BIS Working Papers
No 1091

FinTech, investor sophistication and financial portfolio choices

by Leonardo Gambacorta, Romina Gambacorta and Roxana Mihet

Monetary and Economic Department

April 2023

JEL classification: G1, G5, G4, D83, L8, O3.

Keywords: inequality, inclusion, FinTech, innovation, Matthew Effect.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

FinTech, investor sophistication and financial portfolio choices

Leonardo Gambacorta[*], Romina Gambacorta[†] and Roxana Mihet[‡§]

Abstract

This paper analyses the links between advances in financial technology, investors' sophistication, and the composition and returns of their financial portfolios. We develop a simple portfolio choice model under asymmetric information and derive some theoretical predictions. Using detailed microdata from Banca d'Italia, we test these predictions for Italian households over the period 2004-20. In general, heterogeneity in portfolio composition and in returns between sophisticated and unsophisticated investors grows with improvements in financial technology. This heterogeneity is reduced only if financial technology is accessible to everyone and if investors have a similar capacity to use it.

Keywords: Inequality; Inclusion; FinTech; Innovation; Matthew Effect.

JEL-Codes: G1, G5, G4, D83, L8, O3

[*]Bank for International Settlements and CEPR. Contact: leonardo.gambacorta@bis.org

[†]Banca d'Italia. Contact: romina.gambacorta@bancaditalia.it

[‡]Swiss Finance Institute at HEC Lausanne and CEPR. Contact: roxana.mihet@unil.ch

[§]We thank Robert Marquez and Anthony Zhang, for extremely constructive and helpful feedback. We also thank participants at the RCFS Winter Conference, and Andreas Fuster, Tarun Ramadorai, Vatsala Shreeti and Laura Veldkamp for their useful comments. We are also indebted to Ginette Eramo, Marco Langiulli and Paola Rossi for helping us to better understand and use the data of Banca d'Italia Regional Lending Survey and Supervisory Reports of Italian banks. The views in this paper are those of the authors only and do not necessarily reflect those of Banca d'Italia or the Bank for International Settlements.

1 Introduction

Financial technology (FinTech) allows for more efficient use of data to solve problems of asymmetric information. The application of artificial intelligence techniques in finance (i.e. machine learning and big data) can increase financial inclusion, while reducing the costs of financial services at the same time. Recent research indicates substantial benefits of financial technology in alleviating discrimination in mortgage markets (Bartlett et al. (2022)), in facilitating the calculation of credit scoring for opaque borrowers (Berg et al. (2020)), in reducing minimum investment requirements and capital charges (Abraham et al. (2019)), and in allowing users access to an expanding array of information useful for investment decisions remotely and effortlessly (Katona et al. (2018)).

However, financial technology can also increase discrimination among investors groups, especially if they have a different access to (or use of) the new technology. For example, FinTech can create opportunities for sophisticated market players to acquire better data and formulate profitable trading strategies at the expense of less sophisticated ones. Early articulations of this argument were proposed in Becker (1967) and Arrow (1987). Becker (1967) suggests that households who can secure systematically higher financial returns become wealthier, and Arrow (1987) proposes that individuals with lower costs of information acquisition buy more information and therefore get higher rates of return. Subsequent variations of these arguments, with wealth effects and fixed stock market participation costs (Peress (2004)), pooling through mutual funds in order to share the cost of data acquisition (Mihet (2018)), or heterogeneity in data processing abilities (Kacperczyk et al. (2018), Azarmsa (2019)) have all reached the same theoretical conclusion: general progress in financial information technologies benefits sophisticated investors to the detriment of less sophisticated ones. This is at odds with the general optimism in the financial industry. Thus, whether advances in financial technology are democratizing finance and leveling-out the playing field is unclear.

To contribute to this debate, in this paper we analyse the links between advances in financial technology, investors' sophistication levels and financial portfolios' composition and returns using standard portfolio theory which we test with novel micro-level data. In particular, we use a workhorse portfolio choice model with asymmetric information, where investors differ in their sophistication levels as measured by their capacity to process information, and choose which assets to learn about and invest in. We use the model to derive several testable predictions on the impact of absolute and relative increases in sophistication on investors' shares of risky asset classes and ex-post returns. We then test these predictions using a unique micro-level data set merging information on Italian households' and banks' characteristics. Data on households are gathered from Banca d'Italia's Survey on Household Income and Wealth (SHIW) from 2004 to 2020. We complement these data with the Regional Bank

Lending Survey (RBLS) which collects information on the digitalization of the Italian banking sector. Particularly, we succeed in building a set of indicators for the degree of bank investment in FinTech technologies for the provision of financial services dating back to 2004.

The main contribution of this paper is to combine novel data with a theory-motivated empirical approach to investigate the potential channels that could explain the observed financial return and portfolio choice heterogeneity between investors with different levels of sophistication, which we proxy by financial literacy, and access to FinTech services, which we measure using the characteristics of the investor's bank. The theoretical predictions are tested, one at a time, using microdata that allows us to control for households and banks characteristics without relying on *ad hoc* assumptions.

Three predictions of the model seem particularly relevant in light of observed trends. The first implies that capital return heterogeneity between sophisticated and unsophisticated investors grows with enhancements in aggregate financial information technology that hold constant or amplify the gap in investor sophistication. Second, the model predicts a growing presence of sophisticated investors in risky asset classes, and a retrenchment of less sophisticated investors from trading and stock market ownership in general. Third, and in a departure from existing work, the theoretical framework implies that innovations that lower the gap in investor sophistication levels also lower capital income inequality.

We empirically verify these predictions one by one. Examining the realized rates of return and the portfolio composition of investors who differ in their financial literacy levels, and controlling for a multitude of factors – such as households' risk-aversion, age, gender, access to remote banking, as well as time and region fixed effects – we find that heterogeneity in financial returns and in the share of risky assets between sophisticated and unsophisticated investors increases with innovations in financial technology, conditional on the existence of a growing gap in investor sophistication. In other words, advances in financial technology amplify inequalities in the two groups of investors if, at the same time, they do not reduce the investors' sophistication gap.

The policy implication of our findings is that financial technology enhancements need also to bridge the gap in investor sophistication levels in order to leave no one behind. Merely providing access to innovative sources of financial information and advice is not enough. As traditional and non-traditional financial institutions move more of their services and initiatives online, the least sophisticated risk of being left further behind. Avoiding this requires making financial technologies not only accessible to everyone but also usable by everyone. Development initiatives need to incorporate innovative elements in order to expand the use of technological advances to the most marginalised; build the capacity of the digitally disadvantaged to help them catch up with the rest; and curtail the capacity and intent of malicious actors using digital technologies to manipulate information and markets.[1]

Our paper is related to three main strands of literature.

First, a number of studies look into the impact of technological innovation on inequality. Acemoglu (2002) discusses how technical change in the latter half of the 20th century has been "skills-biased", leading to greater differentials between skilled and unskilled labour. Jaumotte et al. (2013) discuss the relative contributions of technological progress and globalisation in explaining rising inequality across countries in the period 1981–2003; they find that technological progress has been a more important factor. Differently from these papers, we look at how progress in financial technology could contribute to financial income inequality.

A second strand of the literature looks at the so called *"Matthew effect"*, the mechanism of the well-endowed receiving further privilege, e.g. the rich getting richer (Merton, 1968).[2] Using administrative data from Norway, Fagereng et al. (2020) find higher returns on financial assets and net wealth among wealthier households, which they call "scale dependence". Frost et al. (2022) find that while households of all wealth deciles benefit from the effects of financial development and financial technology, these benefits are larger when moving toward the top of the wealth distribution. Differently from these papers, we look at the specific mechanisms through which more sophisticated investors (not necessarily the wealthiest) could benefit from more advances in financial technologies.

A third strand of the literature looks at the effects of financial literacy. For Sweden, Calvet et al. (2007) find that more sophisticated households are more likely to have financial investments and invest efficiently. Likewise, using Dutch household survey data, Deuflhard et al. (2019) find that a one-standard deviation increase in financial literacy is associated with a 12 per cent increase in returns on saving accounts. They find that online accounts are one channel through which financial literacy has a positive correlation with returns. Differently from these papers that focus on the link between financial literacy and portfolio choice, we consider also the effects of advances in digital financial services on such a link.

The rest of the paper is organised as follows. Section (2) describes the model and derives the testable predictions. Section (3) discusses the data and documents some stylized facts. The empirical analysis is developed in section (4), including some robustness checks and extensions. The last section summarises the main conclusions.

[1] Given the growing interest in cryptocurrencies, there is a pressing need to increase financial literacy and bridge the gap in investor sophistication levels. Cryptocurrencies are a highly volatile asset, particularly popular among younger and male investors, who are commonly identified as the most risk-seeking segment of the population (Auer et al., 2022).

[2] The name of the *"Matthew effect"* derives from the New Testament Book of Matthew (25:29), in which it is written: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath".

2 Theoretical Model and Predictions

We present a one-period general equilibrium portfolio choice model with endogenous learning about assets' stochastic payoffs in the spirit of Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk et al. (2018), where investors' utility function is CRRA with respect to end of period wealth, investors learn with information capacity constraints, and there are a finite number of uncorrelated risky assets.

Our goal is not to develop innovative theory. Rather, the goal is to write down a simple framework rooted in workhorse models of portfolio choice with endogenous information acquisition to quantify the channels and implications of progress in financial information technologies and provide some clean testable predictions. Our model has three key features: (i) Investors' preferences belong to the constant relative risk aversion (CRRA) and implicitly the decreasing absolute risk aversion (DARA) family, implying risky portfolio shares that increase with investors' wealth. This is consistent with our data (see our summary statistics table in the empirical section) and with a long behavioural and experimental economics literature that supports DARA preferences instead of the more commonly used CARA preferences in finance. CRRA preferences, as opposed to CARA preferences, not only reflect reality better, but also allow us to separate (theoretically and empirically) the effects due to progress in financial information technologies from wealth heterogeneity and risk-aversion and isolate each of their contribution on stock market portfolio choices and the evolution of financial income and wealth. (ii) The financial market consists of multiple risky assets; this assumption allows us to have predictions on asset ownership across different risky asset classes. (iii) Investors differ in their financial literacy levels, which we model as heterogeneity in investors' information capacity constraints (because we can observe financial literacy empirically).

After we fully characterize the equilibrium, we derive several theoretical predictions for the impact of growth in information processing capacity. In a departure from existing work, we focus on three distinct cases: 1) an asymmetric (relative) growth in the information processing capacity of sophisticated and unsophisticated investors that favours sophisticated investors; (2) a symmetric (absolute) growth in information processing capacity of sophisticated and unsophisticated investors; and (3) an asymmetric (relative) growth in the information processing capacity of sophisticated and unsophisticated investors that favours unsophisticated investors.

2.1 Set-up

This is a static model divided into two periods, time t and time $t+1$. In the first period, at time t, the investor chooses the precision of signals about asset payoffs, subject to an information capacity constraint. At the beginning of the second period, at time $t+1$, the investor observes signals and then chooses which assets to purchase. At the end of the second period, the investor receives the asset payoffs and realizes his utility.

Investors' preferences. There is a unit mass of perfectly competitive investors indexed by j. Investors have CRRA preferences with respect to their terminal wealth $W_{j,t+1}$ and a relative risk-aversion coefficient $\rho > 0$, as in Van Nieuwerburgh and Veldkamp (2010) and Brunnermeier (2001):

$$max\ \mathbb{E}\frac{1}{1-\rho}W_{j,t+1}^{1-\rho} \tag{1}$$

This specification guarantees that an investor increases his investment in risky stocks as his wealth increases.[3] Moreover, an investor with a CRRA utility function always invests a constant fraction of his wealth in the risky portfolio.

Independent assets. Without loss of generality, the financial market consists of one riskless asset, with a small rate of return r and a price normalized to 1, and N risky assets, indexed by i with prices p_i and stochastic payoffs which are i.i.d. and normally distributed according to $z_i \sim N(\bar{z}_i, \sigma_i^2)$. [4] The supply of the risky assets is stochastic and provided by liquidity traders $x_i \sim N(\bar{x}_i, \sigma_{xi}^2)$, and is independent of payoffs and across assets. The riskless asset is in unlimited supply.

Information allocation choice. There are two aspects to the information choice: how much information to acquire, and how to allocate that information among assets (whether to acquire information about one or all of the risky assets). Information is obtained in the form of signals, which are then used to update the beliefs that inform the investors' portfolio allocations. We assume that each investor receives a separate, independent signal s_{ij} on each

[3]The coefficient of absolute risk aversion, $-u''/u = \rho W_{j,t+1}^{-1}$, is decreasing in wealth, implying that the CRRA utility considered belongs to the DARA utility family. Moreover, the coefficient of relative risk aversion is a constant, ρ.

[4]In our case, the covariance between any two assets i and i' is zero, $\sigma_{i,i'} = 0$. For any set of correlated assets with a full rank variance-covariance payoff matrix, we can extract $n \leq N$ independent principal components as linear combinations of those correlated assets. The solution to the correlated asset problem would be exactly the same, as long as we define $[q_1, q_2, ..., q_n]$, $[p_1, p_2, ..., p_n]$ and $[z_1, z_2, ..., z_n]$ to be the quantities invested, prices and payoffs of the n risk factors. Investors learn and invest in the risk factors just as if they were independent assets. Investing more into one asset is equivalent to investing in one risk factor, which is a portfolio of assets with correlated payoffs. While the solution remains unchanged, only the interpretation is different with correlated assets.

of the assets' payoffs z_i:[5]

$$z_i = s_{ji} + \epsilon_{ji}, \text{ where } s_{ji} \sim N(\bar{z}_i, \sigma^2_{sji}) \text{ and } \epsilon_{ji} \sim N(0, \sigma^2_{\epsilon ji}) \tag{2}$$

with $\sigma_i^2 = \sigma^2_{sji} + \sigma^2_{\epsilon ji}$. The intuition for this specification, unlike the literature that assumes an additive noise signal structure[6] is that investors simplify the state (the uncertainty in the world) rather than amplify it with noise. This formulation of the state as the sum between a signal and some data loss due to compression of the random variable z_i is equivalent to the formulation of an additive noise signal structure (see Cover and Thomas (2006)). Moreover, we also allow investors to learn from prices p_i.

The standard measure of information in information theory is entropy. It is often used to model limited information processing by individuals. Following Grossman and Stiglitz (1980) and Van Nieuwerburgh and Veldkamp (2009), we model the amount of information learned as an entropy reduction. In other words, the quantity of information an investor observes is his capacity, $\tilde{K_j}$. We take $\tilde{K_j}$ as given and focus on the information allocation problem. Prior variances σ_i^2 are not random; they are given. The posterior (conditional) variances $\hat{\sigma_i}^2$, which measure investors' uncertainty about asset payoffs, are also not random; these are choice variables that summarize the investors' optimal information decisions. Learning about asset i makes the conditional variance $\hat{\sigma_i}^2$ lower than its unconditional variance σ_i^2. Because payoffs and signals are independent across assets, the information capacity constraint is a continuous function that maps the product of prior and posterior variances across all assets i into a level of capacity unique for each investor j, $\prod_{i=1}^N \sigma_i / \prod_{i=1}^N \hat{\sigma}_i \leq \tilde{K_j}$. It will become more convenient to work with an equivalent log form of this constraint where $\tilde{K_j} = exp(2K_j)$:

$$\prod_{i=1}^{N} \frac{\sigma_i}{\hat{\sigma}_i} \leq exp(2K_j) \tag{3}$$

We assume that all investors in the economy have the ability to learn about the N risky assets, but some are better than others at learning. Specifically, a measure λ of investors have high capacity for processing information and thus, they are sophisticated, while a measure $1-\lambda$ of investors have low capacity for processing information and are referred to as unsophisticated investors. Thus, $0 \leq K_U \leq K_S \leq \overline{K}_{max}$.[7]

[5]For each asset, investor j decomposes the payoff z_i into a signal component, s_{ji}, and a residual component, ϵ_{ji} that represents data lost due to the compression of the random variable z_i

[6]An additive noise signal structure is modelled as the signal being an unbiased predictor of the stochastic payoff, $s_{ji} = z_i + \epsilon_{ji}$ with $\epsilon_{ji} \sim N(0, \sigma^2_{\epsilon ji})$. For applications of this particular signal structure see Grossman and Stiglitz (1980), Brunnermeier (2001) and Van Nieuwerburgh and Veldkamp (2010).

[7]$\overline{K}_{max}$ needs to be bounded away from infinity, because financial innovation, which is assumed to increase the capacity constraint, becomes meaningless if capacity is unbounded already.

For sophisticated investors, the capacity constraint takes the form $\prod_{i=1}^{N} \sigma_i/\hat{\sigma}_i \leq exp(2K_S)$. Unsophisticated investors have a capacity constraint as $\prod_{i=1}^{N} \sigma_i/\hat{\sigma}_i \leq exp(2K_U)$. Capacity allocation is a one-time decision, not a sequential choice. An investor is not allowed to re-optimize the learning choice based on signal realizations. Rather, we imagine that at time t, the investor decides what news articles about the assets to download, and at time $t+1$ the investor observes these news articles and decides how to invest.

Lastly, we assume a no-forgetting constraint, as in Van Nieuwerburgh and Veldkamp (2010). In other words, the variance of each signal must be non-negative. Without this constraint, an investor could choose to forget what he learned about one asset in order to obtain a more precise signal about another, without violating the capacity constraint. This constraint implies that the posterior variance on any asset i cannot exceed the prior variance:

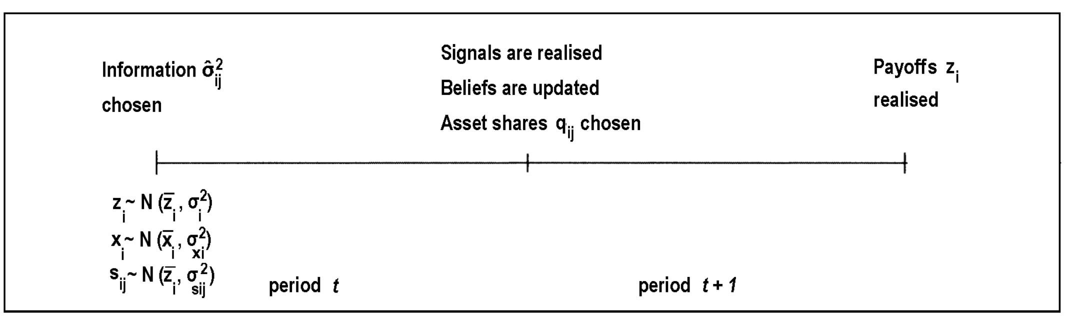
$$0 < \hat{\sigma_i}^2 \leq \sigma_i^2 \tag{4}$$

Portfolio allocation choice. Given his posterior beliefs, an investor chooses the quantities of each asset that he chooses to hold, $q_{ji} = [q_{1,1}, q_{1,2}, ..., q_{1,N}]$. The investor takes as given the risk-free return r and the asset prices $[p_1, p_2, ..., p_N]$. In making this choice, the investor is constrained by his budget constraint:

$$W_{t+1,j} = rW_{t,j} + \sum_{i=1}^{N} q_{ji}\,(z_i - rp_i) \tag{5}$$

Timing. In the first subperiod, investors solve the information acquisition problem subject to their information capacity constraint. In the second subperiod, investors learn and choose their portfolio allocations. The sequence of events is depicted in Figure 1.

Figure 1: Sequence of events



2.2 Solution

As standard in the information-based general equilibrium theoretical literature, we solve the model backwards, starting with the investors' portfolio problem in subperiod $t+1$, for a given information structure, and then solving for the information choice in period t.

Portfolio choice. Each investor j chooses to invest quantity q_i in each asset i to maximize:

$$max\ \mathbb{E}_{t+1,j} \frac{1}{1-\rho} W_{t+1,j}^{1-\rho} \tag{6}$$

subject to the budget constraint

$$W_{t+1,j} = rW_{t,j} + \sum_{i=1}^{N} q_{ji} \left(z_i - rp_i \right) \tag{7}$$

where $\mathbb{E}_{2j}$ denotes the expected utility conditional on investor j's information set at $t+1$, wealth $W_{t,j}$ is wealth at period t, and q_{ji} is the quantity invested in asset i by investor j. This terminal wealth has a Normal distribution if the z_i's are Normal.

Proposition 1 The solution to this problem is given by

$$q_i = \frac{(E[z_i|I_j] - rp_i)W_{t,j}}{\rho \hat{\sigma_i}^2} \tag{8}$$

where $E[z_i|I_j]$ is the conditional expectation of asset payoff i conditional on an investor's information set I_j, and $\hat{\sigma_i}^2 = var[z_i|I_j]$ is the expected variance of asset i's payoff conditional on the investor's information set. The proof is in Appendix *A.1.1*.

Information choice. Following Van Nieuwerburgh and Veldkamp (2010), with CRRA preferences and entropy-learning technology, investors use their entire capacity to learn about the different assets. As the total amount of information any given investor has is exogenous, the only question is how to allocate this information among assets. We proceed by first computing ex-ante expected utility, as in Brunnermeier (2001). To ease exposition, denote the unconditional excess return as $R_i = z_i - rp_i$ with conditional mean $\hat{R_{ji}}$, and the ex-ante volatility of $R_{ji} - \hat{R_{ji}}$ as $\hat{V_{ji}}$ expressed below:

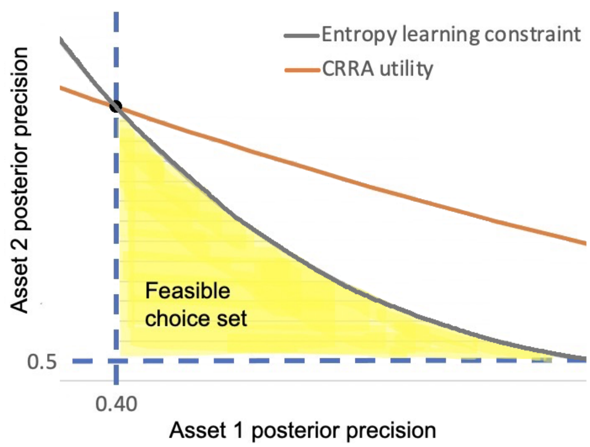
$$\hat{R}_{ji} = E[z_i|I_{ji}] - rE[p_i|I_{ji}] \tag{9}$$

$$\hat{V}_{ji} = Var(R_i - \hat{R}_i) \tag{10}$$

Proposition 2 Each investor specializes by allocating his entire capacity to learning about a single asset.

The complete proof is in Appendix *A.1.2*. Intuitively, an investor's information choice problem can be expressed as a convex objective function decreasing in the choice variable, the posterior standard error $\hat{\sigma}_i$. Thus, an investor invests all his capacity K_j in learning about one single asset and nothing about any other asset as in Figure (2). With CRRA utility and an information capacity constraint, there are increasing returns to devoting additional capacity to learning about a given asset (specialization). Furthermore, the asset that is most valuable to learn about is the one with the highest maximal expected gains (a high expected return and low initial uncertainty). If multiple assets have equal maximal expected gains, the investor randomizes between them.

Figure 2: CRRA preferences and information capacity constraints



Legend: The yellow area represents the set of posterior variances that are feasible and satisfy the entropy – information capacity – learning constraint. The orange line represents the investor's utility. The black dot at the intersection of the feasible information choice set and the investor's utility represents the optimal information choice. This corner solution is due to a convex objective function decreasing in the choice variable, the information choice $\hat{\sigma_1}^{-1}$ and $\hat{\sigma_2}^{-1}$. In this example, the expected return on both assets is the same, but the investor is initially less uncertain about asset 2. The prior precision on asset 2 is 0.5, higher than the prior precision on asset 1 (0.4). The investor specializes in learning all about asset 2 and none about asset 1.

The preference for specialization is an artefact of the learning technology and investors' preferences.[8] We use CRRA preferences with an entropy constraint to obtain specialization,

[8]The most commonly used learning technologies are the entropy constraint, the additive precision constraint, or the linear, concave or convex costs of acquiring precision. Depending on investors' preferences, these technologies give rise to either specialized learning, where the learning portfolio is concentrated or to generalized learning. For example, CARA preferences deliver less of a preference for specialization than CRRA preferences used in much of the rest of the finance literature. CRRA means

because there is extensive empirical evidence that retail investors' portfolios are highly concentrated in a small number of stocks (Barber and Odean (2000), Campbell (2006), Calvet et al. (2007), and Luo et al. (2020)). And the best assets to acquire information about are the ones the investor expects to hold.

Investors' posterior beliefs about payoffs depend on whether they learn or not. For assets that are passively traded, the posterior belief equals the prior belief (i.e., nothing is learned). For assets that are actively traded, the posterior mean equals the received signal and the posterior variance is strictly lower and decreasing in capacity (i.e., the higher the capacity, the more precise the posterior gets). Most important for the results is that the higher the capacity of an investor, the larger the weight that the investor's signal puts on the realization of the signal relative to the prior (i.e., this means that sophisticated investors will respond more strongly to positive/negative signals).

$$\hat{\sigma}_i = \begin{cases} \hat{\sigma}_i = exp(-2K_j)\sigma_i & \text{if the investor learns (if } i \in A) \\ \sigma_i & \text{if the investor does not learn (if } i \notin A) \end{cases} \tag{11}$$

$$\text{and } E[z_{ij}|I_j] = \begin{cases} s_{ij} & \text{if the investor learns (if } i \in A) \\ \bar{z}_i & \text{if the investor does not learn (if } i \notin A) \end{cases} \tag{12}$$

We next present the asset market equilibrium, given the solution to an individual investor's information allocation problem in equations (11) and (12). We follow Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk et al. (2018) in deriving the asset market equilibrium.

Corollary 1 The asset market equilibrium is given by a market clearing condition that equates the demand and supply for each asset

$$(1-\phi_i)\left(\frac{\bar{z}_i - rp_i}{\rho\sigma_i^2}\right) + \int_0^1 \left(\frac{s_{ji} - rp_i}{\rho\sigma_i^2 exp(-2K_j)}\right) dj = x_i \tag{13}$$

and a price for each asset that is a linear combination of the asset's stochastic return and its stochastic supply $p_i = a_i + b_i z_i - c_i x_i$ where

$$a_i = \frac{\bar{z}_i}{r\left(1+\phi_i\mathcal{C}\right)}; \quad b_i = \frac{\phi_i\mathcal{C}}{r\left(1+\phi_i\mathcal{C}\right)}; \quad c_i = \frac{\rho\sigma_i^2}{r\left(1+\phi_i\mathcal{C}\right)} \tag{14}$$

that absolute risk aversion fluctuates, depending on the investor's realized wealth. This fluctuation works to hedge the risk that specialized learning entails (Van Nieuwerburgh and Veldkamp (2010)).

and thus the price of each asset i takes the form

$$p_i = \frac{\bar{z}_i + z_i \phi_i \mathcal{C} - x_i \rho \sigma_i^2}{r\left(1 + \phi_i \mathcal{C}\right)} \tag{15}$$

where ϕ_i denotes the measure of investors that learn about asset i that we still have to solve for, K_j denotes the investor's capacity which can be K_u or K_s for unsophisticated and sophisticated investors respectively, and $\mathcal{C} = \left[\lambda\left(e^{2K_s} - 1\right) + (1-\lambda)\left(e^{2K_u} - 1\right)\right]$ is a proxy for the total capacity in the economy.[9] The proof is in Appendix ($A.1.5$).

We next determine which assets are learned about in equilibrium, and how much information capacity is allocated across these assets. We again follow Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk et al. (2018) in deriving the information allocation.

Proposition 3 All assets that are actively learned about in equilibrium (actively traded) belong to the set of assets with maximal expected gains, $\mathcal{A}$, or in other words, the ones with the highest expected return and lowest initial uncertainty. Not all assets are learned about in equilibrium (but they can be traded passively).

The complete proof is in Appendix $A.1.4$. The optimal information acquisition strategy uses all capacity to learn first about the asset with the highest maximal expected gain (i.e., highest idiosyncratic volatility), then learn (with the capacity left) about the asset with the second highest expected gain (i.e., second highest idiosyncratic volatility), and so on and so forth. An investor's objective function is to choose the variance $\sigma^2_{\epsilon ji}$ to maximize ex-ante utility:

$$\max_{\sigma^2_{\epsilon ji}} \sum_{i=1}^{N} \frac{\hat{\mathcal{V}}_i + \hat{R}^2_{ji}}{\sigma^2_{\epsilon ji}} \tag{16}$$

$$\text{s.t.} \prod_{i=1}^{N} \frac{\sigma_i^2}{\sigma^2_{\epsilon ji}} \leq exp(2K_j) \tag{17}$$

where $\hat{\mathcal{V}}_i = (1-2rb_i)\sigma_i^2 + r^2\sigma^2_{pi}$ and $\hat{R}_i = \bar{z}_i - r\bar{p}_i$ are components common across all investors as these are average ex-ante variances and expectations of posterior excess returns. As we have seen already in Proof ($A.1.2$), the solution to this problem is a corner solution because the objective function is convex and decreasing in the choice variable $\sigma^2_{\epsilon ji}$. Thus, an investor

[9] $\mathcal{C}$ is the summation of the total capacity of sophisticated investors, in measure λ, and the total capacity of unsophisticated investors in measure $1-\lambda$, minus 1. This transformation is easy to work with in our exponential/log derivations, and does not materially change any of our results.

allocates his entire capacity to learning about a single asset and all assets that are actively traded in equilibrium belong to the set of assets with maximal expected gains:

$$\mathcal{A} = \left\{ i | i \in arg \max_i A_i \right\}, \text{ where}$$
$$A_i = \frac{\hat{\mathcal{V}}_i + \hat{R}_{ji}^2}{\sigma_{ji}^2} \tag{18}$$

Proposition 4 Among the assets that are actively traded, $i \in \mathcal{A}$, not all are given the same attention by investors. The market endogenously learns about a select number of assets, and the mass of investors ϕ_i choosing to learn about asset i will vary with the volatility and liquidity of the asset.

The complete proof is in Appendix *A.1.6*. Using the price equation (15), we can derive

$$\hat{R}_i = \frac{\bar{x}_i \rho \sigma_i^2}{(1 + \phi_i \mathcal{C})}; \quad \hat{\mathcal{V}}_i = \frac{\sigma_i^2 (1 + \rho^2 \sigma_i^2 \sigma_{xi}^2)}{(1 + \phi_i \mathcal{C})^2}; \quad \hat{V}_{ji} = \frac{\sigma_i^2 (1 + \rho^2 \sigma_i^2 \sigma_{xi}^2) - \sigma_{\epsilon ji}^2 (1 - \phi_i^2 \mathcal{C}^2)}{(1 + \phi_i \mathcal{C})^2} \tag{19}$$

and substituting them into equation (18), the gain factor becomes

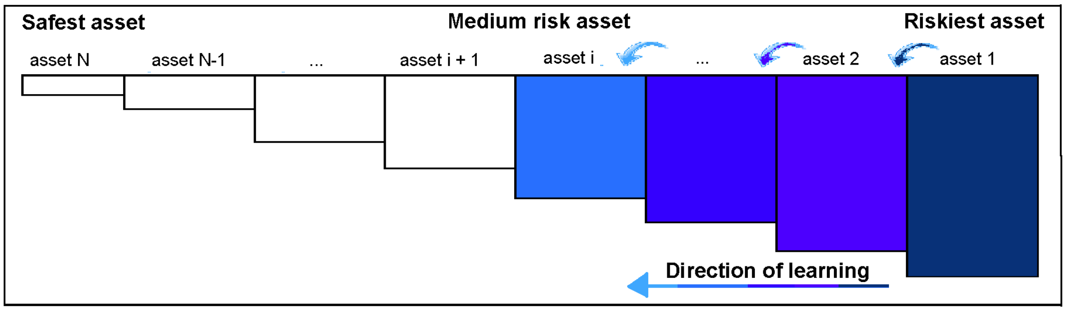
$$A_i = \frac{1 + \rho^2 \sigma_i^2 (\sigma_{xi}^2 + \bar{x}_i^2)}{(1 + \phi_i \mathcal{C})^2} \tag{20}$$

where $\mathcal{C} = \left[\lambda \left(e^{2K_s} - 1\right) + (1 - \lambda) \left(e^{2K_u} - 1\right)\right]$ is the total capacity in the economy. Since $\phi_i \geq 0$ and $\mathcal{C} > 0$, the gain factor, derived in equation (20), is decreasing in the total capacity of the economy, $\mathcal{C}$, as well as in the mass ϕ_i of investors who learn about asset i. Intuitively, the more investors learn, the lower the gains from trading the asset. Investors prefer to trade in an environment where there are few informed investors. However, the gain factor increases with the volatility of the asset σ_i^2, as well as with the volatility of the stochastic supply σ_{xi}^2. Intuitively, the more noise traders there are, the higher the gains from trading with them. Likewise, the riskier the asset, the higher the gains from trade.

Corollary 2 Investors start learning about a first asset with the highest idiosyncratic risk $\sigma_i^2(\sigma_{xi}^2 + \bar{x}_i^2)$. As capacity $\mathcal{C}$ increases, investors start learning about new assets in a decreasing order of idiosyncratic risk, going from the riskiest asset they can afford into less risky assets.

Figure (3) displays the order of assets in which investors learn.

Figure 3: Learning equilibrium. Investors first learn about the most volatile asset, and then expand their learning towards the next highest volatility asset and so on.



Legend: Investors start learning about a first asset with the highest idiosyncratic risk (i.e. the riskiest, most volatile asset). As capacity increases, investors start learning about new assets in a decreasing order of idiosyncratic risk, going from the riskiest asset into less risky assets (moving in the left direction of learning).

The proof is in Appendix A.1.7. Intuitively, investors start learning about the asset with the highest gain, which is the asset with the highest idiosyncratic risk. When aggregate information capacity increases, investors move down the pecking-order of learning into the second highest idiosyncratic risk asset, and so on and so forth, due to strategic substitutability. At low capacity, both investor types learn about the most volatile asset. Yet, as capacity grows, the gains from learning about this asset decline (through a general equilibrium effect that works through prices), which pushes some investors to learn about less volatile assets. This happens as part of pinning down the market equilibrium, moving from one asset to another as greater demand raises the price of that asset, reducing the benefit of holding that asset. The threshold that pins down the selection of investors who learn about the first $1 \leq n \leq N$ assets is given below.

Corollary 3. The selection of investors who learn about the first $n \leq N$ assets is pinned down by

$$\phi_{n+1} = ... = \phi_N = 0 \tag{21}$$

$$\phi_i = \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} \left(\frac{1}{\mathcal{C}_n} + \frac{1 - \frac{1}{\mathcal{C}_n} \sum_{i=1}^{n \leq N} \left(\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1 \right)}{\sum_{i=1}^{n \leq N} \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}} \right) - \frac{1}{\mathcal{C}_n} \tag{22}$$

$$\phi_1 = \frac{1 - \frac{1}{\mathcal{C}_n} \sum_{i=1}^{n \leq N} (\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1)}{\sum_{i=1}^{n \leq N} \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}} \tag{23}$$

where $i \in \{2, 3, ..., n\} \subset \mathcal{A}$, ρ denotes the average risk-aversion, $\phi_i \geq 0$ denotes the measure of investors who learned about asset i, $\Sigma_i = \sigma_i^2(\sigma_{xi}^2 + \bar{x}_i^2)$ denotes the idiosyncratic volatility of asset i, and $\mathcal{C}_n = \left[\lambda\left(e^{2K_s} - 1\right) + (1-\lambda)\left(e^{2K_u} - 1\right)\right] \geq \mathcal{C}_1$ is the total capacity in the economy in which $1 < n \leq N$ assets are learned about in equilibrium. The proof is in Appendix *A*.1.8.

Lastly, denote the sets of sophisticated and unsophisticated investors that learn about asset i as M_{Sj} and M_{Uj} respectively. These sets are of measure $\lambda\phi_i$ and $(1-\lambda)\phi_i$, where ϕ is defined in terms of the primitives of this model as in equation (21). This completes the full characterization of the equilibrium of this economy.

2.3 Theoretical predictions

Given that we have characterized the full equilibrium in the previous section, we now turn to predictions related to heterogeneity in capital income and to the impact that progress in financial information technologies have on heterogeneity in capital income and portfolio holdings in the context of this particular theoretical setting.

Prediction 1. Heterogeneity in information sophistication leads to differences in realized and expected capital income, as well as in expected portfolio holdings. That is, if $K_S > K_U$, then $\sum_{i=1}^N \Pi_{S,i,t} > \sum_{i=1}^N \Pi_{U,i,t}$, and $\sum_{i=1}^N E[\Pi_{S,i,t}] > \sum_{i=1}^N E[\Pi_{U,i,t}]$ and $E[H_{S,i,t}/\lambda] \geq E[H_{S,i,t}/(1-\lambda)]$.

The complete proof is in Appendix *A*.2.1. Capital income heterogeneity is given by the difference in profits between sophisticated and unsophisticated investors

$$\sum_{i=1}^{n} \Pi_{S,i,t} - \Pi_{U,i,t} = \sum_{i=1}^{n} \left[\frac{\phi_i(e^{2K_S} - e^{2K_U})(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{24}$$

where $\Pi_{S,i,t}$ and $\Pi_{U,i,t}$ denote the per capita average profit of sophisticated and unsophisticated investors respectively, ϕ_i is the mass of investors learning about asset i, K_S and K_U are the capacities of sophisticated and unsophisticated investors. Thus, if $K_S > K_U$ this implies $\sum_{i=1}^n \Pi_{S,i,t} > \sum_{i=1}^n \Pi_{U,i,t}$.

Heterogeneity in portfolio composition is given by differences in holding levels of a given asset by sophisticated and unsophisticated investors.

$$H_{S,i,t} = \lambda\left[\frac{(\bar{z}_i - rp_{i,t}) + \phi_i(e^{2K_S} - 1)(z_i - rp_{i,t})}{\rho\sigma_i^2}\right] \tag{25}$$

$$H_{U,i,t} = (1-\lambda)\left[\frac{(\bar{z}_i - rp_{i,t}) + \phi_i(e^{2K_U}-1)(z_i - rp_{i,t})}{\rho\sigma_i^2}\right] \tag{26}$$

Note that $H_{S_i,t}$ and $H_{U,i,t}$ are the holding levels of asset i for each investor type at time t. While both types of investors make the same returns on the passive assets and have the same holdings per capita of passive assets, sophisticated investors make higher expected returns and have higher expected holdings per capita of active assets (i.e., the ones with maximal expected gains) relative to unsophisticated investors. Thus, if $K_S > K_U$ this implies $E[H_{S,i,t}/\lambda] > E[H_{S,i,t}/(1-\lambda)]$, which implies $\sum_{i=1}^{N} E[\Pi_{S,i,t}] > \sum_{i=1}^{N} E[\Pi_{U,i,t}]$. Heterogeneity in information capacity between investors (the fact that $K_S > K_U$) drives differences in capital returns. Sophisticated investors generate higher capital returns than unsophisticated investors because of two reasons: (i) they achieve higher profits because they hold a different average portfolio (i.e., sophisticated investors have better information to identify profitable assets, so they invest in high risk - high return assets more), and (ii) they benefit more from positive shocks and are more sheltered from negative shocks (i.e. because they tilt their portfolios towards profitable assets more than unsophisticated investors do). In other words, sophisticated investors, relative to unsophisticated investors, learn more so they gain more (and lose less) from positive (negative) shocks relative to their prior expectations.

Prediction 2. An absolute (symmetric) increase in information sophistication leads to growing expected capital income heterogeneity. That is, if both K_S and K_U increase proportionally by the same per cent difference, $K'_S = (1+\xi)K_S$ and $K'_U = (1+\xi)\gamma K_S$, where $0 < \xi \leq 1$ and $0 < \gamma < 1$ expected income heterogeneity increases, $d\left(E[\Pi'_{S,i,t}] - E[\Pi'_{U,i,t}]\right)/dK_S > 0$.

The complete proof and other derivations are in Appendix A.2.2. The expected capital income differential increases with the overall growth in aggregate market sophistication, which can be interpreted as general progress in information processing technology. This holds even when relative sophistication stays constant. The intuition for this is that the more an investor learns and knows, the easier it is for this investor to learn on the margin. Thus, the effect of achieving even higher profits is strengthened because sophisticated investors already start from a higher level of capacity to process information to begin with.

Prediction 3. A relative (asymmetric) increase in information sophistication leads to greater realized capital income heterogeneity. That is, if K_S and K_U increase by δ_S and δ_U respectively, such that $K'_S = K_S + \delta_S$ and $K'_U = K_U + \delta_U$, where $\delta_S > \delta_U$, realized capital income heterogeneity increases, $\partial\,(\Pi_{S,i,t} - \Pi_{U,i,t})/\partial(e^{2K_S} - e^{2K_U}) > 0$.

As long as $K_S > K_U$ and there is learning in equilibrium about asset i, such that $\phi_i \neq 0$, capital income heterogeneity is increasing in the capacity gap $K_S - K_U$.

$$\frac{\partial\ (\Pi_{S,i,t} - \Pi_{U,i,t})}{\partial(e^{2K_S} - e^{2K_U})} = \left[\frac{\phi_i(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] > 0 \tag{27}$$

The proof is in Appendix A.2.3. When sophisticated investors become even more sophisticated, and the unsophisticated become even less sophisticated, capital income heterogeneity increases. This result is somewhat mechanical. Intuitively, the higher the dispersion in information capacity, $K_S' - K_S > K_U' - K_U$, the better the information that sophisticated investors receive and the more they use it to achieve higher returns.

Prediction 4. An increase in absolute (and relative) sophistication predicts a growing presence of sophisticated investors in risky asset classes. In other words,

$$\frac{dE\left[\frac{H_{S,i}}{\lambda} - \frac{H_{U,i}}{(1-\lambda)}\right]}{dK_S} > 0 \quad \text{and} \quad \frac{dE\left[\frac{H_{S,i}}{\lambda} - \frac{H_{U,i}}{(1-\lambda)}\right]}{d\mathcal{C}} > 0 \tag{28}$$

where $H_{S,i}$ is the holding level of a sophisticated investor of asset i, and $H_{U,i}$ is the holding level of an unsophisticated investor of asset i.

The complete proof is in Appendix A.2.4. This result holds both for an absolute increase in sophistication (i.e., where K_S and K_U grow by the same percentage difference), and also – trivially – for a relative increase (i.e., where K_S and K_U grow by δ_S and $\delta_U < \delta_S$ respectively). The intuition for this result is that sophisticated investors increase their ownership of equities in a specific order: they first start with the most volatile stocks they can afford, and then continue with stocks with medium and lower volatility. Thus, when their sophistication increases, they can afford to learn about and invest in even riskier asset classes.

Prediction 5. An absolute (and relative) increase in sophistication predicts a retrenchment of less sophisticated investors from trading and stock market ownership in general.

This follows from Prediction 4, as $dE\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{1-\lambda}\right]/dK_S > 0$. The intuition for this result is through a general equilibrium effect which involves prices adjusting to an increased demand from sophisticated investors. In particular, investors expand their risky portfolio holdings by moving down in the asset volatility space as total capacity in the economy expands. Unsophisticated investors perceive an informational disadvantage in trading an asset after sophisticated investors enter, because the new price adjusts to the greater demand from these sophisticated investors. Therefore, unsophisticated investors withdraw into safer assets and away from active trading in general. As asset prices increase, $E[p_i]$ is increasing in $\phi_i\mathcal{C}$, while $E[z_i - rp_i]$ is decreasing in $\phi_i\mathcal{C}$, both types of investors see their returns go down, but only unsophisticated investors will choose to decrease their portfolios because their signals are not sufficiently good enough to sustain their previous position as an optimal choice anymore. The proof is in Appendix *A*.2.5.

3 Data Description

The theoretical framework offers rich cross-sectional and time-series predictions on financial returns and portfolio composition of sophisticated and unsophisticated investors. However, empirically assessing the impact of progress in financial information technology on these variables is difficult because of data limitations (i.e., one needs very detailed panel-level household portfolio holdings data that are usually only contained in sensitive tax records with detailed trading information), but mostly because of identification (i.e., one needs local variation in the diffusion of FinTech technologies).

The existing household finance literature has focused mainly on specific FinTech developments, such as robo-advisors. Reher and Sokolinski (2021) find robo-advisors do not benefit poorer households; Rossi and Utkus (2020) find robo-advisors move poorer households into bond holdings and wealthier households in high-fee active mutual funds that experience significant performance gains. In this paper, we take a broader perspective and analyse the interaction between investors (with different degrees of sophistication) and financial intermediaries offering heterogeneous forms of digital financial services.

Our empirical analysis uses a unique dataset merging information on household and bank-specific characteristics. Data on households are gathered from Banca d'Italia's Italian Survey on Household Income and Wealth (SHIW) from 2004 to 2020.[10] This survey contains socio-demographic characteristics of households members (such as the province of birth and residence, age, gender, education, marital status, work status, and risk aversion), detailed information about all households' income sources and wealth components (real and financial assets, split into deposits, bonds, MMFs, public equity, private equity, etc., and debts). The survey also collects information on households' banks and on their remote connection to banks, which we include as a relevant control to study the overall level of technological development.[11] Figure (4) shows the diffusion of remote banking connections throughout Italy over time.

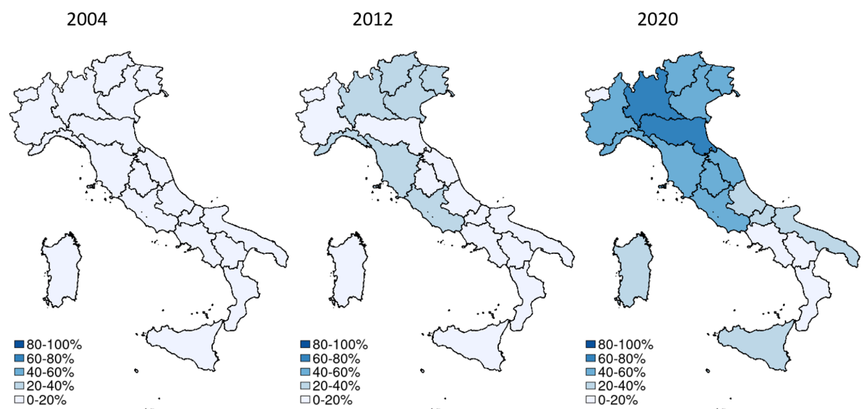
We complement these data with Banca d'Italia's Regional Bank Lending Survey (RBLS) which collects information on the digitalization of the Italian banking sector from a sample of about 280 Italian banks, which cover the Italian credit market almost entirely.[12] Using these

[10] We restrict our analysis to potential investors, i.e. households clients of banks, which represent about 85 per cent of the total sample.

[11] For more details about the methodological aspects of the SHIW see Banca d'Italia (2022).

[12] The sample covers almost 90 per cent of deposits and 85 per cent of loans to firms and households.

Figure 4: Early-stage FinTech: Remote banking connections



Source: Banca d'Italia's Survey on Household Income and Wealth (SHIW).

data we build an indicator that represents a bank's attitude to offer digital services to households. In particular, we use the share of digital services to households, among peer-to-peer payments through mobile, loans for house purchases, consumer credit, and asset management, offered by the household's banks in a specific year.[13]

The average values of the digital financial services indicator for households in 2004, 2012, and 2020 across Italian regions are reported in Figure (5).

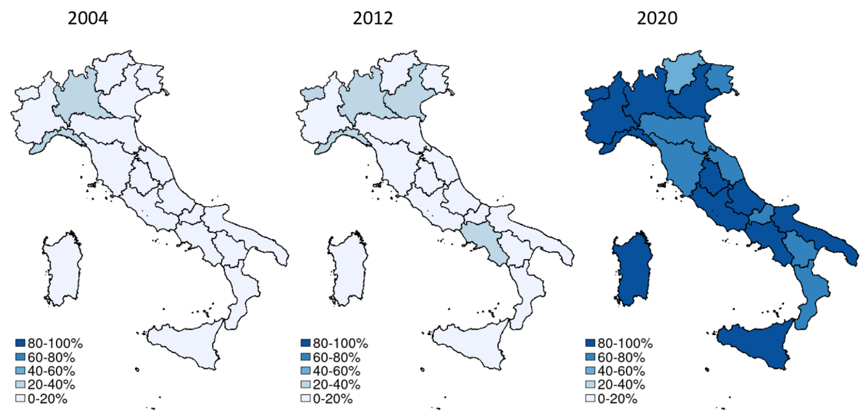
As for the analysis of investors' level of sophistication, we rely on survey information on financial literacy. Financial literacy has been collected in the 2006, 2008, 2010, 2018 and 2020 editions of the SHIW using different sets of questions.[14] Based on this information, we use an indicator of financial literacy calculated as the share of correct answers provided by the respondents. For missing years this indicator has been reconstructed using the predictions of a weighted regression with relevant covariates, such as respondents' socio-demographic characteristics and household's economic conditions.[15] Using this indicator we define sophisticated investors as those having a level of financial literacy in the last quartile; conversely, unsophisticated investors have a level of financial literacy in the first quartile. Table B1 in Appendix

[13]For banks in the SHIW not surveyed in the RBLS, we reconstruct the indicator using a simple regression model. In particular, following the results in Arnaudo et al. (2022), which show that the intensity of technological innovation depends on bank-specific characteristics, we use the prediction of a simple regression model where the digitalization index is estimated as a function of bank capital adequacy (tier1 ratio), profitability (return on equity) and size, including time fixed effects. On the methodology used to construct the combined dataset of the SHIW with the Supervisory Reports of Italian banks see Eramo et al. (2022).

[14]The questions regarded the knowledge of the impact on the purchasing power of inflation, the calculation of compound interests, the awareness of the degree of risk of different kinds of financial assets, the ability to read a bank statement or to distinguish among different types of mortgages.

[15]For more details about the financial literacy indicator see Eramo et al. (2022).

Figure 5: FinTech: digital financial services to households



Source: Banca d'Italia's Regional Bank Lending Survey.

B reports summary statistics of all the variables used in the analysis for the two groups of investors (sophisticated and unsophisticated) and for the total sample. Sophisticated investors have a higher level of income, have more children (when living in couples), are middle-aged (41–65 years), live in larger cities, have a higher level of education, are more likely to be employed or self-employed than unsophisticated investors.

3.1 Stylized facts: matching the model with the data

Before discussing the empirical strategy to test the model's hypotheses, we first present some preliminary evidence on ownership across risky asset classes, market values, stock turnover, and trading intensity and discussion of the cross-sectional implications of the theoretical framework.

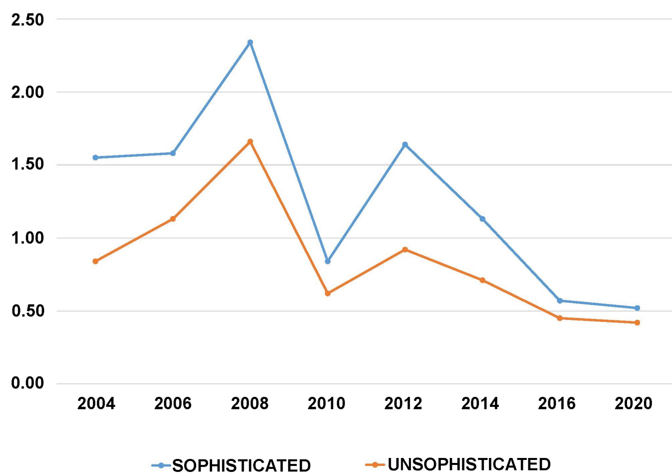
In the model, financial returns increase with investors' sophistication via the "information channel" Arrow (1987). This happens because information matters more to sophisticated investors, as they can better process information, and this can significantly affect returns. This is consistent with empirical evidence from households' tax returns in Scandinavia and India (Fagereng et al. (2020), Di Maggio et al. (2018), and Campbell et al. (2018)).

As predicted by the theoretical model, heterogeneity in information capacity between investors causes differences in their financial returns for Italian households.[16] Figure (6) reports the (unconditional) average rate of financial returns for sophisticated and unsophis-

[16]The SHIW imputes financial returns by assuming that the family held a constant value of financial assets throughout the year (equivalent to the reported amount at the end of the year) and assuming a fixed rate of return for each asset type. For more information on how financial returns are estimated in the SHIW, please refer to Appendix A in Frost et al. (2022).

ticated investors. The difference in financial returns between the two groups of investors is always positive over the sample period: on average around 42 basis points. Interestingly, this difference vanishes in recent years, which are characterised by low or negative interest rates.

Figure 6: Average rates of financial returns by investors' sophistication



Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.

As risk-tolerant wealthy investors take on higher-risk strategies (Chiappori and Paiella, 2011), it is important to control for the different degrees of risk aversion between sophisticated and unsophisticated investors. From the SHIW we can observe directly this investor-specific characteristic:[17] while 69 per cent of unsophisticated investors have high risk aversion, this percentage drops to 46 for sophisticated investors.

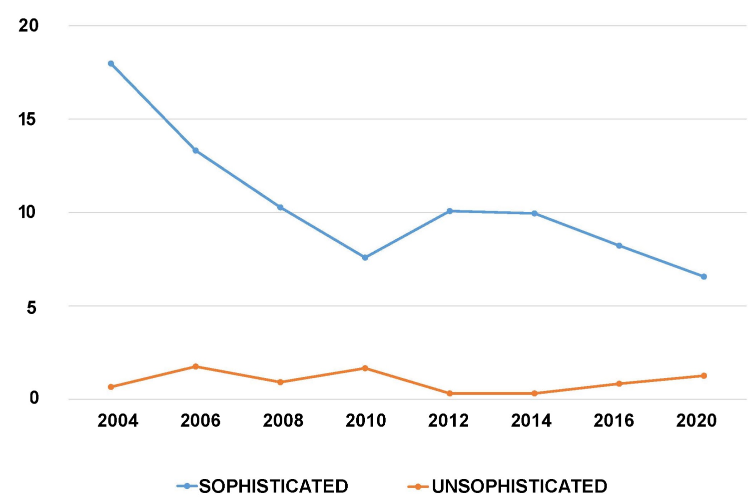
The lower degree of risk aversion for sophisticated investors is associated with a higher percentage of risky investments in their portfolios. Figure (7) shows the (unconditional) average share of risky assets for sophisticated and unsophisticated investors. The difference in the share of risky assets between the two groups of investors is quite remarkable: on average 9 percentage points over the whole sample period. Interestingly, this difference has reduced over time, from 17 percentage points in 2004 to 5 percentage points in 2020. The reduction of the difference in terms of share of risky assets is significant in the period up to the global financial crisis. In the last part of the sample, characterised by very low returns, the difference is instead more stable. To account for the market movements that occurred in the analysed period we include time-fixed effects in our models.

[17]Risk aversion is measured through the following question: *In managing your financial investments, would you say you have a preference for investments that offer: 1 = very high returns, but with a high risk of losing part of the capital (no risk aversion); 2 = a good return, but also a fair degree of protection for the invested capital (low risk aversion); 3 = a fair return, with a good degree of protection for the invested capital (medium risk aversion); 4 = low returns, with no risk of losing the invested capital (high risk aversion).*

An examination of the household portfolio composition by investor financial literacy using SHIW data in Figure (8) confirms that sophisticated households in Italy hold different assets compared to less sophisticated households. Low-risk assets, such as deposits and saving accounts, represent the largest share held by the lowest half of the financial literacy distribution. The lower quartile holds very few medium-risk assets such as government securities or bonds and high-risk assets such as equity shares, fund shares, or other complex securities. The middle classes have more exposure to medium-high risk assets, but low-risk assets still dominate their portfolio. At the top quartile, the share of high-risk assets rises.

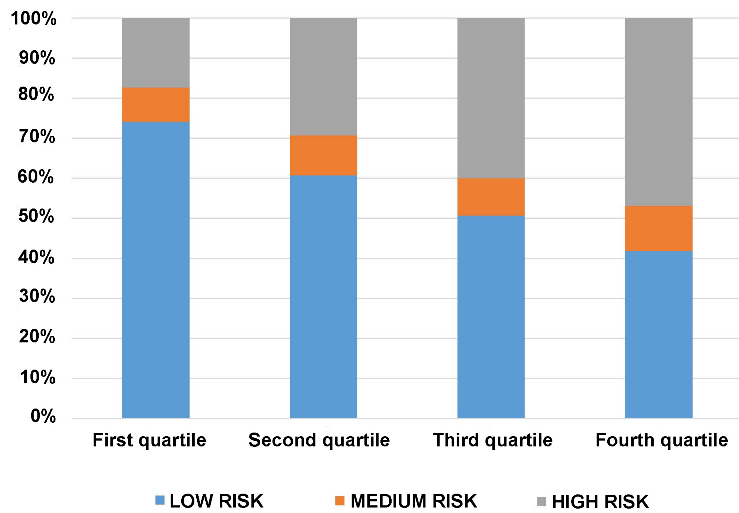
Another prediction of the model is that a general improvement in financial information technology (reflected in the digital services provided by banks) could benefit sophisticated investors more than unsophisticated ones and this could contribute to a decrease in overall household stock-market participation.

Figure 7: Average share of risky assets[1] by investors' sophistication



Legend: (1) Risky financial assets: shares and equity, managed portfolio, funds (in equities, mixed or in foreign currencies), foreign securities and other financial assets (option, futures, etc.). Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.

Figure 8: Portfolio composition across financial literacy quartiles



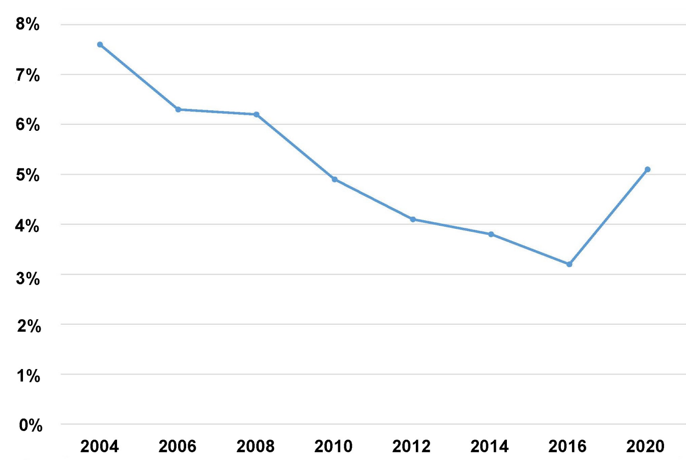
Legend: The figure shows the financial portfolio composition across households' financial literacy quartiles averaged over 2004-20. Low risk: deposits, saving accounts; Medium risk: government securities, bonds, funds (in bonds, money market and liquidity in euros), loans to cooperatives, High risk: shares, equity, managed portfolio, funds (in equities, mixed or in foreign currencies), and other complex securities (options, futures, etc.). Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.

This is consistent with a puzzling phenomenon of the last two decades of a growing retrenchment of less sophisticated, retail investors from trading and stock market ownership in general, as shown in Figure (9). This is also consistent with trends from the US (Wolff (2014) and Mihet (2018)), which are puzzling in light of the fact that direct participation costs have fallen significantly over the last two decades.

In the model, aggregate information technology improvements increase the share of sophisticated investors into high-risk, high return asset classes. Unsophisticated investors perceive their information disadvantage through asset prices and allocate their investments away from the allocations of informed investors. As a result, sophisticated investors earn higher returns, and over time, their financial income diverges from that of unsophisticated investors with relatively lower levels of information.

Figure (10) shows the cumulative financial returns for the two groups of investors (sophisticated and unsophisticated) of 100 Euros invested in 2002. At the end of 2020, sophisticated investors have cumulative financial returns that are 5 percentage points higher than unsophisticated investors, despite the poor performance of the Italian Stock Market in the analysed period (-25%). This model prediction fits well also with empirical evidence for other countries, which documents a presence of sophisticated market players that have doubled in risky asset classes over the last decades (Gompers and Metrick (2001), Garleanu and Pedersen (2018)).

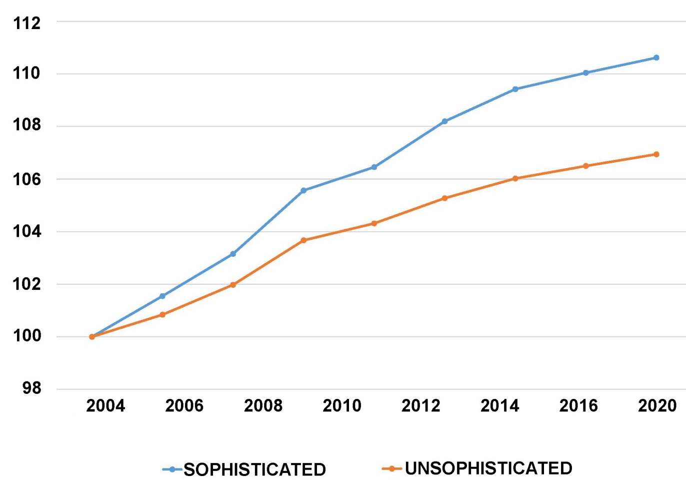
Figure 9: Stock market participation over time



Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.

Figure 10: Cumulative financial returns by investors' sophistication

(100=2002, per cent)



Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.

In the time-series, the dispersion in returns is predicted to grow with advances in financial information technologies. Our theoretical model implies a growing disparity in capital incomes across households. This result is confirmed by computing inequality measures for financial returns in Italy: between 2004 and 2020, the Gini index of financial returns has increased from 0.80 to 0.84 and the quintile ratio (p80/p20) from 16 to 30.

4 Empirical Findings

The patterns described above are broadly in line with the predictions of the theoretical model. However, these findings can only be suggestive as the unconditional means do not control simultaneously for all relevant factors that influence the link between financial technologies, the investors' degree of sophistication, and their portfolio choices and financial returns. This section econometrically analyses the five qualitative predictions of the model discussed in Section 2.3.

Prediction 1. Heterogeneity in information sophistication leads to differences in capital income and portfolio composition.

In the model, we have seen that heterogeneity in information capacity between investors drives differences in capital returns and portfolio composition. Sophisticated investors generate higher capital returns than unsophisticated investors because of two reasons: (i) they achieve higher profits because they hold a different average portfolio (i.e., sophisticated investors have better information to identify profitable assets, so they invest more in high risk-return assets), and (ii) they react better to unexpected shocks. In other words, sophisticated investors, relative to unsophisticated investors, receive higher-quality signals and as a result, they respond more strongly to positive/negative realized excess returns.

In our baseline regressions, we study if sophisticated investors have a systematically higher level of financial returns and invest more in risky assets.

The baseline model specification is the following:

$$Y_{i,j,t} = \alpha_0 \star \text{Remote}_{i,j,t} + \alpha_1 \star \text{Sophisticated}_{i,j,t} + \text{Controls}_{i,j,t} + \theta_j + \xi_t + \epsilon_{i,j,t} \tag{29}$$

where $Y_{i,j,t}$ denotes either financial returns or the portfolio share invested in risky asset classes (such as shares, equity, managed portfolio funds and other complex securities such as options, future, etc.) of household i in the region j at calendar year t. $Remote_{i,j,t}$ is a dummy that takes the value of 1 for investors that have access to remote banking and $Sophisticated_{i,j,t}$ is a dummy that takes the value of 1 for investors in the upper quartile of the distribution of financial education and 0 elsewhere. The vector $Controls_{i,j,t}$, include investor-specific characteristics (gender, education, age, work status, risk aversion, equivalent income quintile, household type, size of the municipality of residence, dummy if the investor is born abroad). θ_j, and ξ_t denote region and time fixed effects. The inclusion of time dummies is particularly relevant to control for general market conditions. The model implies that $\alpha_1 > 0$.

Table 1 reports the results. All models are estimated using a linear regression approach with standard errors adjusted for clustering at the regional level. The first part considers the rate of returns as the dependent variable. Sophisticated investors, other things being equal, earn 13 basis points more than unsophisticated investors. Having access to remote banking increases the rate of returns by an additional 7 basis points. As expected, financial returns are also increasing in the level of education: interestingly, when we control for the level of investor's financial sophistication, the differences among education levels are not so large. An investor with a university degree earns only 7 basis points more than an investor with a primary school education. Financial returns are also increasing in investor's age: controlling for income and other concomitant factors, investors with an age above 65 tend to earn 12 basis points more than those aged 30 or lower. Finally, the results point to the high relevance of the heterogeneity in risk aversion: investors with high risk aversion earn 22 basis points less than those classified with no risk aversion (risk neutral investors).

The second part of Table 1 considers the share of risky assets as the dependent variable.[18] Sophisticated investors, other things being equal, have a share of risky assets that is 4 percentage points larger than unsophisticated investors. Having access to remote banking increases the share of risky assets further by 4 percentage points. Also in this case, the share of risky assets increases with the level of education and age and decreases with risk aversion. Other things being equal, a risk-neutral investor with more than 65 years and with a university degree has a portfolio with a share of risky assets 20 percentage points larger than a risk-averse young investor with very limited education.

The robustness of the above results has been checked in several ways. In particular, we have: (1) controlled for heterogeneity among Italian regions using province dummies; (2) tested for an alternative definition of the sophistication dummy (equal to 1 above the median rather than the upper quartile); (3) used robust standard errors rather than standard errors clustered at the regional level; (4) used in the share of risky assets all risk-bearing financial assets including bonds, investment funds and equity shares. In all cases, results are very similar.

Finally, we tested for the presence of sorting. For example, the results in Table 1 could simply depend upon the fact that sophisticated investors have lower transaction costs than unsophisticated investors or, more in general, the two types of investors match (endogenously) with banks with different characteristics. To test for the possible existence of sorting effects, we have run a linear probability model to represent the probability of an investor having a highly digitalised bank. The right-hand side of the regression includes the dummy *Sophis-*

[18] Similar results are obtained using non-linear models such as Probit and Logit. However, in the baseline regression, we prefer to use linear models because estimates from non-linear models are known to be biased when there are a large number of fixed effects and interaction terms (Lancaster, 2000; Gomila, 2021) and their coefficients are not readily interpretable.

Table 1: Households' rate of returns and portfolio share in risky financial assets[a]

	Dependent variable:	
	(Rate of returns)	(Share of risky assets)
Remote[b]	0.065*** (0.012)	0.036*** (0.005)
Sophisticated[c]	0.126*** (0.010)	0.034*** (0.006)
Education [no schooling]		
Primary school	0.044*** (0.016)	−0.002 (0.003)
Lower secondary school	0.077*** (0.019)	0.005 (0.006)
Upper secondary school	0.125*** (0.020)	0.016*** (0.005)
University degree	0.118*** (0.026)	0.017* (0.010)
Age [30 and under]		
31-40	0.004 (0.016)	0.010*** (0.004)
41-50	0.063*** (0.020)	0.021*** (0.005)
51-65	0.108*** (0.022)	0.027*** (0.008)
Over 65	0.125*** (0.028)	0.031*** (0.010)
Risk aversion [no risk aversion]		
Low risk aversion	−0.107** (0.052)	−0.086** (0.034)
Medium risk aversion	−0.078 (0.048)	−0.110*** (0.034)
High risk aversion	−0.218*** (0.052)	−0.150*** (0.036)
Other covariates [d]	YES	YES
Time fixed effects	YES	YES
Region fixed effects	YES	YES
Observations	52,129	52,129
Pseudo R-square	0.498	0.149

Note: *p<0.1; **p<0.05; ***p<0.01; Standard errors (clustered by region) in brackets.
Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.
[a] Risky financial assets: shares and equity, managed portfolio, funds (in equities, mixed or in foreign currencies), foreign securities and other complex financial assets (option, futures, etc.)
[b] Dummy equal to 1 for investors that have access to remote banking.
[c] Dummy equal to 1 for investors in the upper quartile of the financial literacy distribution.
[d] Other covariates: equivalent income quintile, household type, gender, work status, size of the municipality of residence, born abroad. Individual characteristics refer to the head of the household defined as the main income earner with the exception of risk aversion and financial literacy collected for respondents.

$ticated_{i,j,t}$, a full set of controls for investor-specific characteristics (gender, education, age, work status, risk aversion, equivalent income quintile, household type, size of the municipality of residence, dummy if the investor is born abroad) and time and regional dummies. The results reported in Table B2 indicate that there is no significant correlation between being a sophisticated investor and having a highly digitalised bank.

Prediction 2. An absolute increase in information sophistication leads to a growing expected capital income heterogeneity.

The difference in financial returns between sophisticated and unsophisticated investors increases with the overall growth in the level of financial technology and market sophistication. This should hold even when relative sophistication stays constant. The intuition for this effect is that the more an investor learns and knows, the easier it is for this investor to learn on the margin. Thus, the effect under prediction 1 should be larger for sophisticated investors when the aggregate level of financial technology increases.

The baseline model specification is therefore modified as follows:

$$\begin{aligned}Y_{i,j,t} = \alpha_0 \star \text{Remote}_{i,j,t} + \alpha_1 \star \text{Sophisticated}_{i,j,t} + \alpha_2 \star \text{FinTech}_{i,j,t} + \\ + \beta_1 \star \text{FinTech}_{i,j,t} \star \text{Sophisticated}_{i,j,t} + \\ + \text{Controls}_{i,j,t} + \theta_j + \xi_t + \epsilon_{i,j,t} \end{aligned} \tag{30}$$

where $Y_{i,j,t}$ denotes in this case financial returns of household i in region j at calendar year t. $\text{FinTech}_{j,t}$ is a dummy variable that takes the value of 1 if the investor j's main bank has an index of digital services to households above the regional mean, and 0 otherwise. The interaction term between $\text{FinTech}_{j,t}$ and $\text{Sophisticated}_{i,j,t}$ evaluate the different effects of having a bank with a high level of digitalisation among different types of investors. The prediction of the model implies that $\alpha_0 > 0$, $\alpha_1 > 0$ and $\beta_1 > 0$.

The first column of Table 2 reports the results. Investors with access to remote banking earn, ceteris paribus, 7 basis points more than other investors. Sophisticated investors earn on average 13 basis points more than unsophisticated investors. The effect of having a bank with a high level of digitalisation (FinTech) is of 2 additional basis points for sophisticated investors with respect to other investors.

Prediction 3. A relative increase in information sophistication leads to a growing realized capital income heterogeneity.

When sophisticated investors become even more sophisticated, and/or the unsophisticated become even less sophisticated, capital income heterogeneity increases. The empirical test can be derived by including a specific dummy "Unsophisticated" in equation 29. This dummy takes the value of 1 for those investors in the first quartile of the distribution by financial literacy, and 0 elsewhere.

Table 2: Households' rate of returns and portfolio share in risky financial assets[a]

	Dependent variable:		
	(Rate of returns)		(Share of risky assets)
Remote[b]	0.065*** (0.007)	0.064*** (0.007)	0.036*** (0.002)
Sophisticated[c]	0.116*** (0.008)	0.115*** (0.006)	0.026*** (0.002)
Unsophisticated[d]		−0.042*** (0.007)	−0.0002 (0.002)
FinTech[e]	0.005 (0.006)		−0.005** (0.002)
Interactions:			
Soph.*FinTech	0.020* (0.011)		0.016*** (0.003)
Unsoph.*FinTech			−0.003 (0.003)
Other covariates [f]	YES	YES	YES
Time fixed effects	YES	YES	YES
Regional fixed effects	YES	YES	YES
Observations	52,129	52,129	52,129
Pseudo R-square	0.499	0.499	0.150

Note:*p<0.1; **p<0.05; ***p<0.01; Standard errors (clustered by region) in brackets.
Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.
[a] Risky financial assets: shares and equity, managed portfolio, funds (in equities, mixed or in foreign currencies), foreign securities and other complex financial assets (option, futures, etc.)
[b] Dummy equal to 1 for investors that have access to remote banking.
[c] Dummy equal to 1 for investors in the upper quartile of the financial literacy distribution.
[d] Dummy equal to 1 for investors in the lower quartile of the financial literacy distribution.
[e] Dummy equal to 1 for investors' main bank providing a number of digital services to households above the regional mean.
[f] Other covariates: equivalent income quintile, household type, gender, education, work status, risk aversion, age class, size of the municipality of residence, born abroad. Individual characteristics refer to the head of the household defined as the main income earner with the exception of risk aversion and financial literacy collected for respondents.

The model becomes therefore:

$$Y_{i,j,t} = \alpha_0 \star \text{Remote}_{i,j,t} + \alpha_1 \star \text{Sophisticated}_{i,j,t} + \gamma_1 \star \text{Unsophisticated}_{i,j,t} + \\ + \text{Controls}_{i,j,t} + \theta_j + \xi_t + \epsilon_{i,j,t} \quad (31)$$

In this model, the coefficients of the dummies *Sophisticated* and *Unsophisticated* indicate the difference in financial returns with respect to investors with an intermediate level of financial literacy (those in the second and third quartile of the distribution). The prediction of the model implies that α_1 should be significantly greater than 0 while γ_1 should be significantly lower than 0.

The second column of Table 2 reports the results. Financial returns of an unsophisticated investor that moves to an intermediate level of financial literacy increase by 4 basis points. Moving from an intermediate to sophisticated determines a further increase by 12 basis points.

Prediction 4. An increase in absolute (and relative) sophistication predicts a growing expected presence of sophisticated investors in risky asset classes.

Sophisticated investors increase their ownership of assets in a specific order: they first start with the riskiest assets and then continue with assets with medium and lower risk. The empirical test of this prediction derives from the following equation:

$$\begin{aligned}
Y_{i,j,t} =& \alpha_0 \star \text{Remote}_{i,j,t} + \alpha_1 \star \text{Sophisticated}_{i,j,t} + \alpha_2 \star \text{Unsophisticated}_{i,j,t} \\
& + \gamma_1 \star \text{FinTech}_{i,j,t} + \beta_1 \star \text{FinTech}_{i,j,t} \star \text{Sophisticated}_{i,j,t} + \\
& + \beta_2 \star \text{FinTech}_{i,j,t} \star \text{Unsophisticated}_{i,j,t} + \\
& + \text{Controls}_{i,j,t} + \theta_j + \xi_t + \epsilon_{i,j,t} \qquad (32)
\end{aligned}$$

where $Y_{i,j,t}$ denotes the share of risky assets of household i in region j at year t. The prediction of the model implies that $\alpha_1 > 0$, $\alpha_2 < 0$, $\beta_1 > 0$, and $\beta_2 < 0$.

The third column of Table 2 reports the results. Sophisticated investors have on average 3 percentage points more of risky assets. The effect of having a bank with a high level of digitalisation (FinTech) is statistically different for sophisticated and unsophisticated investors. Sophisticated investors operating with a FinTech bank have a share of risky assets that is 1 percentage larger ($0.016 - 0.005 = 0.011^{***}$), while the effect on unsophisticated investors is negative for around half a percentage point (-0.005^{**}).

Prediction 5. An absolute (and relative) increase in sophistication predicts a retrenchment of less sophisticated investors from trading and stock market ownership in general.

This prediction implies that unsophisticated investors reduce (relative to the sophisticated ones) their risky portfolio holdings by moving down in the asset volatility distribution as absolute financial technology capacity expands. Unsophisticated investors perceive an informational disadvantage in trading an asset after sophisticated investors enter. Therefore, unsophisticated investors will retrench into less risky assets and from active trading in general.

The baseline model specification is therefore modified as follows:

$$Y_{i,j,t} = \alpha_0 \star \text{Remote}_{i,j,t} + \alpha_1 \star \text{Unsophisticated}_{i,j,t} + \alpha_2 \star \text{FinTech}_{i,j,t} +$$

$$+ \beta_1 \star \text{FinTech}_{i,j,t} \star \text{Unsophisticated}_{i,j,t} +$$
$$+ \text{Controls}_{i,j,t} + \theta_j + \xi_t + \epsilon_{i,j,t} \quad (33)$$

where $Y_{i,j,t}$ denotes the share of risky assets of household i in region j at calendar year t.

The results reported in the first column of Table 3 indicate that unsophisticated investors have, on average, 1 percentage point less of risky assets. The effect of having a FinTech bank with a high level of digitalisation is statistically different for unsophisticated investors with respect to other investors. Unsophisticated investors operating with a FinTech bank have a share of risky assets that is in practice equal to other investors not operating with a FinTech bank (0.006-0.008=-0.002, not statistically significant), while the effect of having a FinTech bank on the share of risky assets for other investors is positive for around 1 percentage point (0.006^{***}).

The second column of Table 3 presents the results of a different test, where we do not consider the specific characteristics of the main bank of the investor but the general characteristics of banks operating in the region where the investor resides. In this case, the dummy *FinTech* is replaced by a regional indicator of the level of bank digitalisation (*RegionFinTech*). This indicator is given by the regional average of the share of digital services to households provided by investors' main banks representing the general supply capacity of FinTech technology services in a geographical area. Moreover, the use of a share rather than a dummy (*FinTech*) allows us to include a quadratic term in the specification.

The model becomes:

$$\begin{aligned} Y_{i,j,t} = \alpha_0 \star \text{Remote}_{i,j,t} + \alpha_1 \star \text{Unsophisticated}_{i,j,t} + \alpha_2 \star \text{RegionFinTech}_{j,t} + \\ + \beta_1 \star \text{RegionFinTech}_{j,t} \star \text{Unsophisticated}_{i,j,t} + \\ + \beta_2 \star (\text{RegionFinTech}_{j,t} \star \text{Unsophisticated}_{i,j,t})^2 + \\ + \text{Controls}_{i,j,t} + \xi_t + \epsilon_{i,j,t} \quad (34) \end{aligned}$$

The results in the second column of Table 3 indicate that the share of risky assets for unsophisticated investors is always negatively correlated with the regional level of FinTech. The effect on the share of risky assets for plausible values of the regional bank digitalisation index (0.04-0.90) is between -1.0 and -2.4 percentage points.

Table 3: Households' portfolio share in risky financial assets[a]

	Dependent variable:	
	(Share of risky assets)	
Remote[b]	0.045*** (0.002)	0.046*** (0.002)
Unsophisticated[c]	−0.011*** (0.002)	−0.028*** (0.005)
FinTech[d]	0.006*** (0.002)	
RegionFinTech[e]		0.086*** (0.009)
Interactions:		
Unsophisticated*FinTech	−0.008*** (0.003)	
Unsophisticated*RegionFinTech		0.087*** (0.028)
(Unsophisticated*RegionFinTech)^2		−0.092*** (0.030)
Other covariates [f]	YES	YES
Time fixed effects	YES	YES
Regional fixed effects	YES	NO
Observations	52,129	52,129
Pseudo R-square	0.126	0.128

Note:*p<0.1; **p<0.05; ***p<0.01; Standard errors in brackets.

Source: Banca d'Italia's Survey on Household Income and Wealth 2004-20.

[a] Risky financial assets: shares and equity, managed portfolio, funds (in equities, mixed or in foreign currencies), foreign securities and other complex financial assets (option, futures, etc.)

[b] Dummy equal to 1 for investors that have access to remote banking.

[c] Dummy equal to 1 for investors in the lower quartile of the financial literacy distribution.

[d] Dummy equal to 1 for investors' main bank providing a number of digital services to households above the regional mean.

[e] Regional mean of the index of digital services to households provided by investors' main bank.

[f] Other covariates: equivalent income quintile, household type, gender, education, work status, risk aversion, age class, size of the municipality of residence, born abroad. Individual characteristics refer to the head of the household defined as the main income earner with the exception of risk aversion and financial literacy collected for respondents.

5 Conclusions

By offering an easier access to different savings and investment opportunities, and a multitude of data sources for informed investment, FinTech promises to democratize the investment management sector and level-out the playing field. On the other hand, investors with a different degree of financial sophistication could benefit differently from financial technological advances, with effects on financial income inequality that are *a priori* uncertain.

In this paper, we present a simple micro-founded model that derives testable predictions on the links between financial technologies, investors' degree of sophistication, and their portfolio choices and financial returns. Using microdata from the Survey on Household Income and Wealth conducted by Banca d'Italia over the period 2004-20, we test the theoretical predictions of the model and find that the gaps in financial returns and share of risky assets between sophisticated and unsophisticated investors increase with progress in financial technology. This means that inequality is reduced only if financial technology is accessible to everyone, and if all investors have the same capacity to use it.

References

Abraham, Facundo, Sergio L. Schmukler, and Jose Tessada, "Robo-Advisors : Investing through Machines," Research and Policy Briefs 134881, The World Bank 2019.

Acemoglu, Daron, "Technical change, inequality, and the labor market," *Journal of economic literature*, 2002, *40* (1), 7–72.

Arnaudo, Davide, Silvia Del Prete, Cristina Demma, Marco Manile, Andrea Orame, Marcello Pagnini, Carlotta Rossi, Paola Rossi, and Giovanni Soggia, "The digital trasformation in the Italian banking sector," Bank of Italy Occasional Paper 682 2022.

Arrow, Kenneth J., "The Demand for Information and the Distribution of Income," *Probability in the Engineering and Informational Sciences*, 1987, *1* (1), 3–13.

Auer, Raphael, Cornelli Giulio, Doerr Sebastian, Frost Jon, and Gambacorta Leonardo, "Crypto trading and Bitcoin prices: evidence from a new database of retail adoption," BIS Working Papers 1049, Bank for International Settlements 2022.

Azarmsa, Ehsan, "Investment Sophistication and Wealth Inequality," Working Paper, SSRN February 2019. retrieved from SSRN on October 24, 2019.

Banca d'Italia, "The Survey on Household Income and Wealth," Methods and Sources: metholoogical notes, Bank of Italy 2022.

Barber, Brad M. and Terrance Odean, "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance*, 2000, *55* (2), 773–806.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer-lending discrimination in the FinTech Era," *Journal of Financial Economics*, 2022, *143* (1), 30–56.

Becker, Gary, *Human Capital and the Personal Distribution of Income: An Analytical Approach* Woytinsky lecture, University of Michigan: Institute of Public Administration, 1967.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri, "On the Rise of FinTechs: Credit Scoring Using Digital Footprints," *Review of Financial Studies*, 2020, *33* (7), 2845–2897.

Brunnermeier, Markus K., *Asset pricing under asymmetric information: bubbles, crashes, technical analysis, and herding*, Oxford University Press, 2001.

Calvet, Laurent E, John Y Campbell, and Paolo Sodini, "Down or out: Assessing the welfare costs of household investment mistakes," *Journal of Political Economy*, 2007, *115* (5), 707–747.

Campbell, John Y., "Household Finance," *The Journal of Finance*, 2006, *61* (4), 1553–1604.

Campbell, John Y, Tarun Ramadorai, and Benjamin Ranish, "Do the Rich Get Richer in the Stock Market? Evidence from India," WP 24898, NBER August 2018.

Chiappori, Pierre-André and Monica Paiella, "Relative Risk Aversion Is Constant: Evidence From Panel Data," *Journal of the European Economic Association*, 2011, *9* (6), 1021–52.

Cover, Thomas M. and Joy A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, USA: Wiley-Interscience, 2006.

Deuflhard, Florian, Dimitris Georgarakos, and Roman Inderst, "Financial literacy and savings account returns," *Journal of the European Economic Association*, 2019, *17* (1), 131–164.

Eramo, Ginette, Romina Gambacorta, and Marco Langiulli, "Households' portfolio choices and banks' economic and financial conditions," Bank of Italy, mimeo 2022.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri, "Heterogeneity and persistence in returns to wealth," *Econometrica*, 2020, *88* (1), 115–170.

Frost, Jon, Leonardo Gambacorta, and Romina Gambacorta, "On the nexus between wealth inequality, financial development and financial technology," *Journal of Economic Behavior & Organization*, 2022, *202*, 429–451.

Garleanu, Nicolae and Lasse H. Pedersen, "Efficiently Inefficient Markets for Assets and Asset Management," *The Journal of Finance*, 2018, *73* (4).

Gomila, Robin, "Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis.," *Journal of Experimental Psychology: General*, 2021, *150* (4), 700.

Gompers, Paul A. and Andrew Metrick, "Institutional Investors and Equity Prices*," *The Quarterly Journal of Economics*, 02 2001, *116* (1), 229–259.

Grossman, Sanford and Joseph Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 1980, *70* (3), 393–408.

Jaumotte, Florence, Subir Lall, and Chris Papageorgiou, "Rising income inequality: technology, or trade and financial globalization?," *IMF economic review*, 2013, *61* (2), 271–309.

Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens, "Investor sophistication and capital income inequality," *Journal of Monetary Economics*, 2018.

Katona, Zsolt, Marcus Painter, Panos N. Patatoukas, and Jean Zeng, "On the Capital Market Consequences of Alternative Data: Evidence from Outer Space," Working Paper 3222741, SSRN 2018.

Lancaster, Tony, "The incidental parameter problem since 1948," *Journal of econometrics*, 2000, *95* (2), 391–413.

Luo, Cheng, Enrichetta Ravina, and Luis Viceira, "Retail Investors' Contrarian Behavior Around News and the Momentum Effect," *SSRN Electronic Journal*, 01 2020.

Maggio, Marco Di, Amir Kermani, and Kaveh Majlesi, "Stock Market Returns and Consumption," Working Paper 24262, NBER January 2018.

Mihet, Roxana, "Financial Technology and the Inequality Gap," Working Paper 3474720, SSRN April 2018.

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Information Immobility and the Home Bias Puzzle," *The Journal of Finance*, 2009, *64* (3), 1187–1215.

_ and _ , "Information Acquisition and Under-Diversification," *The Review of Economic Studies*, 2010, *77* (2), 779–805.

Peress, Joel, "Wealth, Information Acquisition, and Portfolio Choice," *Review of Financial Studies*, 2004, *17* (3), 879–914.

Reher, Michael and Stanislav Sokolinski, "Automation and Inequality in Wealth Management," Working Paper 3515707, SSRN 2021.

Rossi, Alberto G. and Stephen P. Utkus, "Who Benefits from Robo-advising? Evidence from Machine Learning," Working Paper 3552671, SSRN 2020.

Wolff, Edward N., "Household Wealth Trends in the United States, 1962-2013: What Happened over the Great Recession?," Working Paper 20733, NBER December 2014.

Appendix A Detailed Proofs and Other Results

A.1 Solving for equilibrium

A.1.1 Portfolio choice

Proof of Proposition 1 The solution to this problem is given by

$$q_i = \frac{(E[z_i|I_j] - rp_i)W_{t,j}}{\rho\hat{\sigma}_i{}^2}$$

Proof: Each investor chooses a portfolio allocation to maximize:

$$max\ \mathbb{E}_{2j}\frac{1}{1-\rho}W_{t+1,j}^{1-\rho}$$

subject to the budget constraint

$$W_{t+1,j} = r\left(W_{t,j} - \sum_{i=1}^{N} p_i q_{ji}\right) + \sum_{i=1}^{N} z_i q_{ji}$$

We first compute expected utility. For this, we need to express terminal wealth in log-terms and then use an approximation of the log return.

$$\overline{W_{t+1,j}} = W_{t,j}e^{log\left[r\left(1-\sum_{i=1}^{N} p_i\frac{q_{ji}}{W_{t,j}}\right)+\sum_{i=1}^{N} z_i\frac{q_{ji}}{W_{t,j}}\right]} =$$
$$= W_{t,j}e^{log\left[r+\frac{1}{W_{t,j}}\sum_{i=1}^{N} p_i q_{ji}\frac{z_i - rp_i}{p_i}\right]}$$

The only stochastic term in the above equation is z. We can define

$$F(z_i - rp_i) = log\left[r + \frac{1}{W_{t,j}}\sum_{i=1}^{N} p_i q_{ji}\frac{z_i - rp_i}{p_i}\right] \text{ and}$$
$$G(q_i) = \left[r + \frac{1}{W_{t,j}}\sum_{i=1}^{N} q_{ji}(\bar{z}_i - rp_i)\right]$$

The second-order Taylor approximation of the F function is

$$F(z_i - rp_i) = f(\bar{z}_i - rp_i) + (z_i - \bar{z}_i)F'(\bar{z}_i - rp_i) + \frac{1}{2}(z_i - \bar{z}_i)^2 F''(\bar{z}_i - rp_i) + \mathcal{O}(z_i - rp_i)$$

Plugging in the terms

$$F' = \frac{q_i}{W_{t,j}}\ \frac{1}{\left[r + \frac{1}{W_{t,j}}\sum_{i=1}^{N} q_{ji}(\bar{z}_i - rp_i)\right]}$$

$$F'' = -\frac{q_i^2}{W_{t,j}^2} \; \frac{1}{\left[r + \frac{1}{W_{t,j}} \sum_{i=1}^{N} q_{ji}(\bar{z}_i - rp_i)\right]^2}$$

we get that the Taylor approximation is

$$\begin{aligned} F(z_i - rp_i) = log\left[r + \frac{1}{W_{t,j}} \sum_{i=1}^{N} q_{ji}(\bar{z}_i - rp_i)\right] + (z_i - \bar{z}_i)\frac{q_i}{W_{t,j}} \; \frac{1}{\left[r + \frac{1}{W_{t,j}} \sum_{i=1}^{N} q_{ji}(\bar{z}_i - rp_i)\right]} \\ - \frac{1}{2}(z_i - \bar{z}_i)^2 \frac{q_i^2}{W_{t,j}^2} \; \frac{1}{\left[r + \frac{1}{W_{t,j}} \sum_{i=1}^{N} q_{ji}(\bar{z}_i - rp_i)\right]^2} + \mathcal{O}(z_i - rp_i) \end{aligned}$$

And substituting in the G function, the Taylor approximation of the F function can be written concisely as

$$F(z_i - rp_i) = logG(q_i) + (z_i - \bar{z}_i)\frac{q_i}{W_{t,j}} \; \frac{1}{G(q_i)} - \frac{1}{2}(z_i - \bar{z}_i)^2 \frac{q_i^2}{W_{t,j}^2} \; \frac{1}{G(q_i)^2}$$

Thus, to compute expected utility, we can write

$$\left(e^{[F(z_i - rp_i)]}\right)^{1-\rho} = G(q_i)^{(1-\rho)} e^{(1-\rho)(z_i - \bar{z}_i)\frac{q_i}{W_{t,j}} \; \frac{1}{G(q_i)} - \frac{(1-\rho)}{2}(z_i - \bar{z}_i)^2 \frac{q_i^2}{W_{t,j}^2} \; \frac{1}{G(q_i)^2}}$$

In the above, we can approximate the term $(z_i - \bar{z}_i)^2$ by its expected conditional volatility ${\hat{\sigma}_i}^2$. At the approximation point, we can ignore variation in the term $G(q_i)$ because it is a constant and thus,

$$\begin{aligned} log\mathbb{E}W_{j,t+1}^{1-\rho} &= c_0(1-\rho)\left[\sum_{i=1}^{N} \frac{q_i}{W_{t,j}} \; (E[z_i|I_j] - rp) + (1-\rho)\sum_{i=1}^{N} \frac{q_i^2 {\hat{\sigma}_i}^2}{W_{t,j}} \; \frac{1}{2W_{t,j}^2} - \sum_{i=1}^{N} \frac{q_i^2 {\hat{\sigma}_i}^2}{2W_{t,j}^2}\right] \\ &= c_0(1-\rho)\left[\sum_{i=1}^{N} \frac{q_i}{W_{t,j}} \; (E[z_i|I_j] - rp) - \rho \sum_{i=1}^{N} \frac{q_i^2 {\hat{\sigma}_i}^2}{2W_{t,j}^2}\right] \qquad (35) \end{aligned}$$

Taking the first order condition of this utility function with respect to the portfolio choice variable q_i gives the equation

$$q_i = \frac{(E[z_i|I_j] - rp_i)W_{t,j}}{\rho {\hat{\sigma}_i}^2} \qquad (36)$$

where ${\hat{\sigma}_i}^2 = var[z_i|I_j]$ is the expected variance of asset i's payoff conditional on the investor's information set.

QED.

A.1.2 Information choice

Proof of Proposition 2 Each investor specializes by allocating his entire capacity to learning about a single asset. Not all assets are learned about in equilibrium (but they can be traded passively. All assets actively traded belong to the set of asset with maximal expected gains.

Proof: We can write in matrix form

$$U = \frac{(rW_{t,j})^{1-\rho}}{1-\rho} \; exp\left\{\frac{1-\rho}{\rho}\frac{1}{2}\left[(R-\hat{R})\hat{\Sigma}^{-1}(R-\hat{R}) + 2\hat{R}\hat{\Sigma}^{-1}(R-\hat{R}) + \hat{R}\hat{\Sigma}^{-1}\hat{R}\right]\right\}$$

Taking expectations, this becomes

$$EU = \frac{(rW_{t,j})^{1-\rho}}{1-\rho} \; |I - 2\hat{V}\frac{(1-\rho)}{2\rho}\hat{\Sigma}^{-1}|^{-\frac{1}{2}} exp\left[\frac{(1-\rho)^2}{2\rho^2}\hat{R}\hat{\sigma}^{-1}\left(I - 2\hat{V}\frac{(1-\rho)}{2\rho}\hat{\Sigma}^{-1}\right)^{-1}\hat{V}\hat{R}\hat{\Sigma}^{-1}\right]$$
$$exp\left[+\frac{(1-\rho)}{2\rho}\hat{R}\hat{\Sigma}^{-1}\hat{R}\right]$$

Expressing the determinant and grouping terms, the above becomes

$$EU = \frac{(rW_{t,j})^{1-\rho}}{1-\rho} \; \Pi_i^{-1/2}\left(1 - \hat{V}_i\frac{(1-\rho)}{\rho}\hat{\sigma_i}^{-1}\right)^{-\frac{1}{2}} exp\left[\frac{(1-\rho)}{2\rho}\sum \hat{R}_i^2\hat{\sigma}^{-1}\left(1 + \frac{(1-\rho)}{2\rho}\hat{V}_i\hat{\sigma}_i^{-1}\right)^{-1}\right]$$

Taking the natural log of the negative of the above expression yields

$$\text{constant} + \frac{1}{2}\sum_{i=1}^{N} log\left(1 + \frac{(1-\rho)}{2\rho}\hat{V}_i\hat{\sigma}_i^{-1}\right) + \sum_{i=1}^{N}\frac{(\rho-1)}{2\rho}\frac{\rho\hat{R}_i^2}{\rho\hat{\sigma}_i + (\rho-1)\hat{V}_i}$$

Therefore, an investor's information choice problem is to maximize

$$\max_{\hat{\sigma}_i^{-1}} \; \frac{1}{2}\sum_{i=1}^{N} log\left(1 + \frac{(1-\rho)}{2\rho}\frac{\hat{V}_i}{\hat{\sigma}_i}\right) + \sum_{i=1}^{N}\frac{(\rho-1)}{2\rho}\frac{\rho\hat{R}_i^2}{\rho\hat{\sigma}_i + (\rho-1)\hat{V}_i}$$
$$\text{subject to } \prod_{i=1}^{N}\frac{\sigma_i}{\hat{\sigma}_i} \leq exp(2K_j)$$

The solution is a corner solution, because the objective function is convex and strictly decreasing in the choice variable $\hat{\sigma}_i$.

QED.

A.1.3 Expected utility

Plugging in the posterior uncertainty into an investor's expected utility function and after some manipulation, we obtain

$$\frac{1}{2}\sum_{i=1}^{N} log\left(1+\frac{(1-\rho)}{2\rho\sigma_i}\hat{V}_i exp(2K_{ji})\right)+\sum_{i=1}^{N}\frac{(\rho-1)}{2}\frac{\hat{R}_i^2}{\rho exp(-2K_{ji})\sigma_i+(\rho-1)\hat{V}_i}$$

The expected utility function is clearly increasing in K_j as the first term of the above expression is increasing in K_j and the second term is also increasing in K_j.

To show that increasing returns arise, we can redefine the choice variable to be the amount of entropy capacity devoted to learning about each asset. The investor j chooses $[K_{j1}, K_{j2}, ..., K_{jN}] \geq 0$, where the choice variable measures the increase in precision $\hat{\sigma}_i^{-1} = exp(2K_{ji})\sigma_i^{-1}$ subject to the constraint that $\sum_i K_{ji} = K_j$ and the no-forgetting constraint. The Lagrangian of this problem is

$$\mathcal{L}=\frac{1}{2}\sum_{i=1}^{N} log\left(1+\frac{(1-\rho)}{2\rho\sigma_i}\hat{V}_i exp(2K_{ji})\right)+\frac{1}{2}\sum_{i=1}^{N}\frac{(\rho-1)\hat{R}_i^2}{\rho exp(-2K_{ji})\sigma_i+(\rho-1)\hat{V}_i}+\xi(K_j-\sum_i K_{ji})+\phi_i K_{ji}$$

where ξ and ϕ are the Lagrange multipliers on the capacity constraint and the no-forgetting constraint. The first-order condition of this problem is

$$\frac{\partial U_j}{\partial K_{ji}}=\frac{-(\rho-1)\hat{V}_i exp(2K_{ji})}{\left(2\rho\sigma_i-(\rho-1)\hat{V}_i exp(2K_{ji})\right)}-\frac{(\rho-1)\rho\sigma_i\hat{R}_i^2 exp(-2K_{ji})}{\left(\rho exp(-2K_{ji})\sigma_i+(\rho-1)\hat{V}_i\right)^2}-\xi+\phi_i=0$$

This first order condition does not characterize the solution to the problem, because the second-order condition is not satisfied. The second-order derivative is positive if $\rho > 1$, hence the Lagrangian is a convex function. Thus, by Proposition 2, we know the investor will devote all capacity to learning about the asset with the highest expected gains.

A.1.4 Maximal expected gain A_i

Proposition 3 All assets that are actively learned about in equilibrium (actively traded) belong to the set of assets with maximal expected gains, $\mathcal{A}$, or in other words, the ones with the highest expected return and lowest initial uncertainty. Not all assets are learned about in equilibrium (but they can be traded passively).

Proof: Plugging in the optimal portfolio q_i from equation (36) into expected utility in (35) gives the following indirect utility function

$$U_{t=2,j} = \frac{1}{2\rho} \sum_{i=1}^{N} \frac{(E[z_{ji}|I_j] - rp_i)^2}{\hat{\sigma_{ji}}^2} \tag{37}$$

which is maximized subject to the capacity constraint

$$\prod_{i=1}^{N} \frac{\sigma_i}{\hat{\sigma}_{ji}} \leq exp(2K_j) \tag{38}$$

Taking expectations w.r.t. period 1 and rearranging using the fact that $var\left(E_2[z_{ji}|I_j] - rp_i\right) = E_1\left((E_2[z_{ji}|I_j] - rp_i)^2\right) - E_1\left(E_2[z_{ji}|I_j] - rp_i\right)^2$, alternatively written as $\hat{V}_{ji} = E_1[(E_2[z_{ji}|I_j] - rp_i)^2] - \hat{R}^2_{ji}$ the investor's ex-ante utility is

$$U_{t=1,j} = \frac{1}{2\rho} \sum_{i=1}^{N} \frac{\hat{V}_{ji} + \hat{R}^2_{ji}}{\hat{\sigma_{ji}}^2} \tag{39}$$

Before we can proceed with the information choice, we need to guess and verify the price p_i of each asset. Let us guess and verify that the price of each asset is a linear combination of the asset's stochastic return and its stochastic supply

$$p_i = a_i + b_i z_i - c_i x_i \tag{40}$$

We will solve for the coefficients a_i, b_i and c_i later. For now, we can derive the time $t = 1$ expectations and variances about excess returns

$$\hat{R}_i = \bar{z}_i - r\bar{p}_i, \text{ and will be the same across all investors } j \tag{41}$$

$$\hat{V}_{ji} = var_{t=1,j}\left(E_{t=2,j}[z_i|I_{ji}]\right) + r^2\sigma^2_{pi} - 2r \times cov\left(E_{t=2,j}[z_i|I_{ji}], p_i\right) \tag{42}$$

Note that the ex-ante variance of posterior beliefs is given by

$$var_{t=1,j}\left(E_{t=2,j}[z_i|I_{ji}]\right) = \frac{cov^2(z_i, s_{ji})}{var(s_{ji})} = \frac{var^2(s_{ji})}{var(s_{ji})} = var_{t=1,j}(s_{ji})$$

Moreover, we can write

$$\begin{aligned} cov\left(E_{t=2,j}[z_i|I_{ji}], p_i\right) &= \frac{cov\left(E_{t=2,j}[z_i|I_{ji}], z_i\right) \times cov\left(z_i, p_i\right)}{\sigma^2_i} = \frac{var(s_{ji})) \times b_i\sigma^2_i}{\sigma^2_i} = \\ &= b_i var(s_{ji})) \end{aligned} \tag{43}$$

Plugging (43) into the variance of posterior beliefs (42)

$$
\begin{aligned}
\hat{V}_{ji} &= var_{t=1,j}(s_{ji}) + r^2\sigma_{pi}^2 - 2rb_i var(s_{ji})) = \\
&= (1-2rb_i)var_{t=1,j}(s_{ji}) + r^2\sigma_{pi}^2 = \\
&= (1-2rb_i)\sigma_{sji}^2 + r^2\sigma_{pi}^2 = (1-2rb_i)(\sigma_i^2 - \sigma_{\epsilon ji}^2) + r^2\sigma_{pi}^2 = \\
&= \underbrace{(1-2rb_i)\sigma_i^2 + r^2\sigma_{pi}^2}_{\hat{\mathcal{V}}_i} - (1-2rb_i)\sigma_{\epsilon ji}^2 = \\
&= \hat{\mathcal{V}}_i - (1-2rb_i)\sigma_{\epsilon ji}^2
\end{aligned}
\tag{44}
$$

where $\hat{\mathcal{V}}_i = (1-2rb_i)\sigma_i^2 + r^2\sigma_{pi}^2$ is a component common across all investors.

The investor's objective function (39) then becomes

$$
\begin{aligned}
U_{t=1,j} &= \frac{1}{2\rho}\sum_{i=1}^{N}\frac{\hat{V}_{ji} + \hat{R}_{ji}^2}{\hat{\sigma_{ji}}^2} = \frac{1}{2\rho}\sum_{i=1}^{N}\frac{\hat{\mathcal{V}}_i - (1-2rb_i)\sigma_{\epsilon ji}^2 + \hat{R}_{ji}^2}{\sigma_{\epsilon ji}^2} = \\
&= \frac{1}{2\rho}\sum_{i=1}^{N}\frac{\hat{\mathcal{V}}_i + \hat{R}_{ji}^2}{\sigma_{\epsilon ji}^2} - \frac{1}{2\rho}\sum_{i=1}^{N}(1-2rb_i)\sigma_{\epsilon ji}^2\frac{1}{\sigma_{\epsilon ji}^2} \\
&= \frac{1}{2\rho}\sum_{i=1}^{N}\frac{\hat{\mathcal{V}}_i + \hat{R}_{ji}^2}{\sigma_{\epsilon ji}^2} - \underbrace{\frac{1}{2\rho}\sum_{i=1}^{N}(1-2rb_i)}_{\text{can ignore as independent of } \sigma_{\epsilon ji}^2}
\end{aligned}
$$

Thus the investor's objective function is to choose the variance $\sigma_{\epsilon ji}^2$ to maximize ex-ante utility:

$$
\max_{\sigma_{\epsilon ji}^2}\sum_{i=1}^{N}\frac{\hat{\mathcal{V}}_i + \hat{R}_{ji}^2}{\sigma_{\epsilon ji}^2} \tag{45}
$$

$$
\text{s.t. } \prod_{i=1}^{N}\frac{\sigma_i^2}{\sigma_{\epsilon ji}^2} \leq exp(2K_j) \tag{46}
$$

where $\hat{\mathcal{V}}_i = (1-2rb_i)\sigma_i^2 + r^2\sigma_{pi}^2$ and $\hat{R}_i = \bar{z}_i - r\bar{p}_i$ are components common across all investors as these are average ex-ante variances and expectations of posterior excess returns.

As we have seen already in Proof (*A*.1.2), the solution to this problem is a corner solution because the objective function is convex and decreasing in the choice variable $\sigma_{\epsilon ji}^2$. Thus, an investor allocates his entire capacity to learning about a single asset and all assets that are actively traded in equilibrium belong to the set of assets with maximal expected gains:

$$
\mathcal{A} = \left\{ i | i \in arg\max_i A_i \right\}, \text{ where}
$$

$$A_i = \frac{\hat{\mathcal{V}}_i + \hat{R}_{ji}^2}{\sigma_{ji}^2} \tag{47}$$

QED.

A.1.5 Asset market equilibrium

We guessed above and we verify in this section that the price of asset i is a linear function of the stochastic asset payoff and the stochastic supply.

$$p_i = a_i + b_i z_i - c_i x_i$$

The market clearing condition for each asset is such that the demand from investors who learn (who can be sophisticated and unsophisticated) plus the demand from investors who do not learn at all has to equal the total supply for each asset. Let ϕ_i denote the mass of investors who learn about asset i. Plugging in the optimal portfolio choice (36) and investors' posterior beliefs about payoffs, (41), which depend on whether they learn or not, the market clearing condition equating demand and supply is

$$\underbrace{(1-\phi_i)\left(\frac{\bar{z}_i - rp_i}{\rho\sigma_i^2}\right)}_{\text{do not learn}} + \underbrace{\int_{\mathcal{M}_{Sj}}\left(\frac{s_{ji} - rp_i}{\rho\sigma_i^2 exp(-2K_S)}\right)dj + \int_{\mathcal{M}_{Uj}}\left(\frac{s_{ji} - rp_i}{\rho\sigma_i^2 exp(-2K_U)}\right)dj}_{\text{sophisticated and unsophisticated who learn}} = x_i \tag{48}$$

Bayes' law gives that each signal received by an investor j will be a weighted average between the prior and its true realization:

$$E(s_{ji}|z_i) = exp(-2K_j)\bar{z}_i + (1 - exp(-2K_j))z_i \tag{49}$$

Given that ϕ_i is the measure of investors learning about asset $i \in \mathcal{A}$, we have

$$\int_{\mathcal{M}_{Sj}} s_{ji}dj = \lambda\phi_i\left[exp(-2K_s)\bar{z}_i + (1 - exp(-2K_s))z_i\right]$$
$$\int_{\mathcal{M}_{Uj}} s_{ji}dj = (1-\lambda)\phi_i\left[exp(-2K_u)\bar{z}_i + (1 - exp(-2K_u))z_i\right]$$

Plugging these back into the market clearing condition and solving for the price coefficients a_i, b_i and c_i for the price of each asset, where $p_i = a_i + b_i z_i - c_i x_i$ we obtain

$$a_i = \frac{\bar{z}_i}{r\left\{1 + \phi_i\left[\lambda\left(e^{2K_s} - 1\right) + (1-\lambda)\left(e^{2K_u} - 1\right)\right]\right\}}$$
$$b_i = \frac{\phi_i\left[\lambda\left(e^{2K_s} - 1\right) + (1-\lambda)\left(e^{2K_u} - 1\right)\right]}{r\left\{1 + \phi_i\left[\lambda\left(e^{2K_s} - 1\right) + (1-\lambda)\left(e^{2K_u} - 1\right)\right]\right\}}$$

$$c_i = \frac{\rho\sigma_i^2}{r\left\{1+\phi_i\left[\lambda\left(e^{2K_s}-1\right)+(1-\lambda)\left(e^{2K_u}-1\right)\right]\right\}}$$

Letting $\mathcal{C} = \left[\lambda\left(e^{2K_s}-1\right)+(1-\lambda)\left(e^{2K_u}-1\right)\right]$ denote the total capacity in the economy, we can compactly rewrite the price coefficients as

$$a_i = \frac{\bar{z}_i}{r\left(1+\phi_i\mathcal{C}\right)}; \quad b_i = \frac{\phi_i\mathcal{C}}{r\left(1+\phi_i\mathcal{C}\right)}; \quad c_i = \frac{\rho\sigma_i^2}{r\left(1+\phi_i\mathcal{C}\right)} \tag{50}$$

and thus the price of each asset i takes the form

$$p_i = \frac{\bar{z}_i + z_i\phi_i\mathcal{C} - x_i\rho\sigma_i^2}{r\left(1+\phi_i\mathcal{C}\right)} \tag{51}$$

where $\mathcal{C} = \left[\lambda\left(e^{2K_s}-1\right)+(1-\lambda)\left(e^{2K_u}-1\right)\right]$ denotes the total capacity in the economy, and ϕ_i is the measure of investors learning about asset $i \in \mathcal{A}$, that we still have to solve for in equilibrium.

QED.

A.1.6 Actively traded assets

Proposition 4 Among the assets that are actively traded, $i \in \mathcal{A}$, not all are given the same attention by investors. The market endogenously learns about a select number of assets, and the mass of investors ϕ_i choosing to learn about asset i will vary with the volatility and liquidity of the asset.

Proof: Solving for $\hat{R}_i$, $\hat{\mathcal{V}}_i$ and the variance of posterior beliefs $\hat{V}_{ji}$ in equations (41) and (44) using the price equation (51), we obtain

$$\hat{R}_i = E[z_i - rp_i] = \bar{z}_i - \frac{\bar{z}_i + \bar{z}_i\phi_i\mathcal{C} - \bar{x}_i\rho\sigma_i^2}{(1+\phi_i\mathcal{C})} = \frac{\bar{x}_i\rho\sigma_i^2}{(1+\phi_i\mathcal{C})} \tag{52}$$

$$\hat{\mathcal{V}}_i = (1-2rb_i)\sigma_i^2 + r^2\sigma_{pi}^2 = \left(\frac{1-\phi_i\mathcal{C}}{1+\phi_i\mathcal{C}}\right)\sigma_i^2 + r^2\sigma_{pi}^2 = \frac{\sigma_i^2(1+\rho^2\sigma_i^2\sigma_{xi}^2)}{(1+\phi_i\mathcal{C})^2} \tag{53}$$

$$\hat{V}_{ji} = \hat{\mathcal{V}}_i - (1-2rb_i)\sigma_{\epsilon ji}^2 = \frac{\sigma_i^2(1+\rho^2\sigma_i^2\sigma_{xi}^2) - \sigma_{\epsilon ji}^2(1-\phi_i^2\mathcal{C}^2)}{(1+\phi_i\mathcal{C})^2} \tag{54}$$

where we plugged in σ_{pi}^2 as

$$\sigma_{pi}^2 = var(p_i) = var\left(\frac{z_i\phi_i\mathcal{C} - x_i\rho\sigma_i^2}{r\left(1+\phi_i\mathcal{C}\right)}\right) = \frac{\sigma_i^2\left(\phi_i^2\mathcal{C}^2 + \rho^2\sigma_i^2\sigma_{xi}^2\right)}{r^2\left(1+\phi_i\mathcal{C}\right)^2} \tag{55}$$

Substituting in equations (52) and (53) for $\hat{R}_i$ and $\hat{\mathcal{V}}_i$, we obtain the gain factor A_i is

$$A_i = \frac{\hat{\mathcal{V}}_i + \hat{R}_{ji}^2}{\sigma_{ji}^2} = \frac{1 + \rho^2 \sigma_i^2 (\sigma_{xi}^2 + \bar{x}_i^2)}{(1 + \phi_i \mathcal{C})^2} \tag{56}$$

Since $\phi_i > 0$ and $\mathcal{C} > 0$, then

$$\begin{aligned}
\frac{\partial A_i}{\partial \mathcal{C}} &= -\frac{2\phi_i \left(1 + \rho^2 \sigma_i^2 (\sigma_{xi}^2 + \bar{x}_i^2)\right)}{(1 + \phi_i \mathcal{C})^3} < 0 \\
\frac{\partial A_i}{\partial \phi_i} &= -\frac{2\mathcal{C} \left(1 + \rho^2 \sigma_i^2 (\sigma_{xi}^2 + \bar{x}_i^2)\right)}{(1 + \phi_i \mathcal{C})^3} < 0 \\
\frac{\partial A_i}{\partial \sigma_i^2} &= \frac{\rho^2 (\sigma_{xi}^2 + \bar{x}_i^2)}{(1 + \phi_i \mathcal{C})^2} > 0 \\
\frac{\partial A_i}{\partial \sigma_{xi}^2} &= \frac{\rho^2 \sigma_i^2}{(1 + \phi_i \mathcal{C})^2} > 0
\end{aligned}$$

Letting $\Sigma_i = \sigma_i^2(\sigma_{xi}^2 + \bar{x}_i^2)$, we can write the gain factor

$$A_i = \left(1 + \rho^2 \Sigma_i\right) (1 + \phi_i \mathcal{C})^{-2} \tag{57}$$

Thus, the gain factor A_i, is decreasing in the total capacity of the economy, $\mathcal{C}$, as well as in the mass ϕ_i of investors who learn about asset i. However, the gain factor increases with the volatility of the asset σ_i^2, as well as with the volatility of the stochastic supply σ_{xi}^2.

QED.

A.1.7 Learning equilibrium

Following Kacperczyk et al. (2018), we sort the N assets in terms of their volatility, $\Sigma_i > \Sigma_{i+1}$, so now can compute the threshold for learning about asset $i+1$ given that asset i has been learned about. If asset i has been learned about, then $\phi_i = 1$ and $\phi_{i+1}, ..., \phi_N = 0$. This implies that the gain for asset i is larger than the gain for the next asset $i+1$ such that $A_i > A_{i+1}$ implying that

$$\frac{\left(1 + \rho^2 \Sigma_i\right)}{(1 + \phi_i \mathcal{C})^2} > \frac{\left(1 + \rho^2 \Sigma_{i+1}\right)}{(1 + \phi_{i+1} \mathcal{C})^2}$$

For asset $i+1$ to be learned about in equilibrium, it has to be that the above inequality holds. Therefore, for any asset learned about in equilibrium, the gain factors must be equated and thus

$$\left(\frac{1+\phi_{i+1}\mathcal{C}}{1+\phi_i\mathcal{C}}\right)^2 = \frac{\left(1+\rho^2\Sigma_{i+1}\right)}{\left(1+\rho^2\Sigma_i\right)} \tag{58}$$

Solving for ϕ_{i+1} we get that the measure of investors that learn about asset $i+1$ is

$$\phi_{i+1} = \sqrt{\frac{\left(1+\rho^2\Sigma_{i+1}\right)\left(1+\phi_i\mathcal{C}\right)^2}{\left(1+\rho^2\Sigma_i\right)\mathcal{C}^2}} - \frac{1}{\mathcal{C}} \tag{59}$$

Any asset $i+1$ that is not learned about in equilibrium has a strictly lower gain factor such that

$$\frac{\left(1+\rho^2\Sigma_i\right)}{\left(1+\phi_i\mathcal{C}\right)^2} > \left(1+\rho^2\Sigma_{i+1}\right) \tag{60}$$

If only one asset is learned about in equilibrium, the one with the highest gain factor, the measure of investors learning about this asset is $\phi_1 = 1$, while $\phi_i = 0, \forall i \in \{2, 3, ..., N\}$ and the capacity of this economy with only one active asset is given by:

$$\mathcal{C}_1 = \sqrt{\frac{\left(1+\rho^2\Sigma_1\right)}{\left(1+\rho^2\Sigma_2\right)}} - 1 \tag{61}$$

If two assets are learned about in equilibrium then the capacity of this economy with two active assets is given by $\mathcal{C}_2$ below

$$\mathcal{C}_2 = \frac{\sqrt{\frac{(1+\rho^2\Sigma_1)}{(1+\rho^2\Sigma_2)}} - 1}{\phi_1 - \phi_2\sqrt{\frac{(1+\rho^2\Sigma_1)}{(1+\rho^2\Sigma_2)}}} = \frac{\mathcal{C}_1}{\phi_1 - \phi_2\sqrt{\frac{(1+\rho^2\Sigma_1)}{(1+\rho^2\Sigma_2)}}} \tag{62}$$

and so on and so forth. Notice that $0 < \mathcal{C}_0 < \mathcal{C}_1 < ... < \mathcal{C}_N$.

This implies that for any aggregate information capacity $\mathcal{C} \geq \mathcal{C}_1$, at least two assets will be learned about in equilibrium. As aggregate capacity increases, investors learn about new assets in a decreasing order of the assets' gains.

QED.

A.1.8 Mass of active investors ϕ_i

We can now fully characterize the equilibrium, following Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk et al. (2018). Let the first $n \leq N$ risky assets be learned about in equilibrium such that

$$\sum_{i=1}^{n\leq N} \phi_i = 1 \tag{63}$$

then all the gain factors on each of the first $n \leq N$ risky assets will be equal in equilibrium. This implies

$$\frac{1+\rho^2\Sigma_1}{(1+\phi_1\mathcal{C})^2} = \frac{1+\rho^2\Sigma_2}{(1+\phi_2\mathcal{C})^2} = ... = \frac{1+\rho^2\Sigma_i}{(1+\phi_i\mathcal{C})^2} = ... = \frac{1+\rho^2\Sigma_n}{(1+\phi_n\mathcal{C})^2}$$

This implies that, labeling $\mathcal{C}_n$ as the total capacity of the economy such as $n \leq N$ assets are learned about in equilibrium

$$\frac{1+\phi_i\mathcal{C}_n}{1+\phi_1\mathcal{C}_n} = \underbrace{\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}}_{\text{Call this } \mathbb{c}_{1i}}, \quad \forall\ i \leq n \leq N$$

So $\mathbb{c}_{1i} = \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}$.
Solving for ϕ_i gives

$$\begin{aligned}
\phi_i &= \frac{\mathbb{c}_{1i}(1+\phi_1\mathcal{C}_n)-1}{\mathcal{C}_n} = \\
&= \frac{(\mathbb{c}_{1i}+\phi_1\mathbb{c}_{1i}\mathcal{C}_n)-1}{\mathcal{C}_n} = \\
&= \phi_1\mathbb{c}_{1i} + \frac{(\mathbb{c}_{1i}-1)}{\mathcal{C}_n} = \\
&= \phi_1\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} + \frac{\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}-1}{\mathcal{C}_n} = \\
&= \frac{\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}(1+\phi_1\mathcal{C}_n)-1}{\mathcal{C}_n}
\end{aligned} \tag{64}$$

and plugging into the constraint equation (63) implies that

$$\begin{aligned}
\sum_{i=1}^{n\leq N}\phi_i = 1 &= \sum_{i=1}^{n\leq N}\frac{\mathbb{c}_{1i}(1+\phi_1\mathcal{C}_n)-1}{\mathcal{C}_n} \implies \\
1 &= \sum_{i=1}^{n\leq N}\frac{\mathbb{c}_{1i}}{\mathcal{C}_n} + \phi_1\sum_{i=1}^{n\leq N}\mathbb{c}_{1i} - \sum_{i=1}^{n\leq N}\frac{1}{\mathcal{C}_n} \implies \\
\phi_1 &= \frac{1-\frac{1}{\mathcal{C}_n}\sum_{i=1}^{n\leq N}(\mathbb{c}_{1i}-1)}{\sum_{i=1}^{n\leq N}\mathbb{c}_{1i}} \implies
\end{aligned}$$

$$\phi_1 = \frac{1 - \frac{1}{\mathcal{C}_n}\sum_{i=1}^{n\leq N}(\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1)}{\sum_{i=1}^{n\leq N}\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}} \tag{65}$$

Substituting ϕ_1 back into (64) we obtain the measure of active investors for each asset i as a function of primitives, assuming that $i = 1, 2, ...n$ are the $n \leq N$ assets learned about in equilibrium and denoting $\mathcal{C}_n$ as the total market capacity where n assets are learned about in equilibrium

$$\begin{aligned}
\phi_i &= \frac{\mathbb{c}_{1i}(1+\phi_1\mathcal{C}_n) - 1}{\mathcal{C}_n} = \frac{\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}(1+\phi_1\mathcal{C}_n) - 1}{\mathcal{C}_n} \\
&= \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}\left(\frac{1}{\mathcal{C}_n} + \phi_1\right) - \frac{1}{\mathcal{C}_n} \\
&= \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}\left(\frac{1}{\mathcal{C}_n} + \frac{1 - \frac{1}{\mathcal{C}_n}\sum_{i=1}^{n\leq N}\left(\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1\right)}{\sum_{i=1}^{n\leq N}\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}}\right) - \frac{1}{\mathcal{C}_n}
\end{aligned} \tag{66}$$

Lastly, the measures of sophisticated and unsophisticated investors that learn about asset i as

$$M_{Sj} = \lambda\phi_i = \lambda\left[\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}\left(\frac{1}{\mathcal{C}_n} + \frac{1 - \frac{1}{\mathcal{C}_n}\sum_{i=1}^{n\leq N}\left(\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1\right)}{\sum_{i=1}^{n\leq N}\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}}\right) - \frac{1}{\mathcal{C}_n}\right] \tag{67}$$

$$M_{Uj} = (1-\lambda)\phi_i = (1-\lambda)\left[\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}\left(\frac{1}{\mathcal{C}_n} + \frac{1 - \frac{1}{\mathcal{C}_n}\sum_{i=1}^{n\leq N}\left(\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1\right)}{\sum_{i=1}^{n\leq N}\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}}\right) - \frac{1}{\mathcal{C}_n}\right] \tag{68}$$

We are now done characterizing the equilibrium of this economy.

QED.

A.2 Proofs of Theoretical Predictions

A.2.1 Prediction 1

Heterogeneity in information sophistication leads to differences in realized and expected capital income, as well as in expected portfolio holdings. That is, if $K_S > K_U$, then $\sum_{i=1}^{N}\Pi_{S,i,t} > \sum_{i=1}^{N}\Pi_{U,i,t}$, and $\sum_{i=1}^{N}E[\Pi_{S,i,t}] > \sum_{i=1}^{N}E[\Pi_{U,i,t}]$ and $E[H_{S,i,t}/\lambda] \geq E[H_{S,i,t}/(1-\lambda)]$.

Proof: Let $\Pi_{j,i,t}$ be the average profit per capita for investor j from trading asset i at time t:

$$\Pi_{S,i,t} = \frac{H_{S,i,t}(z_{i,t} - rp_{i,t})}{\lambda} \text{ for sophisticated investors and} \tag{69}$$

$$\Pi_{U,i,t} = \frac{H_{U,i,t}(z_{i,t} - rp_{i,t})}{(1-\lambda)} \text{ for unsophisticated investors} \tag{70}$$

where $H_{j,i,t}$ is the holding level of asset i for investors of type j at time t, obtained by integrating optimal portfolio choices q_{ji} across investors of type j, and ϕ_i denotes the mass of investors learning about asset i

$$H_{S,i,t} = \lambda \left[\frac{(\bar{z}_i - rp_{i,t}) + \phi_i(e^{2K_S} - 1)(z_i - rp_{i,t})}{\rho\sigma_i^2}\right] \tag{71}$$

$$H_{U,i,t} = (1-\lambda) \left[\frac{(\bar{z}_i - rp_{i,t}) + \phi_i(e^{2K_U} - 1)(z_i - rp_{i,t})}{\rho\sigma_i^2}\right] \tag{72}$$

Substituting $H_{j,i,t}$ in the profit functions we get

$$\Pi_{S,i,t} = \left[\frac{(\bar{z}_i - rp_{i,t})(z_{i,t} - rp_{i,t}) + \phi_i(e^{2K_S} - 1)(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{73}$$

$$\Pi_{U,i,t} = \left[\frac{(\bar{z}_i - rp_{i,t})(z_{i,t} - rp_{i,t}) + \phi_i(e^{2K_U} - 1)(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{74}$$

Capital income heterogeneity is thus given by the equation below

$$\sum_{i=1}^{N} \Pi_{S,i,t} - \Pi_{U,i,t} = \sum_{i=1}^{N} \left[\frac{\phi_i(e^{2K_1} - e^{2K_2})(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{75}$$

Thus, if $K_S > K_U$, $e^{2K_S} > e^{2K_U}$ and $\frac{H_{S,i,t}}{\lambda} > \frac{H_{U,i,t}}{(1-\lambda)}$ which implies $\sum_{i=1}^{N} \Pi_{S,i,t} > \sum_{i=1}^{N} \Pi_{U,i,t}$.

We can also compute expected per capita profits for sophisticated and unsophisticated investors.

$$\begin{aligned}
E\left[\Pi_{S,i,t}\right] &= E\left[\frac{(\bar{z}_i - rp_{i,t})(z_{i,t} - rp_{i,t}) + \phi_i(e^{2K_S} - 1)(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] \\
&= \frac{\rho^2\Sigma_i - \phi_i\mathcal{C}\sigma_i^2 + \phi_i(e^{2K_S} - 1)(\rho^2\Sigma_i + \sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2} \\
E\left[\Pi_{U,i,t}\right] &= E\left[\frac{(\bar{z}_i - rp_{i,t})(z_{i,t} - rp_{i,t}) + \phi_i(e^{2K_U} - 1)(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right]
\end{aligned} \tag{76}$$

$$= \frac{\rho^2\Sigma_i - \phi_i\mathcal{C}\sigma_i^2 + \phi_i(e^{2K_U} - 1)(\rho^2\Sigma_i + \sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2} \tag{77}$$

The fact that sophisticated investors achieve higher profits is because they are better able to identify profitable assets (so they invest in high risk, high return assets more). Assume that the first $1 < n \leq N$ risky assets are learned about in equilibrium (remember that we sorted the assets in terms of their idiosyncratic volatility such that $\Sigma_1 > \Sigma_2 > ... > \Sigma_n > ... > \Sigma_N$) then,

$$\phi_i = 0, \ \forall i \in \{n+1, ..., N\} \implies \frac{H_{S,i,t}}{\lambda} = \frac{H_{U,i,t}}{(1-\lambda)} \implies \sum_{i=n+1}^{N} \left(\Pi_{S,i,t} - \Pi_{U,i,t}\right) = 0 \tag{78}$$

$$\phi_i > 0, \ \forall i \in \{1, ..., n\} \implies E\left[\frac{H_{S,i,t}}{\lambda}\right] > E\left[\frac{H_{U,i,t}}{(1-\lambda)}\right] \implies \sum_{i=n+1}^{N} \left[E[\Pi_{S,i,t} - \Pi_{U,i,t}\right] > 0 \tag{79}$$

So, while both types of investors make the same returns on the passive assets and have the same holdings per capita of passive assets, sophisticated investors make higher expected returns and have higher expected holdings per capita of active assets compared to unsophisticated investors. As we saw before, active assets are high-risk, high-return assets (with maximal gains). And investors on average tilt their portfolios more towards profitable assets compared with unsophisticated investors

$$\text{if } i, i' \in \{1, ..., n\} \text{ with } E[z_i - rp_i] > E[z_{i'} - rp_{i'}] \implies E\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{S,i',t}}{\lambda}\right] > E\left[\frac{H_{U,i,t}}{1-\lambda} - \frac{H_{U,i',t}}{1-\lambda}\right] \tag{80}$$

Lastly, we mentioned that sophisticated investors also achieve larger gains from positive shocks and smaller losses from negative shocks

$$\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)} = \frac{\phi_i(e^{2K_S} - e^{2K_U})(z_i - rp_i)}{\rho\sigma_i^2} \tag{81}$$

$$\text{and } \frac{\partial \frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}}{\partial(z_i - rp_i)} = \frac{\phi_i(e^{2K_S} - e^{2K_U})}{\rho\sigma_i^2} > 0 \tag{82}$$

QED.

A.2.2 Prediction 2

An absolute (symmetric) increase in information sophistication leads to growing expected capital income heterogeneity. That is, if both K_S and K_U increase proportionally by the same percent difference, $K_S' = (1+\xi)K_S$ and $K_U' = (1+\xi)\gamma K_S$, where $0 < \xi \leq 1$ and $0 < \gamma < 1$ expected income heterogeneity increases, $d\left(E[\Pi'_{S,i,t}] - E[\Pi'_{U,i,t}]\right)/dK_S > 0$.

Proof. Assume K_S and $K_U = \gamma K_S$, with $\gamma < 1$. For example, let $K_S = 100$ and $\gamma = 0.8 < 1$, which implies $K_U = 80$. Now, assume a symmetric increase in capacity, such that the new $K_S' = (1+0.5)K_S = 150$, a 50% increase, and the new $K_U' = (1+0.5)K_U = 120$, also a 50% increase. We are thus interested in comparative statics with respect to K_S. Then new total capacity in the economy becomes

$$\mathcal{C}' = \left[\lambda\left(e^{2K_S} - 1\right) + (1-\lambda)\left(e^{2\gamma K_S} - 1\right)\right]$$

The difference in profits is

$$\sum_{i=1}^{n} \Pi_{S,i,t} - \Pi_{U,i,t} = \sum_{i=1}^{n}\left[\frac{\phi_i(e^{2K_S} - e^{2\gamma K_S})(z_i - rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{83}$$

So let's take the derivative of ϕ_i with respect to aggregate capacity $\mathcal{C}$. From equation (66)

$$\phi_i = \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}\left(\frac{1}{\mathcal{C}_n} + \frac{1 - \frac{1}{\mathcal{C}_n}\sum_{i=1}^{n\leq N}\left(\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} - 1\right)}{\sum_{i=1}^{n\leq N}\sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}}}\right) - 1 \tag{84}$$

$$= \mathbb{c}_{1i}\left(\frac{1}{\mathcal{C}} + \frac{1 - \frac{1}{\mathcal{C}}\sum_{i=1}^{n\leq N}(\mathbb{c}_{1i} - 1)}{\sum_{i=1}^{n\leq N}\mathbb{c}_{1i}}\right) - 1 \tag{85}$$

where $\mathbb{c}_{1i} = \sqrt{\frac{1+\rho^2\Sigma_i}{1+\rho^2\Sigma_1}} < 1$ so

$$\frac{d\phi_i}{d\mathcal{C}} = -\mathbb{c}_{1i}\left(\frac{1}{\mathcal{C}^2} + \frac{1}{\mathcal{C}^2}\frac{\sum_{i=1}^{n\leq N}(1-\mathbb{c}_{1i})}{\sum_{i=1}^{n\leq N}\mathbb{c}_{1i}}\right) < 0 \tag{86}$$

and thus

$$\frac{d\phi_1}{d\mathcal{C}} = -\frac{1}{\mathcal{C}^2}\frac{\sum_{i=2}^{n\leq N}(1-\mathbb{c}_{1i})}{\sum_{i=1}^{n\leq N}\mathbb{c}_{1i}} < 0 \tag{87}$$

Equations (87) and (63) imply that for at least one asset $i \leq n \leq N$, it has to be that $\frac{d\phi_i}{d\mathcal{C}} > 0$. Moreover, the amount of total information devoted to asset i is $\mathbb{c}_{1i}\left(1 + \frac{\mathcal{C} - \sum_{i=1}^{n\leq N}(\mathbb{c}_{1i} - 1)}{\sum_{i=1}^{n\leq N}\mathbb{c}_{1i}}\right) - 1$

and its derivative with respect to market capacity is

$$\frac{d\mathcal{C}\phi_i}{d\mathcal{C}} = \mathbb{c}_{1i}\left(\frac{1}{\sum_{i=1}^{n\leq N}\mathbb{c}_{1i}}\right) > 0 \tag{88}$$

We are first going to show that $\frac{d\phi_i(exp(2K_S)-1)}{d\mathcal{C}} > 0$ and that $\frac{d\phi_i(exp(2K_S)-1)}{d\mathcal{C}} > \frac{d\phi_i(exp(2K_U)-1)}{d\mathcal{C}}$. With this, we can then show that aggregate (symmetric) growth in financial technology modeled as a common growth rate of K_S and K_U leads to growing capital income heterogeneity.

The derivatives of the amount of total information sophisticated and unsophisticated investors respectively devote to asset i with respect to market capacity are

$$\frac{d\phi_i(e^{2K_S}-1)}{d\mathcal{C}} = \frac{d\phi_i(e^{2K_S}-1)}{d\mathcal{C}} = 2e^{2K_S}\phi_i + \frac{d\phi_i}{d\mathcal{C}}(e^{2K_S}-1)\frac{d\mathcal{C}}{dK_S} \tag{89}$$

$$\frac{d\phi_i(e^{2K_U}-1)}{d\mathcal{C}} = \frac{d\phi_i(e^{2\gamma K_S}-1)}{d\mathcal{C}} = 2\gamma e^{2K_S}\phi_i + \frac{d\phi_i}{d\mathcal{C}}(e^{2\gamma K_S}-1)\frac{d\mathcal{C}}{dK_S} \tag{90}$$

$$\text{where } \frac{d\mathcal{C}}{dK_S} = 2\lambda(e^{2K_S}) + 2(1-\lambda)\gamma(e^{2\gamma K_S}) > 0 \tag{91}$$

Because $\phi_i > 0$, and $\frac{d\phi_i}{d\mathcal{C}} > 0$, and $\frac{d\mathcal{C}}{dK_S} > 0$ and $e^{2K_S} > \gamma e^{2\gamma K_S}$, it follows that

$$\frac{d\phi_i(e^{2K_S}-1)}{d\mathcal{C}} > \frac{d\phi_i(e^{2K_U}-1)}{d\mathcal{C}} \tag{92}$$

We can now show that expected profits diverge for sophisticated and unsophisticated investors. Using equations (76) and (77) for the expected profits of sophisticated and unsophisticated investors, we get that

$$\begin{aligned}
&E\left[\Pi_{S,i,t}\right] - E\left[\Pi_{U,i,t}\right] = \\
&= \frac{\rho^2\Sigma_i - \phi_i\mathcal{C}\sigma_i^2 + \phi_i(e^{2K_S}-1)(\rho^2\Sigma_i+\sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2} - \frac{\rho^2\Sigma_i - \phi_i\mathcal{C}\sigma_i^2 + \phi_i(e^{2K_U}-1)(\rho^2\Sigma_i+\sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2} = \\
&= \frac{\phi_i(e^{2K_S}-e^{2K_U})(\rho^2\Sigma_i+\sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2} \\
&= \phi_i(e^{2K_S}-e^{2\gamma K_S})(\rho^2\Sigma_i+\sigma_i^2)\rho^{-1}(1+\phi_i\mathcal{C})^{-2}
\end{aligned}$$

Taking the derivative of the expected profit gap, we obtain that

$$\frac{d(E\left[\Pi_{S,i,t}\right]-E\left[\Pi_{U,i,t}\right])}{dK_S}=\frac{(\rho^2\Sigma_i+\sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2}\left(\underbrace{\frac{d\phi_i(e^{2K_S}-1)}{d\mathcal{C}}-\frac{d\phi_i(e^{2K_U}-1)}{d\mathcal{C}}}_{\text{positive by eqn. 92}}+\underbrace{2\phi_i(e^{2K_S}-e^{2\gamma K_S})}_{\text{positive}}\right)>0 \tag{93}$$

Notice that this also holds for

$$\frac{d(E\left[\Pi'_{S,i,t}\right]-E\left[\Pi'_{U,i,t}\right])}{dK_S}= \tag{94}$$

$$=\frac{(\rho^2\Sigma_i+\sigma_i^2)}{\rho(1+\phi_i\mathcal{C})^2}\left(\underbrace{\frac{d\phi_i(e^{2(1+\xi)K_S}-1)}{d\mathcal{C}}-\frac{d\phi_i(e^{2(1+\xi)K_U}-1)}{d\mathcal{C}}}_{\text{positive by eqn. 92}}+\underbrace{2\phi_i(e^{2(1+\xi)K_S}-e^{2\gamma(1+\xi)K_S})}_{\text{positive}}\right)>0 \tag{95}$$

Q.E.D.

A.2.3 Prediction 3

A relative (asymmetric) increase in information sophistication leads to greater realized capital income heterogeneity. That is, if K_S and K_U increase by δ_S and δ_U respectively, such that $K'_S = K_S + \delta_S$ and $K'_U = K_U + \delta_U$, where $\delta_S > \delta_U$, realized capital income heterogeneity increases, $\partial\ (\Pi_{S,i,t} - \Pi_{U,i,t})/\partial(e^{2K_S} - e^{2K_U}) > 0$.

Proof: In other words, consider an increase in the capacity gap $K'_S - K'_U > K_S - K_U$. Plugging the holdings into the average profits yields

$$\Pi_{S,i,t}=\left[\frac{(\bar{z}_i-rp_{i,t})(z_{i,t}-rp_{i,t})+\phi_i(e^{2K_S}-1)(z_i-rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{96}$$

$$\Pi_{U,i,t}=\left[\frac{(\bar{z}_i-rp_{i,t})(z_{i,t}-rp_{i,t})+\phi_i(e^{2K_U}-1)(z_i-rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{97}$$

Capital income heterogeneity is thus given by the equation below

$$\Pi_{S,i,t}-\Pi_{U,i,t}=\left[\frac{\phi_i(e^{2K_S}-e^{2K_U})(z_i-rp_{i,t})^2}{\rho\sigma_i^2}\right] \tag{98}$$

As long as $K_S > K_U$ and there is learning in equilibrium about asset i, such that $\phi_i \neq 0$, equation (99) is increasing in the capacity gap $K_S - K_U$.

$$\frac{\partial \Pi_{S,i,t} - \Pi_{U,i,t}}{\partial (e^{2K_S} - e^{2K_U})} = \left[\frac{\phi_i (z_i - rp_{i,t})^2}{\rho \sigma_i^2} \right] > 0 \tag{99}$$

QED.

A.2.4 Prediction 4

An increase in absolute (and relative) sophistication predicts a growing presence of sophisticated investors in risky asset classes. In other words, $\frac{dE\left[\frac{H_{S,i}}{\lambda} - \frac{H_{U,i}}{(1-\lambda)}\right]}{dK_S} > 0$ and $\frac{dE\left[\frac{H_{S,i}}{\lambda} - \frac{H_{U,i}}{(1-\lambda)}\right]}{d\mathcal{C}} > 0$, where $H_{S,i}$ is the holding level of a sophisticated investor of asset i, and $H_{U,i}$ is the holding level of an unsophisticated investor of asset i.

Proof: The proof for a relative increase in sophistication is trivial. We need to show that the result also holds for an absolute increase in sophistication levels.

$$\frac{dE\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}\right]}{dK_S} > 0 \tag{100}$$

Note that the excess return is given by

$$z_i - rp_i = \frac{z_i\left(1 + \phi_i \mathcal{C}\right) - \bar{z}_i - z_i \phi_i \mathcal{C} + x_i \rho \sigma_i^2}{(1 + \phi_i \mathcal{C})} = \frac{z_i - \bar{z}_i + x_i \rho \sigma_i^2}{(1 + \phi_i \mathcal{C})} \tag{101}$$

Taking expectations we obtain expected excess returns as

$$E[z_i - rp_i] = \frac{\bar{x}_i \rho \sigma_i^2}{(1 + \phi_i \mathcal{C})} \tag{102}$$

and differencing with respect to K_S we get

$$\frac{\partial \frac{\bar{x}_i \rho \sigma_i^2}{(1+\phi_i \mathcal{C})}}{\partial K_S} = (-1) \bar{x}_i \rho \sigma_i^2 \left(1 + \phi_i \mathcal{C}\right)^{-2} \underbrace{\frac{\partial \phi_i \mathcal{C}}{\partial K_S}}_{<0} > 0 \tag{103}$$

In equation (81) we calculated the difference in the holdings of an active asset $i \in \mathcal{A}$. Substituting in the price formula (15), we obtain

$$\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1 - \lambda)} = \frac{\phi_i (e^{2K_S} - e^{2K_U})(z_i - rp_i)}{\rho \sigma_i^2} \tag{104}$$

$$= \frac{\phi_i(e^{2K_S} - e^{2K_U})(z_i - \bar{z}_i + x_i \rho \sigma_i^2)}{\rho \sigma_i^2 \left(1 + \phi_i \mathcal{C}\right)} \tag{105}$$

Taking expectations

$$E\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}\right] = \frac{\phi_i(e^{2K_S} - e^{2K_U})(\bar{x}_i \rho \sigma_i^2)}{\rho \sigma_i^2 \left(1 + \phi_i \mathcal{C}\right)} \tag{106}$$

$$= \frac{\phi_i(e^{2K_S} - e^{2K_U})\bar{x}_i}{(1 + \phi_i \mathcal{C})} \tag{107}$$

and differencing with respect to K_S

$$\frac{dE\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}\right]}{dK_S} = \frac{d\frac{\phi_i(e^{2K_S}-e^{2K_U})\bar{x}_i}{(1+\phi_i\mathcal{C})}}{dK_S} \tag{108}$$

$$= \frac{[d\phi_i(e^{2K_S} - e^{2K_U})\bar{x}_i/dK_S \times (1 + \phi_i \mathcal{C}) - \phi_i(e^{2K_S} - e^{2K_U})\bar{x}_i \times d\phi_i \mathcal{C}/dK_S}{(1 + \phi_i \mathcal{C})^2} = \tag{109}$$

$$= \frac{[d\phi_i(e^{2K_S} - e^{2K_U})\bar{x}_i/dK_S \times (1 + \phi_i \mathcal{C})}{(1 + \phi_i \mathcal{C})^2} - \frac{\phi_i(e^{2K_S} - e^{2K_U})\bar{x}_i \times d\phi_i \mathcal{C}/dK_S}{(1 + \phi_i \mathcal{C})^2} \tag{110}$$

Note that we can write

$$\frac{dE\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}\right]}{dK_S} = \frac{\partial E\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}\right]}{\partial E[z_i - rp_i]} \times \frac{\partial E[z_i - rp_i]}{\partial K_S} = \tag{111}$$

$$= \underbrace{\frac{\partial E\left[\frac{H_{S,i,t}}{\lambda} - \frac{H_{U,i,t}}{(1-\lambda)}\right]}{\partial E[z_i - rp_i]}}_{>0 \text{ from eqn. (82)}} \times \underbrace{\frac{\partial \frac{\bar{x}_i \rho \sigma_i^2}{(1+\phi_i \mathcal{C})}}{\partial K_S}}_{>0 \text{ from eqn. (103)}} > 0 \tag{112}$$

Q.E.D.

Equation (111) implies that when aggregate sophistication goes up, that is when K_S increases (remember that $K_U = \gamma K_S$ and all comparative statics are done without loss of generality with respect to K_S), the difference in the holdings of an active asset $i \in A$ increases, implying that sophisticated investors will a larger share, while unsophisticated investors will hold a smaller share of active assets. The order in which learning happens suggests that these active assets are the riskiest assets (see Appendix *A.1.6*, which shows that the active assets are those with the highest gain factor, which increases with the volatility of the asset σ_i^2).

A.2.5 Prediction 5

An absolute (and relative) increase in sophistication predicts a retrenchment of less sophisticated investors from trading and stock market ownership in general.

Proof: This follows nicely from equation (111). Equation (111) implies that when aggregate sophistication goes up, that is when K_S increases (remember that $K_U = \gamma K_S$ and all comparative statics are done without loss of generality with respect to K_S), the difference in the holdings of an active asset $i \in A$ increases, implying that sophisticated investors will a larger share, while unsophisticated investors will hold a smaller share of active assets. The order in which learning happens suggests that these active assets are the riskiest assets (see the end of Appendix *A*.1.6, which shows that the active assets are those with the highest gain factor, which increases with the volatility of the asset σ_i^2).

The mechanism works through the price such that prices increase with sophistication.

The expected price and excess return, given by

$$E[p_i] = \frac{\bar{z}_i(1 + \phi_i \mathcal{C}) - \bar{x}_i \rho \sigma_i^2}{(1 + \phi_i C)^2} \tag{113}$$

$$E[z_i - rp_i] = \frac{\bar{x}_i \rho \sigma_i^2}{(1 + \phi_i \mathcal{C})} \tag{114}$$

are increasing and decreasing respectively in $\phi_i \mathcal{C}$ given that we have already proved that $\phi_i \mathcal{C}$ is increasing in $\mathcal{C}$.

QED.

Appendix B Tables

Table B1: Summary statistics[a]

	Sophisticated[b]		Unsophisticated[c]		All Sample	
	Mean	Std	Mean	Std	Mean	Std
Financial literacy	0.91	0.10	0.18	0.15	0.57	0.30
FinTech[d]	0.50	0.49	0.34	0.48	0.44	0.50
RegionFinTech[e]	0.31	0.23	0.37	0.27	0.32	0.23
Remote [f]	0.40	0.48	0.10	0.31	0.22	0.42
Share of risky assets[g]	0.10	0.23	0.01	0.08	0.05	0.17
Rate of financial returns	1.26	0.88	0.78	0.61	1.04	0.78
Equivalised income quintile						
first quintile	0.06	0.24	0.26	0.45	0.16	0.37
second quintile	0.11	0.30	0.28	0.46	0.20	0.40
third quintile	0.16	0.36	0.22	0.42	0.21	0.41
fourth quintile	0.25	0.42	0.15	0.36	0.21	0.41
fifth quintile	0.42	0.48	0.09	0.29	0.22	0.41
Risk aversion						
No risk aversion	0.01	0.11	0.01	0.08	0.01	0.09
Low risk aversion	0.12	0.32	0.11	0.32	0.11	0.32
Medium risk aversion	0.41	0.48	0.19	0.40	0.29	0.45
High risk aversion	0.46	0.49	0.69	0.47	0.59	0.49
Household type						
Single person over 65	0.05	0.21	0.29	0.46	0.14	0.35
Single person under 65	0.14	0.34	0.12	0.32	0.14	0.34
Couple without children	0.22	0.40	0.21	0.42	0.22	0.42
Couple with children	0.49	0.49	0.25	0.44	0.39	0.49
Single parent with children	0.07	0.25	0.09	0.29	0.08	0.27
Other household type	0.04	0.18	0.03	0.18	0.04	0.19
Age class						
30 and under	0.05	0.21	0.04	0.21	0.05	0.23
31-40	0.19	0.38	0.12	0.33	0.17	0.37
41-50	0.26	0.43	0.14	0.35	0.21	0.41
51-65	0.32	0.46	0.19	0.40	0.26	0.44
over 65	0.18	0.38	0.51	0.51	0.31	0.46
Municipality size (inhabitants)						
Up to 20.000	0.42	0.48	0.50	0.51	0.47	0.50
From 20.000 to 40.000	0.11	0.31	0.16	0.37	0.14	0.35
From 40.000 to 500.000	0.32	0.45	0.23	0.43	0.27	0.44
Over 500.000	0.15	0.35	0.11	0.32	0.13	0.33

Continued on next page

Table B1 – *Continued from previous page*

	Sophisticated[b]		Unsophisticated[c]		All Sample	
	Mean	Std	Mean	Std	Mean	Std
Gender						
Male	0.74	0.43	0.55	0.50	0.67	0.47
Female	0.26	0.43	0.45	0.50	0.33	0.47
Born abroad	0.04	0.18	0.11	0.32	0.07	0.26
Education						
No schooling	0.00	0.06	0.10	0.30	0.03	0.18
Primary school	0.06	0.23	0.37	0.49	0.18	0.39
Lower secondary school	0.26	0.43	0.33	0.48	0.37	0.48
Upper secondary school	0.41	0.48	0.15	0.37	0.28	0.45
University degree	0.27	0.43	0.05	0.22	0.13	0.34
Work status						
Employee	0.55	0.49	0.35	0.48	0.48	0.50
Self-employed	0.19	0.38	0.05	0.23	0.12	0.32
Not employed	0.26	0.43	0.60	0.50	0.40	0.49

Source: Bank of Italy Survey on Household Income and Wealth 2004-2020.

[a] Individual characteristics refer to the head of the household defined as the main income earner with the exception of risk aversion and financial literacy collected for respondents.

[b] Investors in the upper quartile of the financial literacy distribution.

[c] Investors in the lower quartile of the financial literacy distribution

[d] Dummy equal to 1 for investors with main bank providing a number of digital services to households above the regional mean.

[e] Regional mean of the index of digital services to households provided by investors' main bank.

[f] Dummy equal to 1 for investors that have access to remote banking.

[g] Risky financial assets: shares and equity, managed portfolio, funds (in equities, mixed or in foreign currencies), foreign securities and other complex financial assets (option, futures, etc.)

Table B2: Households' selection in high digitalised banks

	Dependent variable:
	(FinTech[a])
Sophisticated[b]	−0.003 (0.008)
Other covariates [c]	YES
Time fixed effects	YES
Region fixed effects	YES
Observations	52,129
Pseudo R-square	0.064

Note:*p<0.1; **p<0.05; ***p<0.01; Standard errors (clustered by region) in brackets.
Source: Bank of Italy Survey on Household Income and Wealth 2004-2020.
[a] Dummy equal to 1 for investors with main bank providing a number of digital services to households above the regional mean.
[b] Dummy equal to 1 for investors in the upper quartile of the financial literacy distribution.
[c] Other covariates: equivalent income quintile, household type, gender, work status, size of the municipality of residence, born abroad. Individual characteristics refer to the head of the household defined as the main income earner with the exception of risk aversion and financial literacy collected for respondents.

Previous volumes in this series

All volumes are available on our website www.bis.org.