. (2017). An investor's income beta is defined as the sensitivity of her labor income growth to real GDP growth. We retrieve the real GDP series from Statistics Norway.[7]

We construct the sample of workers as follows. For a given year t, we include all individuals who meet two conditions: (i) a minimum income of NOK 5,000, and (ii) a reported income above NOK 5,000 in at least three out of the five previous years. We distinguish between retirees and individuals who are between 26 and 65 years old and are actively working.

We define an individual's real earnings, $Y_{i,t}$, as the ratio of her nominal earnings to the consumer price index, and we denote log real earnings by $y_{i,t} = \ln(Y_{i,t})$. The first difference, $\Delta\, y_{i,t}$, is our measure of real earnings growth. We estimate an individual's permanent income by the average of real earnings $Y_{i,t}$ over the 5-year period between years $t-6$ to $t-2$. This choice of dates ensures that there is no overlap between the period over which earnings growth is computed (years $t-1$ and t) and the period over which average earnings are computed (years $t-6$ to $t-2$). This gap allows us to rule out any mechanical correlation between earnings growth in year t and historical average earnings.

We calculate an individual's income risk in several steps. First, we classify individuals into $g = 1, 2, 3, ..., G-1$ groups based on permanent income (12 groups) and employment sector (20 groups), where the employment sector groups are discussed in Section II.A of this Appendix.

[7] This measure refers to Mainland Norway and excludes general government.

We also include retirement as a separate group, which produces 241 groups in total. Consistent with the main text, the 12 permanent-income groups include the first 9 deciles of the permanent income distribution (groups 1-9), the 90$^{\text{th}}$-99$^{\text{th}}$ percentiles (group 10), the 99$^{\text{th}}$-99.9$^{\text{th}}$ percentiles (group 11), and the top 0.1% (group 12). Occupational sectors are described in Section II.A of this Internet Appendix. We filter out groups with fewer than 1000 individuals and end up with 220 groups.

For each group, we run a panel regression of the annual income growth of investor i in year t, denoted by $\Delta y_{i,t}$, on real GDP growth in the same year:

$$\Delta y_{i,t} = a_g + \beta_g^{GDP} \Delta GDP_t + \varepsilon_{i,t}. \tag{IA-31}$$

The regression yields a slope coefficient β_g^{GDP} for each group. We assign this coefficient to all individuals in the group and use it as a proxy for their exposure to macroeconomic risk.

Figure IA.2 displays the distribution of income betas for four permanent income groups: the 40$^{\text{th}}$-50$^{\text{th}}$ percentiles, the 60$^{\text{th}}$-70$^{\text{th}}$ percentiles, the 80$^{\text{th}}$-90$^{\text{th}}$ percentiles, and the top 0.1%s. For each group, the distribution in income betas is economically significant and ranges from an income beta of -0.5 (retirees) to more than unity (household services, finance, petroleum, technical services). We note that income betas also differ across the permanent income distribution. For example, the income beta of individuals working in finance exceeds 2 for high-income individuals but is below 1 for the median-income individual.

REFERENCES

Bhuller, Manudeep, 2009, Inndeling av norge i arbeidsmarkedsregioner, *Notater 2009/24,* Statistics Norway.

Butikofer, Aline, Antonio Dalla-Zuanna, and Kjell G. Salvanes, 2018, Breaking the links: Natural resource booms and intergenerational mobility, *Review of Economics and Statistics* forthcoming.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cohen, Randolph B., and Christopher Polk, 1998, An investigation of the impact of industry factors in asset-pricing tests, Working paper, Harvard Business School.

Cohen, Randolph B, Christopher Polk, and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.

Fama, Eugene F, and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Guvenen, Fatih, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo, 2017, Worker betas: Five facts about systematic earnings risk, *American Economic Review* 107, 398–403.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Oedegaard, Bernt Arne, 2020, Empirics of the Oslo Stock Exchange: Basic descriptive results, 1980-2019, Working paper, University of Stavanger.

Stambaugh, Robert, and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

Table IA.1
Industry Classification

This table reports our industry classification based on the A64 Rev. 2 version of Norway's National Accounts and Eurostat's NACE industry codes.

Our Industries	A64 Rev. 2	NACE
Resource industries	Crop and animal production, hunting and related service activities	1
Resource industries	Forestry and logging	2
Resource industries	Fishing	3.3
Resource industries	Aquaculture	3.2
Resource industries	Mining and quarrying	5, 7-8, 9.9
Resource industries	Electricity, gas, steam and air conditioning supply	35
Petroleum	Oil and gas extraction	6
Petroleum	Transport via pipelines	49.5
Petroleum	Service activities incidental to oil and gas	9.1
Consumer manufacturing	Manufacture of food products, beverages and tobacco products	10-12
Consumer manufacturing	Manufacture of textiles, wearing apparel and leather products	14-15
Consumer manufacturing	Printing and reproduction of recorded media	18
Consumer manufacturing	Manufacture of furniture	31-32
Material manufacturing	Manufacture of wood and wood products, except furniture	16
Material manufacturing	Manufacture of paper and paper products	17
Material manufacturing	Refined petroleum, chemical and pharmaceutical products	19-21⋆⋆
Material manufacturing	Chemical commodities	20.11-20.15
Material manufacturing	Manufacture of rubber and plastic products	22
Material manufacturing	Manufacture of other non-metallic mineral products	23
Material manufacturing	Manufacture of basic metals	24
Material manufacturing	Fabricated metal products, except machinery and equipment	25
Technological manufacturing	Manufacture of computer, electronic and optical products	26
Technological manufacturing	Manufacture of electrical equipment	27
Technological manufacturing	Manufacture of machinery and equipment n.e.c.	28
Technological manufacturing	Manufacture of motor vehicles, trailers and semi-trailers	29
Technological manufacturing	Building of ships, oil platforms and modules	30
Technological manufacturing	Repair and installation of machinery and equipment	33
Public administration	Water collection, treatment and supply	36
Public administration	Sewerage	37-39
Public administration	Public administration central/local government	84
Public administration	Defence	84.2
Construction	Construction	41-43
Trade	Wholesale and retail trade and repair of motor vehicles	45
Trade	Wholesale trade, except of motor vehicles	46
Trade	Retail trade, except of motor vehicles	47
Transportation and logistics	Land transport, except transport via pipelines	49.1-49.4
Transportation and logistics	Inland water transport and supply	50.1-50.4⋆
Transportation and logistics	Air transport	51
Transportation and logistics	Warehousing and support activities for transportation	52
Transportation and logistics	Postal and courier activities	53
Tourism	Ocean transport	50.101, 50.201
Tourism	Accommodation and food service activities	55-56
Tourism	Travel agency and tour operator reservation service	79

(Continued)

Table IA.1 - *Continued*

Our Industries	A64 Rev. 2	NACE
Media and ICT	Publishing activities	58
Media and ICT	Motion picture and video program production, broadcasting	59-60
Media and ICT	Telecommunications	61
Media and ICT	Computer programming and related activities	62-63
Finance	Insurance, except compulsory social security	65
Finance	Activities auxiliary to financial services and insurance activities	66
Finance & consulting	Financial service activities (not insur.)	64
Finance& consulting	Other prof., scientific activities	74-75
Finance & consulting	Real estate activities	68
Knowledge-based business services	Legal and accounting activities	69-70
Knowledge-based business services	Scientific research and development	72
Knowledge-based business services	Advertising and market research	73
Technical services	Architectural and engineering consultancy activities	71
Business support services	Rental and leasing activities	77
Business support services	Employment activities	78
Business support services	Security and investigation activities	80-82
Education	Education	85
Health and social services	Human health activities	86
Health and social services	Social work activities	87-88
Non-profit and household services	Creative, arts and entertainment activities	90-92
Non-profit and household services	Sports activities and amusement and recreation activities	93
Non-profit and household services	Activities of membership organisations	94,99
Household services	Repair of computers and personal and household goods	95
Household services	Other personal service activities	96
Household services	Activities of households as employers	97

Table IA.2
Return Performance of Firm Factors

This table reports statistics on the return performance of firm factors estimated from the universe of Norwegian stocks in 1997-2018. We consider value-weighted factors at the monthly frequency and report the average return, standard deviation, Sharpe Ratio, CAPM-alpha, and the t-statistic of the CAPM-alpha. The market factor, MKT_t, is the value-weighted portfolio of Norwegian stocks. We sort stocks by market capitalization to obtain the size factor, SMB_t, by industry adjusted book-to-market ratio to obtain the value factor, HML_t, by profit margin to obtain the profitability factor, RMW_t, and by total asset growth to obtain the investment factor, CMA_t. To construct the momentum factor, MOM_t, we consider the geometric return over the previous 12 months where the most recent month is left out. The estimates are based on the 1997-2018 period.

	Pricing Factor					
	MKT	SMB	HML	RMW	CMA	MOM
Mean return (% per month)	0.56	-0.13	0.09	0.85	0.54	0.94
Standard deviation (% per month)	5.95	4.07	5.10	5.86	5.42	7.44
Sharpe ratio (annual units)	0.32	-0.11	0.06	0.50	0.34	0.44
CAPM-alpha (per month)		-0.02	0.16	0.96	0.66	1.05
t(CAPM-alpha)		-0.09	0.51	2.70	2.01	2.32

Table IA.3
Alternative Constructions of Investor Factors

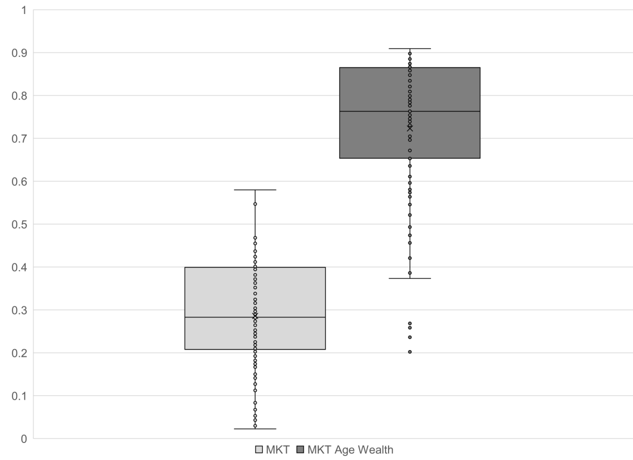
This table reports statistics on the return performance of the combined age-wealth investor factor, AW_t, obtained from different constructions of the pricing factors. For each factor construction methodology, we report the combined age-wealth factor's Sharpe ratio ("SR") and CAPM-alpha ("Alpha"). Columns 1 and 2 focus on the baseline method used in the main text. In columns 3 and 4 ("Direct Holdings"), the underlying age and wealth factors are constructed from investor portfolios, as explained in Section II of the main text. Specifically, the age portfolio is long the portfolio of investors above 60 and short the portfolio of investors below 30. The wealth portfolio is long the portfolio of investors in the top 10% of the net worth distribution and short the portfolio of investors in the bottom 30%. In columns 5 and 6 ("Fewer Stocks"), we construct the age and wealth factors by following the two-step method discussed in Section III of the main text, but we exclude from the stock universe Equinor ASA, Norsk Hydro, Telenor, DnB, Entra, Yara, and Kongsberg gruppen, which have the highest share of government ownership. In columns 7 and 8 ("Fewer Investors"), the age and wealth factors are constructed from the two-step method on a random sample of 2/3 of the investors described in Section IV.A of the main text.

	Baseline Method		Direct Holdings		Fewer Stocks		Fewer Investors	
	SR	Alpha	SR	Alpha	SR	Alpha	SR	Alpha
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimate	0.68	1.05	0.17	0.32	0.67	1.14	0.45	0.73
Standard error	0.07	0.29	0.06	0.18	0.07	0.31	0.06	0.30
t-value	9.98	4.15	2.75	1.81	9.85	3.67	6.96	2.43
Number of months	264	264	264	264	264	264	264	264

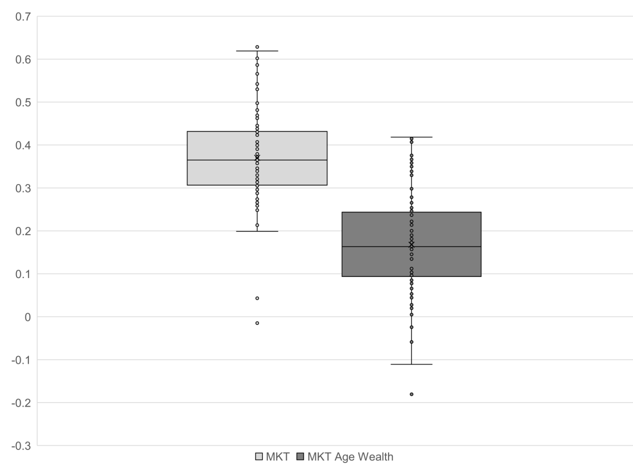
Table IA.4
Drivers of Out-of-Sample Performance of Factor Models

This table reports OLS regressions of a factor model's *out-of-sample* performance measure on the model's *in-sample* statistics. The analysis is based on the 22 factor models described in Section III.E of the main text. We generate 100,000 bootstrap simulations from historical factor returns, where each simulation provides both an in-sample and an out-of-sample series of return factors, as described in Section III.E of the main text. For each factor model, we estimate for each simulation the tangency portfolio by mean-variance optimization of in-sample factor returns and a shrinkage estimator of the covariance matrix; we then compute the out-of-sample Sharpe ratio of the estimated tangency portfolio. The dependent variable is the *out-of-sample* average Sharpe ratio divided by the *in-sample* average Sharpe ratio of the tangency portfolio, where averages are taken over the 100,000 simulations. The independent variables include the mean and standard deviation of the in-sample Sharpe ratio of the tangency portfolio, and the mean and standard deviation of the in-sample average return of the tangency portfolio. Statistical significance is indicated by ***, **, and * for the 0.01, 0.05, and 0.10 levels.

	Dependent Variable: Out-of-Sample SR / In-Sample SR			
	(1)	(2)	(3)	(4)
In-Sample Statistics				
Mean of Sharpe ratio	1.042**			
	(0.445)			
Standard deviation of Sharpe ratio		−5.750***		
		(1.389)		
Mean of average return			10.643	
			(6.790)	
Standard deviation of average return				−0.097***
				(0.019)
Constant	0.368***	0.966***	0.490***	0.632***
	(0.093)	(0.096)	(0.062)	(0.021)
Observations	22	22	22	22
Adjusted R^2	0.176	0.435	0.065	0.533
F Statistic	5.485**	17.144***	2.456	24.939***

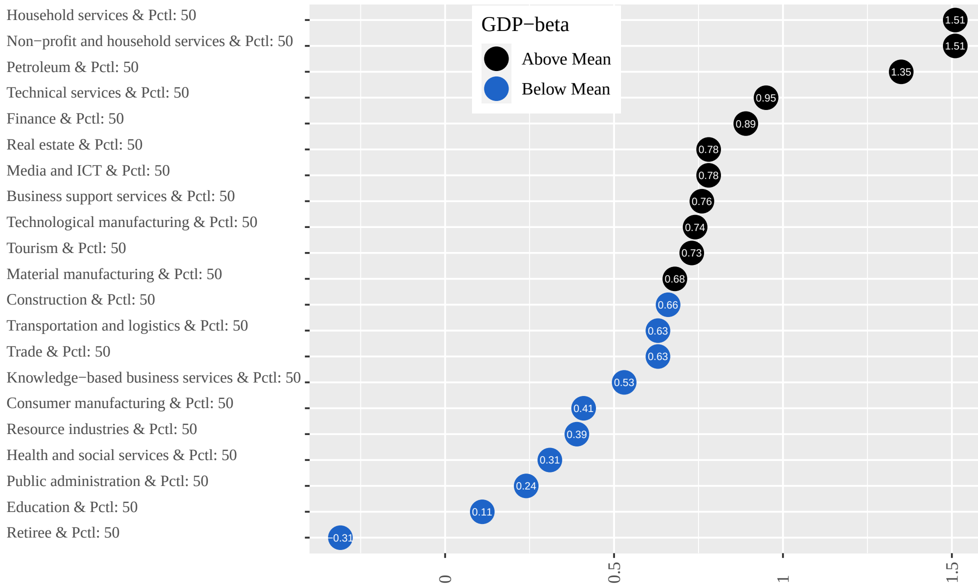


(a) R^2 Coefficient

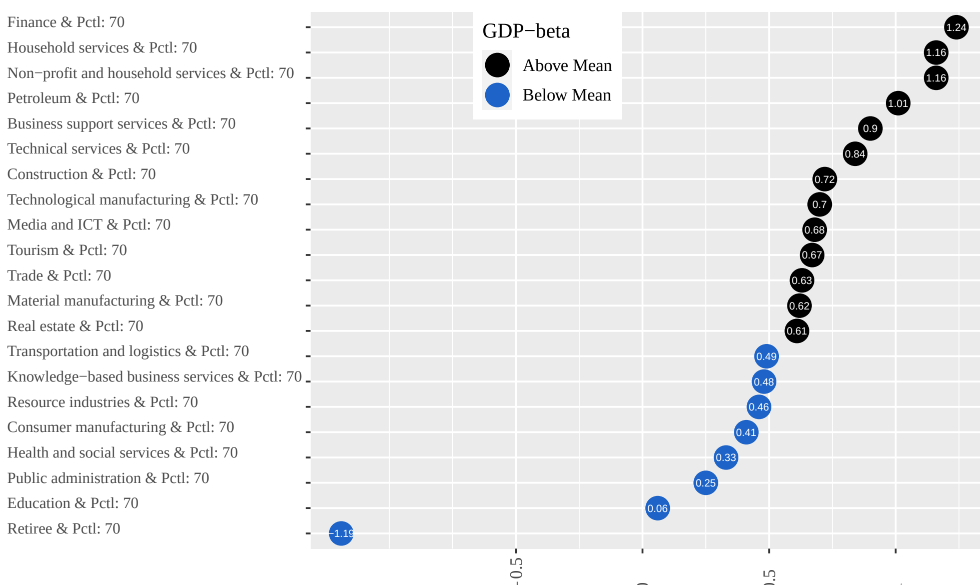


(b) Intercept

Figure IA.1 Explanatory power of age, wealth, and market portfolios. This figure plots the distribution of the R^2 coefficient and the intercept obtained from running pooled OLS regressions of the portfolio holdings of 93 investor groups on (i) the weights of the market (light grey), and (ii) the weights of the market, age, and wealth factor portfolios (dark grey). Investor groups are sorted by socioeconomic and geographic indicators. The age portfolio is long the portfolio of investors above 60 and short the portfolio of investors below 30. The wealth portfolio is long the portfolio of investors in the top 10% of the net worth distribution and short the portfolio of investors in the bottom 30%. The estimation is run over the 2002-2017 period.

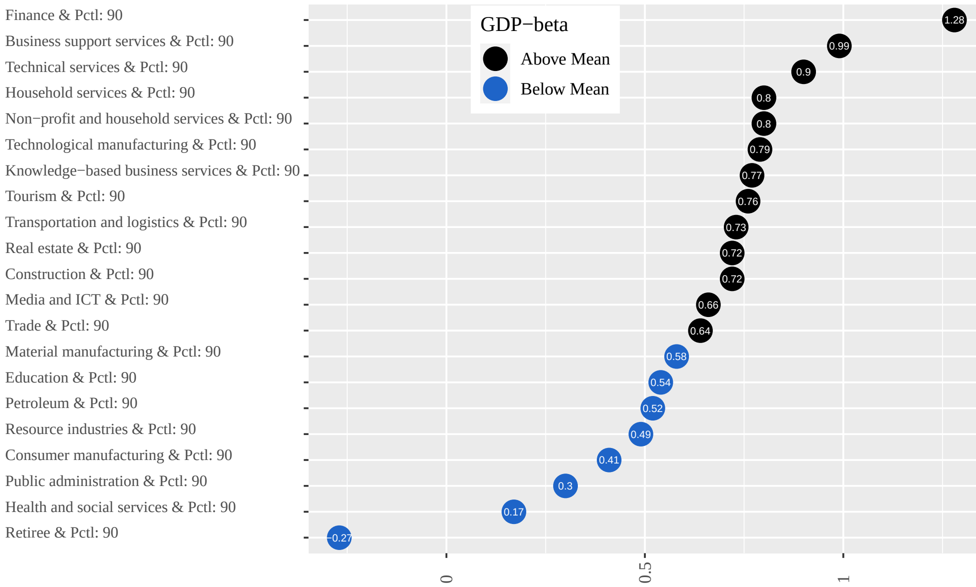


(a) Income Beta of Investors in 40th-50th Percentiles of Permanent Income

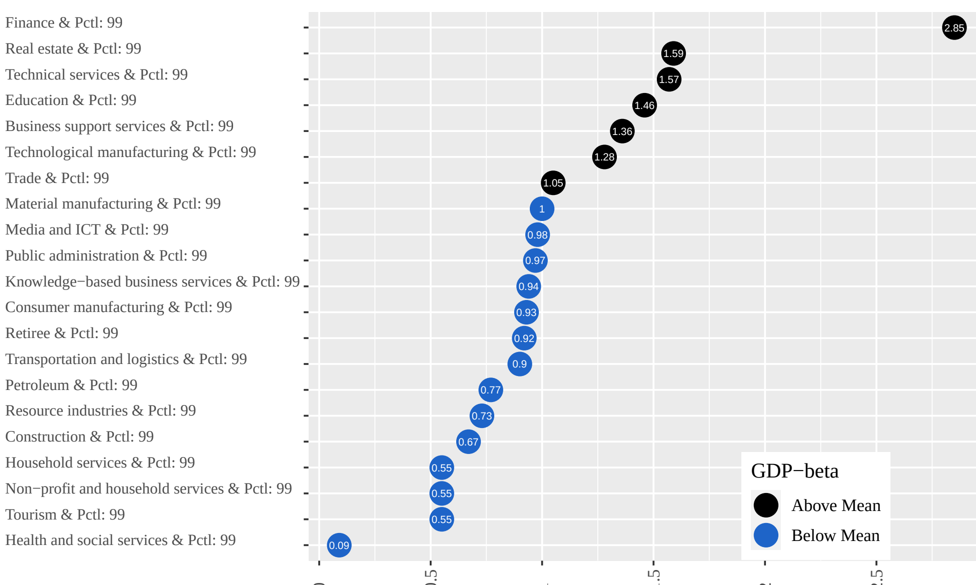


(b) Income Beta of Investors in 60th-70th Percentiles of Permanent Income

Figure IA.2 Income beta across occupational sectors. This figure plots income betas across occupational sectors for four groups of investors sorted by permanent income. An income beta is the slope coefficient from a panel regression of an investor's annual income growth on real GDP growth, where the estimation is conducted within a group of investors in the same occupational sector and labor income bracket. The estimation is run on a panel of Norwegian individual investors in 1997-2018.



(c) Income Beta of Investors in 80th-90th Percentiles of Permanent Income



(d) Income Beta of Investors in 99th-99.9th Percentiles of Permanent Income

Figure IA.2 Income beta across occupational sectors.– *Continued*