The investment behavior of Dutch households

What is the role of preferences?

Master's Thesis
Financial Economics

Name:	Sten te Vogt
Student number:	427180sv
Supervisor:	Dr. M. Montone
Date:	29/08/2017

Abstract

The field of household finance studies how households use financial markets and instruments to achieve their ambitions. I investigate investment behavior of Dutch households with a focus on how rank-dependent and/or skewness preferences can explain under-diversification of equity portfolios. Using data from the DNB Household Survey (DHS)[1] I find that households in the Netherlands hold undiversified portfolios of individual stocks and at the same time invest in well-diversified mutual funds. However, both theories fail to (fully) explain the observed under-diversification. Future research should gain information as detailed as possible on the holdings of households and should study how theories of under-diversification interact.

Key words: households, Dutch, financial assets, equity, portfolios, DNB Household Survey, rank-dependent preferences, (under)diversification, skewness

[1] In this paper use is made of data of the DNB Household Survey conducted by CentERdata. I am grateful to CentERdata for providing the data for this thesis.

Table of Contents

List of Tables

List of Figures

1. Introduction

Individual investor behavior is a popular and often studied topic in the field of finance. This field studies both how individuals should invest their money and how individuals actually do invest. Expected utility theory was one of the first (normative) theories on the decision-making of investors (Von Neumann & Morgenstern, 1944). Expected utility assumes a function that is linear in the probabilities (linear transformation of probabilities). Experimental and empirical research show that expected utility theory is systematically violated. Therefore, researchers have developed non-expected utility models. These functions assume non-linear transformation in probabilities. Examples of non-expected utility are rank-dependent utility and (cumulative) prospect theory.

Another field in finance, that has gained interest in more recent years, is the field of household finance. This field studies how households use financial markets and instruments to achieve their ambitions. A large body of research in this domain shows that the majority of households make (costly) investment mistakes that are often not in accordance with standard economic theory. This has consequences for the way financial products should be designed and regulated (Von Gaudecker, 2015). In this paper I combine the fields of individual investor behavior and household finance. I investigate actual investment behavior of Dutch households with a focus on how under-diversification of investment portfolios can be explained. I study the ability of both rank-dependent preferences as proposed by Polkovnichenko (2005) as well as skewness preference as suggested by Mitton and Vorkink (2007) to explain households' investment behavior. Polkovnichenko (2005, p. 1467) reports two main patterns inconsistent with expected utility, but consistent with rank-dependent preferences: "(i) many households simultaneously invest in well-diversified funds and in poorly-diversified portfolios of stocks; and (ii) some households with substantial savings do not invest anything in equities". Mitton and Vorkink (2007) find that (i) portfolio returns of under-diversified investors are more (positively) skewed than those of their diversified counterparts; (ii) under-diversified investors give up mean-variance efficiency for higher exposure to skewness; and (iii) it is not a coincidence that under-diversification is related to skewness.

I make use of data from the DNB Household Survey (DHS). The DHS is an annual survey monitored by CentERdata that consists of a representative sample of the Dutch population. The DHS comprises six modules that cover economic and psychological determinants of saving behavior. I use data from five DHS waves to identify changes over time. The DHS is a very detailed

survey that allows to gain insight in the holdings of households. In the first part of this paper I follow the methodological approach of Polkovnichenko (2005) and look into the cross-section of equity ownership. It appears that the percentage of households that own stocks ranges from approximately 10% to approximately 17%. Most striking about the finding that many households do not participate in the stock market, is that a fraction of these same households have substantial financial assets in other accounts. Meanwhile, households that do invest move more and more toward owning equity indirectly. Direct ownership of equity is characterized by the fact that the majority of investors own very few stocks. To provide some insightful numbers, approximately 95% of the households own shares of five or less companies and the amount of total financial assets owned by this group is around 90%. Similar to the findings of Polkovnichenko (2005) for households in the U.S., it seems that households in the Netherlands hold undiversified portfolios of individual stocks and at the same time invest in well-diversified mutual funds. Unlike Polkovnichenko (2005), I cannot confirm that households are aware of the higher levels of risk that are associated with undiversified portfolios. Hence, I cannot conclude that rank-dependent preferences conclusively explain the investment behavior of Dutch households.

Therefore, I explore using the approach of Mitton and Vorkink (2007) whether a preference for skewness can explain the under-diversification of Dutch households. To this end, I use detailed information from the DHS to construct the portfolios of households and to match their portfolio holdings with the relevant return distributions obtained from Datastream. Methodologically, I construct three measures of diversification and a coefficient that captures portfolio skewness. Unlike Mitton and Vorkink (2007), I do not find that undiversified households have substantially higher levels of skewness. Moreover, both the portfolio diversification level as well as portfolio skewness are not conclusively negatively related to mean-variance efficiency (measured by the Sharpe ratio of a household's portfolio). Thus, a preference for skewness can also not explain the under-diversification of Dutch households.

Both theories fail to (fully) explain the observed investment behavior. Therefore, most likely other theories of under-diversification are playing a role. This does not imply that both theories of preference do not exist or that that they are irrelevant, it merely shows that the theories do not hold universally. Alternatively, the fact that I do not find evidence in favor of one of the theories could also be the consequence of limitations related to the methodology and/or data of this study. The use of survey data, even though the DHS is a highly sophisticated survey, implies

the use of imperfect data and associated constraints and consequences. Further research should attempt to validate and/or link this data with other data sources. This would enable researchers to work with more accurate data and larger sample sizes, resulting in qualitatively superior analyses and studies. Moreover, future research should study the coexistence of and/or interaction between alternative theories of under-diversification of household portfolios.

This paper proceeds as follows: Section 2 explores the previous literature on investment behavior and household finance and provides the first set of hypotheses. Section 3 discusses the data source(s) and methodology of this paper and presents the first descriptive and summary statistics that are key to the further continuation of this research. Thereafter, section 4 provides the results of the main analyses. Section 5 concludes by discussing the results, its implications, the limitations of this paper and thoughts on future research.

2. Literature review

2.1 Expected utility vs. non-expected utility

Individual investor behavior is one of the main fields of research in (behavioral) finance. Economists have developed models about how individuals should make investments decisions. However, these models often do not reflect actual behavior. According to one of the first theories in decision-making literature, expected utility (EU) theory, individuals make decisions by comparing the expected utility values of risky or uncertain prospects (Von Neumann & Morgenstern, 1944). Expected utility values are calculated by multiplying the utility values of prospects with their respective probabilities. Risk aversion under EU is modelled as the result of the concavity of the utility function over wealth (Rabin, 2000). This implies that investors are averse of large-scale risk, but also that individuals are more or less risk neutral for smaller stakes. Expected utility theory as modelled by Von Neumann and Morgenstern (1944) assumes uniform attitude towards risk.

However, attitude towards risk is often dependent, not uniform. Many scholars and theorists developed alternative models that more accurately describe and reflect actual behavior for decision-making under risk and/or uncertainty. Friedman and Savage (1948) observe a puzzling finding inconsistent with expected utility theory: individuals who insure themselves often buy lottery tickets as well. To accommodate this finding the researchers propose a utility function that

consists of both a concave and convex part. This pioneering work explored portfolio selection with non-expected utility preferences and gained traction in the years following. Markowitz (1952) criticized and modified the utility function of Friedman and Savage (1948). In order to account for some issues (see Markowitz, 1952), one of the inflection points of the utility function is placed at the status quo. In the context of portfolio choice this implies that investors are interested in having both risk protection and a "shot at riches". At the same time Roy (1952) developed his safety-first portfolio theory. In this model investors have the objective to minimize the probability of ruin with ruin meaning that wealth W is below subsistence level s (Shefrin & Statman, 2000). Thus, investors should first take care of their subsistence needs. These early studies have been influential. Kahneman and Tversky (1979) built their prospect theory on the work of Markowitz (1952b) and Lopes (1987) developed the SP/A theory on the foundation of the safety-first model.

To advance with critics on the building blocks of expected utility theory, several axioms of expected utility are violated in practice. A violation of the independence axiom of expected utility is known as the Allais paradox (Allais, 1953). In addition, procedural invariance, also known as preference reversal, and description invariance (framing effects) are other types of violations of expected utility. Hence, expected utility has several proven (theoretical and empirical) shortcomings.

Expected utility takes the objective value of a probability and the subjective value of an outcome. However, violations of EU and experimental evidence are at odds with the expected utility framework. Probability weighting has been shown to be non-linear. Hence, it is more realistic to assume the subjective value of a probability. Quiggin (1982) accommodated this by introducing rank-dependent utility (RDU) and Kahneman and Tversky (1979; 1992) developed (cumulative) prospect theory. These models assume that individuals transform objective probabilities by means of a weighting function. RDU makes use of decision weights. The weight attached to a state of nature depends on its probability as well as on its ranking. Figure 1 presents a common probability weighting function for non-expected utility models. Characteristics of this function are that small probabilities are overweighted and large probabilities are underweighted. Another aspect is that individuals are more insensitive for probabilities in the area in between the end points 0 and 1. This non-linear transformation of probabilities can explain a strong preference of investors for improbable large gains and aversion for improbable large losses (He, Kouwenberg & Zhou, 2017). In economic terms, rank-dependent preferences have the ability to accommodate

probability-dependent risk attitude: it can explain both risk averse as well as risk seeking behavior (Polkovnichenko, 2005).

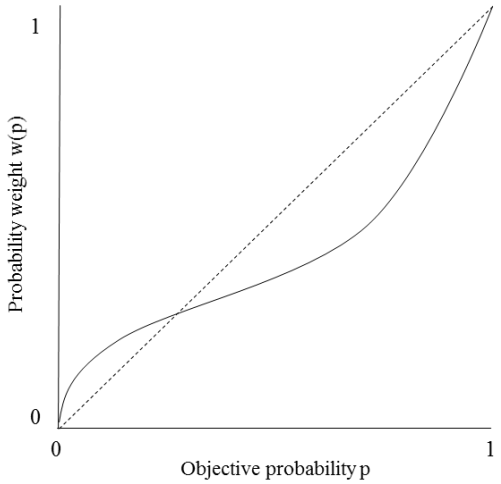


Figure 1 Probability weighting function. The dashed line represents the objective probability weight (linear weighting).

2.2 Household finance

Household finance is a field that studies how households use financial markets and instruments to achieve their ambitions (Guiso & Sodini, 2013). Households have to make many decisions about how to spend their money, how to save and invest their income and how to deal with various risks related to their assets and health. According to standard economic theory households should hold risky assets as part of a well-diversified portfolio (Von Gaudecker, 2015). Merton (1969) shows in his model that investors should both invest in risky markets and the market portfolio. However, studies that analyze portfolio choices of households observe that decisions made are mostly inconsistent with standard asset allocation models (Kapteyn & Teppa, 2009).

Households behave heterogeneously and make many 'mistakes' while investing. Campbell (2006) shows that the majority of households do not hold well-diversified portfolios. For those who do diversify their holdings this is often done naively, for example by allocating money using 1/*n* as a rule of thumb (Benartzi & Thaler, 2001). Another finding is that a lot of households do not even participate in the stock market, also known as the 'participation puzzle' (Haliassos & Bertaut, 1995). This has been a research topic for a long time. Explanations on why households do not participate are among others the costs of participation (Vissing-Jorgensen, 2003), non-standard

preferences of households (Barberis, Huang & Thaler, 2006), the role of trust (Guiso, Sapienza & Zingales, 2008) and financial illiteracy (Benartzi & Thaler, 2001). Households that do participate in the stock market trade excessively (Barber & Odean, 2000). These households would often earn more if they would trade less frequently. In addition, two related but distinct biases known as home (local) bias and familiarity bias appear in household trading data. The former refers to the observation that investors prefer to invest in domestic markets rather than foreign markets (this also holds for regional vs. non-regional stocks (French & Poterba, 1991; Cooper & Kaplanis, 1994)). The latter relates to the finding that a substantial group of (US) households allocate a significant fraction of their holdings in stock of their employer (Benartzi, 2001). Research shows that these investment mistakes are costly and result in welfare losses for households (Calvet, Campbell & Sodini, 2007; Guiso & Jappelli, 2006; Haliassos, 2003).

It should be noted that although these investment mistakes are ubiquitous, heterogeneity among individuals investors exist. Certain types of households are more susceptible to (psychological) biases and beliefs than others and therefore, the magnitude of all aforementioned mistakes varies. For example, Goetzmann and Kumar (2008) show that age, wealth, experience and financial sophistication positively affect the diversification level of stock portfolios. Graham, Harvey and Huang (2009) find that being more comfortable with financial products, i.e. being more financially sophisticated, predicts holding more internationally diversified portfolios.

The literature studying (under)diversification of household portfolios can be divided into two (methodological) strands. One strand studies this topic using administrative data from mainly Sweden, for instance Calvet, Campbell & Sodini (2007, 2009). Swedish authorities provide researchers with very detailed information about large amounts of households. A drawback of using these databases is the lack of covariates (Von Gaudecker, 2015). The other strand in the literature makes use of an abundance of covariates by drawing data from household surveys. Survey data has its own drawbacks such as missing observations and the impossibility of quantitative analysis.

2.3 Empirical findings

The academic literature has acknowledged the shortcomings of expected utility, however it still dominates the field of finance in terms of models of preferences (Polkovnichencko, 2005). This despite anecdotal (Friedman & Savage, 1948), experimental (Kahneman & Tversky, 1979; Lopes, 1987) and theoretical (Shefrin & Statman, 2000) evidence arguing otherwise. Two reasons

are underlying the continuation of EU in financial models: first, there is a lack of evidence that is non-experimental and second, the additional complexity of alternative models is not crucial in applying the models (Polkovnichencko, 2005). The lack of non-experimental evidence mostly implies a lack of empirical evidence. Polkovnichenko (2005) steps into this gap and provides empirical arguments and data for rank-dependent preferences. Using the Survey of Consumer Finances (SCF) data, a cross-sectional survey of US households, Polkovnichenko (2005, p. 1467) reports two main patterns inconsistent with expected utility: "(i) many households simultaneously invest in well-diversified funds and in poorly-diversified portfolios of stocks; and (ii) some households with substantial savings do not invest anything in equities. Households are shown to be aware of the risk associated with holding individual stocks. This idiosyncratic risk is not overlooked (Polkovnichenko, 2005). Other findings in this paper are that some households withhold themselves from owning equity although having a sufficient amount of savings, and that the majority of households owns a very low number of stocks. Polkovnichenko (2005) relates the under-diversification of households to rank-dependent preferences. Other researchers, such as Mitton and Vorkink (2007) explain under-diversification by a preference for skewness. They find that, using data from a large broker, undiversified investors have higher levels of skewness in their portfolio returns, which translates to a higher probability of very high payoffs, and that investors deliberately obtain return skewness. Mean-variance efficiency is sacrificed to obtain this higher level of (positive) skewness, for example by picking stocks that increase the skewness of their expected portfolio returns (Mitton & Vorkink, 2007).

In order to extend the current literature and to provide more empirical arguments this paper explores Dutch household data using the DNB Household Survey (DHS). More specifically, I will investigate whether Dutch households show investment behavior similar to their (American) counterparts and if so, how this behavior can possibly be explained by either rank-dependent preferences and/or a preference for skewness. Taking into account the (costly) investment mistakes made by individual investors it is relevant to study and better understand what drives and explains investment behavior of households in general and under-diversified portfolios specifically. By studying the investment behavior of households this paper enriches the (Dutch) household finance literature. Consequently, this leads to my research question:

How do (Dutch) households invest their assets and what is the role of preferences in this behavior?

To answer this research question I construct and test several hypotheses based on the academic literature.

The first hypothesis is derived from Polkovnichenko's (2005) finding that households simultaneously invest in well-diversified funds and in poorly-diversified portfolios of stocks.[2] This implies that there is a positive correlation between owning the two accounts. This results in hypothesis 1.

Hypothesis 1: Owning equity directly and indirectly is positively correlated.

Hypothesis 2 is closely related to the first hypothesis. Polkovnichenko (2005) provides evidence that households are aware of the risk associated with holding individual stocks. Section 3.4 will elaborate on the method to determine and test whether this hypothesis holds.

Hypothesis 2: Households are aware of the higher risk associated with undiversified portfolios of individual stocks.

Based on the finding of Polkovnichenko (2005) that some households with substantial savings do not invest in equities I construct my third hypothesis. In itself this observation is not testable. In order to be able to actually test this hypothesis I have to reformulate and generalize this finding. Underlying the observation by Polkovnichenko (2005) is the assumption that higher savings (financial assets) is related to investing in equity. Hence I will test the following hypothesis:

Hypothesis 3: Higher financial assets are positively related to participating in the stock market.

[2] Throughout this paper I use the word *stock(s)* to refer to the fact that a household has holdings in a company. This does not refer to the *number of shares* owned by a household, which is the amount of shares owned of a particular stock. Thus, a household owning five stocks has holdings in five individual and distinct companies.

3. Data and methods

3.1 DNB Household Survey

This paper uses data from the DNB Household Survey (DHS) of CentERdata. Among others, the DHS has been used (and proven valuable) to study the effect of risk aversion (Kapteyn and Teppa, 2009), financial literacy and financial advice (Von Gaudecker, 2005) and stock market expectation (Hurd, Van Rooij & Winter, 2011). Annually, economic data is collected through the CentERpanel that consists of approximately 2,000 households. The DHS comprises six modules[3] that cover economic and psychological determinants of saving behavior. Topics range from work and pensions to economic and psychological concepts. In terms of observable characteristics the panel is a representative sample of the Dutch population. To avoid selection bias, those having less are provided with the equipment and facilities to fill out the survey. Since the DHS is a micro-level panel dataset it allows for in-depth analysis of households' financial circumstances and economic behavior. I am particularly interested in the module *Wealth* that contains questions on households' assets and liabilities. It should be noted that a wave of year *t* refers to survey data from year *t-1* (the preceding year). Throughout this study I make use of different samples of the DHS data, expanding and limiting the scope in terms of waves, but this is indicated in the related paragraphs.

3.2 Household financial assets

In the first part of this study I analyze households' wealth and portfolios with the ultimate goal of investigating whether Dutch households have rank-dependent preferences. To this end I follow the methodological approach of Polkovnichenko (2005). I only include observations that are present in all six DHS modules of a wave. Moreover, I use data from five non-consecutive DHS waves to identify changes over time. Events such as the financial crisis might have affected households' attitude towards financial markets and hence impact their portfolios. I select a three year interval as this corresponds to the Survey of Consumer Finances (SCF) used by Polkovnichenko (2005) which is a three-yearly survey. The waves in my sample are 2004, 2007, 2010, 2013 and 2016. Economic behavior and more specifically, portfolio selection and diversification are intuitively influenced by the level of wealth of a household. In order to control

[3] See appendix 1 for a brief description of the six questionnaires.

for this, I construct four groups of households. This is done by dividing the sample of households based on their amount of financial assets (FA). Included in my selection of financial assets are checking-, savings- or deposit-accounts, deposit books, employer-sponsored savings plans, savings certificates, annuities, endowment insurance policies, mutual funds, bonds, shares, and put and call options. For more details and the classification of these assets please refer to Table 1. The classification of the accounts has been done according to Von Gaudecker (2015).

The four groups and their value ranges are:

1. *Financial assets* < €10,000
2. €10,000 < *Financial assets* < €100,000
3. €100,000 < *Financial assets* < €500,000
4. *Financial assets* > €500,000

Table 1 Portfolio accounts

Level 0	Level 1	Level 2
Checking accounts	Checking and saving accounts	Safe financial assets
Employer-sponsored savings plans	Cash value of insurances	Safe financial assets
Savings or deposit accounts	Checking and saving accounts	Safe financial assets
Deposit books	Checking and saving accounts	Safe financial assets
Savings certificates	Checking and saving accounts	Safe financial assets
Single-premium annuity insurance policies	Cash value of insurances	Safe financial assets
Savings or endowment insurance policies	Cash value of insurances	Safe financial assets
Growth funds	Mutual and growth funds	Risky financial assets
Mutual funds and/or mutual fund accounts	Mutual and growth funds	Risky financial assets[1]
Bonds and/or mortgage bonds	Bonds	Risky financial assets[2]
Stocks and shares	Shares	Risky financial assets
Put options	Options	Risky financial assets
Call options	Options	Risky financial assets

Note: 1 and 2 in the table refer to the fact that government bonds and mutual funds that invest in government bonds are not regarded as risky financial assets. These type of investments are classified as safe financial assets. In case households did not specify in which type of bond or mutual fund they invest, these households are assigned to the default. The default for mutual funds is risky financial assets and the default for bonds is also risky financial assets. Similarly, I was unable to assess the type of bond owned by a household for the years 2004 and 2007 as the DHS did not include a relevant question in those years. Bonds in these years are assigned to risky assets, which is the most common type of bond in subsequent years. Safe financial assets and risky financial assets are the two components of the account financial assets.

In order to construct the account 'financial assets', all separate accounts are added up. Respondents have either the option to enter a specific amount for an account or to indicate a value range in case they do not know the exact amount of an account. In the latter case, the average of

that range is computed and inserted as the value for that particular account. For the ranges *0 - X* and *X >* the lower bound is selected.[4] All values are adjusted for inflation, using the CPI and 2004 as base year. Households with financial assets having a value of 0 are deleted from the sample as naturally their allocation cannot be analyzed.

One of the key concepts to differentiate between households is 'stockholder'. Stockholders are defined as households with a share in a company either through stocks and/or put/call options. This definition and the definitions of other key concepts and variables in this paper can be found in Table 2.

Table 2 Definitions of key concepts and variables

Variable/concept	Definition
Stockholder / Owner of equity	Household owning equity either directly through shares and/or put/call options or indirectly through mutual funds or both.
Direct owner of equity / Direct investor	Household owning equity directly through shares and/or put/call options
Indirect owner of equity / Indirect investor	Household owning equity indirectly through mutual funds (mutual funds that invest in bonds excluded)
Direct equity value	Value of directly owned equity
Risky financial assets	Value of shares + put/call options + mutual funds (bond funds excluded) + bonds (non-government bonds only)

The definitions specified in Table 2 are key to understand the results shown in Table 3 and beyond. Table 3 reports data on the distributions of financial assets and equity of households. The first two columns of the table show the median values of financial assets per group in a specific year for both non-stockholders and stockholders. Across all groups and all years stockholders have higher amounts of financial assets than non-stockholders. The only exception is the median value for group 4 in 2016, but this is due to a low amount of observations in that particular group. The values remain fairly stable over time and it is hard to observe a definite trend. The next two columns show the percentage of total financial assets owned by a particular group in a year for both, all households, and for stockholders only. As non-stockholders naturally also own financial assets the percentages for stockholders do not necessarily add up to 100%. Over the past decade, the proportion of total financial assets owned by stockholders fluctuated around 50%. The percentage owned by a particular group in a particular year is mainly explained by the amount of households in that sample group in that year. However, the pattern seems stable over time. Groups

[4] The range X> refers to the range of values where X is the lower bound and the upper bound is undefined (for example the range 50,000 >). In those cases the lower bound X (in the example 50,000) is selected.

1 and 4 account for a small fraction of total financial assets, and groups 2 and 3 respectively account for most of the financial assets.

The next three columns focus on the distribution of equity among the groups in the sample. The first column shows the percentage of total equity that is held by a particular group. In the two subsequent columns this percentage is split up between indirectly and directly owned equity. Directly owned equity is defined as owning stocks and/or put options and/or call options. Indirectly owned equity respectively represents the money invested in equity through mutual funds. In light of this study the overall division between direct and indirect equity is most interesting. Toward 2010 it seems that holding direct equity gains popularity. At least, the value of directly held equity increases in that period to approximately 52% of total equity value. However, this trend reverses, leading to a significant increase in holdings of indirect equity. In 2016, more than 70% of equity value is held indirectly and the value of directly held equity almost halved relative to 2010. This pattern can be seen in the last three columns of Table 3 as well. These columns show the distribution of the population over the different groups. Firstly, the percentage of households in a group are shown. Subsequently, respectively the percentage of households in a group that are stockholders and direct owners of stock are presented. The percentage of stockholders in the population increases in the years before 2010 and decreases in the years after. Similarly, the proportion of households that are direct investors slightly increases before 2010 and drops afterwards. It seems that a decreasing proportion of all households holds stocks (directly) in recent years.

Table 3 Distributions of financial assets and equity

	Financial assets (FA)									
	Median Values (€)		% of aggregate		Equity (% of aggregate)			Households (% of population)		
DHS year and group	Non-stockholders	Stockholders	All	Stockholders	All	Indirect	Direct	All	Stockholders	With direct stocks
2004										
1	2,826	6,477	4.2	0.6	1.1	0.6	0.6	44.1	4.0	3.0
2	25,121	35,301	40.4	15.4	20.0	11.2	8.8	46.5	15.1	9.2
3	144,032	151,170	41.1	31.1	69.1	36.5	32.6	8.9	6.4	4.2
4	-[5]	731,660	14.3	7.8	9.8	7.7	2.1	0.5	0.4	0.4
All	8,600	41,931	100.0	54.9	100.0	56.0	44.1	100.0	25.9	16.8
2007										
1	2,298	3,661	3.0	0.3	0.4	0.3	0.1	43.4	3.3	2.4
2	24,516	45,705	36.2	15.8	19.3	13.1	6.2	45.8	15.6	7.3
3	136,229	169,905	23.5	28.8	42.4	26.6	15.8	9.6	7.0	4.4
4	539,631	660,724	37.3	16.3	38.0	8.7	29.3	1.2	0.9	0.7
All	7,939	57,452	100.0	61.2	100.0	48.7	51.4	100.0	26.8	14.8
2010										
1	2,393	6,051	2.3	0.3	0.4	0.3	0.1	34.8	3.1	1.9
2	26,044	46,015	38.7	16.7	18.3	12.7	5.6	53.3	18.5	9.7
3	143,230	145,262	35.5	21.2	28.3	18.3	9.9	10.8	6.6	4.4
4	-	694,985	24.4	24.4	53.0	16.3	36.7	1.1	1.1	0.1
All	11,803	56,957	100.0	52.6	100.0	47.6	52.4	100.0	30.3	16.1
2013										
1	1,890	4,889	3.3	0.4	0.7	0.4	0.3	42.0	2.9	1.9
2	27,543	40,522	45.7	16.0	26.3	20.1	6.2	48.0	14.0	6.5
3	145,689	153,865	42.1	26.0	58.6	33.1	25.6	9.5	5.9	3.2
4	506,949	615,666	9.0	7.5	14.3	11.8	2.5	0.6	0.4	0.2
All	9,444	53,596	100.0	49.9	100.0	65.4	34.6	100.0	23.2	11.8
2016										
1	2,154	4,760	4.1	0.4	0.8	0.4	0.4	45.0	2.6	1.7
2	24,132	42,617	48.7	16.8	24.3	18.0	6.3	47.0	12.4	6.4
3	147,306	151,262	38.1	27.3	55.8	40.9	14.9	4.5	5.1	2.6
4	910,236	702,904	9.1	7.2	19.1	12.5	6.6	0.4	0.4	0.2
All	8,655	55,301	100.0	51.7	100.0	71.8	28.2	100.0	20.5	10.9

Note: Financial assets and group division are as defined in the paper. Stockholder is defined as holding equities (both directly and/or indirectly).

[5] This group only consists of one observation with an extremely high value of 2,471,938. Including this observation would distort the results. Therefore, this observation is omitted from the table.

These findings are not completely consistent with trends that are observed in other reports. The *Fact Sheet Loans, Savings and Investments* of the Dutch Banking Association (2017) acknowledges and underlines the increasing importance of private capital accumulation. Accordingly, over the past years an increase in both direct and indirect investment is observed among Dutch households (Schroders Investment Barometer, 2015).

An interesting observation that can be made based on the data is the non-participation in the stock market. Table 4 shows the number of observations per investment type. For example, over the years, the amount of households that own stocks ranges from approximately 10% to approximately 17%. Most striking about the finding that many households do no participate is that these same households have substantial financial assets in other accounts. Although the median values for financial assets are higher among stockholders, the median values among non-stockholders are substantial. This finding is consistent with Polkovnichenko's (2005) results for American households. As noted by him, this is inconsistent with expected utility theory and its frictionless models. Theoretically and to some extent practically, participation costs might prevent households from investing in the stock market (Gomes & Michaelides, 2005). However, for the richer cohorts, this argument seems invalid. In sections 3.6 and 4.3 I will investigate the determinants of stock market participation more in-depth. Table 4 also shows that only a small subset of the investors chooses to (partly) invest their money in foreign stocks. From a theoretical point of view, this cannot be explained. However, taking into account home (local) bias, this behavior can be better understood.

Table 4 Number of observations per type of investment

Variable	2004	2007	2010	2013	2016
Total number of households	943	1017	891	901	1381
# investing in stocks	156	147	150	107	147
# investing in stocks foreign companies	-	-	38	24	36
# investing in mutual funds	185	217	224	175	212
# investing in bonds	48	55	39	29	34
# investing in put/call options	3	9	6	5	7

Note: The DHS data for waves 2004 and 2007 does not allow to determine whether a household invests in foreign companies

3.3 Household equity portfolios

The depth of the questions in the DHS allows to take a closer look at the equity portfolios of households. In the first column of Table 5 the fraction of equity that is directly invested is shown per group. Again, directly owned equity is defined as the investments in stocks, put, and/or call options. It can be seen that for all years the least wealthy group in terms of financial assets mainly invests their money directly. Additionally, the accompanying column on the right shows that these groups have a very limited amount of stocks. Hence, most of the (direct) equity owners in these groups are badly diversified in terms of their stock portfolio. The number of stocks owned is increasing in the groups with a higher amount of financial assets. Across all years, the wealthiest group has (not surprisingly) the highest number of stocks. However, it seems that the median number of stocks for all groups is decreasing over time.

The level of diversification cannot be analyzed and assessed only on the basis of the equity portfolios of households. It must be seen in perspective of the total financial asset portfolio. Therefore, in the three most right columns equity portfolios are presented relative to the total financial assets of households. Median fractions of all equity, directly held equity and indirectly held equity relative to financial assets are reported. It seems that all fractions have remained fairly stable over the 12–year timeframe covered by this study. However, the median fractions invested directly in equity slightly decreased whereas the indirect fractions slightly increased. This is in line with the results of paragraph 3.2. Interesting to observe is that the median investor in household cohort 1 in year 2016 invests all equity directly into one stock and this investment accounts for 63% of the investor's total financial assets. This implies that a significant part of their (financial) assets is invested in an undiversified and risky portfolio.

Panel A of Table 6 shows the percentage of direct investors that combines investing in individual stocks with investments in mutual funds. The percentage of households investing in both is substantial, ranging from 44.2% to 63.3%, and the data seems to show a similar pattern over time as is observed for overall stock market participation. To investigate more formally if households simultaneously invest in well-diversified funds and badly diversified portfolios of individual stocks I investigate the correlation between these two variables.

Table 5 Equities in household portfolios from the DNB household survey (DHS)

	Equity portfolio of direct investors		*Equity relative to total financial assets*		
DHS year and cohort	Direct fraction (relative to all equity owned)	Number of stocks	All equity	Indirect	Direct
Panel A: 2004					
1	1.00	1	0.44	0	0.05
2	0.93	3	0.22	0.08	0.01
3	0.73	3	0.23	0.11	0.07
4	0.42	5.5	0.40	0.13	0.03
All	1.00	2	0.26	0.09	0.03
Panel B: 2007					
1	1.00	1	0.53	0.19	0
2	1.00	2	0.29	0.13	0
3	0.62	3	0.42	0.16	0.02
4	0.72	10	0.51	0.15	0.24
All	0.83	2	0.34	0.15	0
Panel C: 2010					
1	1.00	1	0.36	0.21	0
2	0.5	2	0.23	0.18	0
3	0.44	2	0.30	0.19	0.01
4	0.77	6	0.81	0.10	0.15
All	0.57	2	0.26	0.18	0
Panel D: 2013					
1	1.00	1	0.39	0.06	0
2	1.00	1	0.22	0.11	0
3	0.62	3	0.32	0.16	0
4	0.36	4	0.34	0.31	0.03
All	1.00	2	0.27	0.13	0
Panel E: 2016					
1	1.00	1	0.63	0.07	0
2	1.00	1	0.28	0.14	0
3	0.64	2	0.44	0.27	0
4	0.50	4	0.60	0.50	0.07
All	1.00	1	0.31	0.18	0

Note: The values shown in the table are median values. Only observations having a nonzero amount of equity (either direct or indirect) are taken into account for constructing this table.

Panel B and C of Table 6 show the correlation coefficients between directly owning and indirectly owning equity. The results in panel B show a (weak to moderate) positive correlation when all direct owners are included. Restricting the sample to households owning one individual stock, which is by definition undiversified, results in a (weak) positive correlation as well. At least this indicates that investing in one of the two accounts does not rule out investing in the other (this is what a negative correlation would suggest). Instead the results suggest that it is more likely that a household invests in one of the two accounts if the household is already investing in the other. Hereby, this is in line with hypothesis 1.

Table 6 Combining direct and indirect ownership of equity

Measure	2004	2007	2010	2013	2016
Panel A					
% direct investors investing indirectly	52.6%	52.4%	63.3%	50.5%	44.2%
Panel B					
Correlation coefficient - All direct owners	0.37	0.31	0.40	0.29	0.28
Panel C					
Correlation coefficient - Owners of one stock	0.17	0.14	0.18	0.10	0.08

Note: Panel A shows the percentage of direct investors that also invests indirectly. Panel B and Panel C show the correlation coefficients for the variables *direct_owner and indirect_owner* across the five waves. Panel B presents the correlation coefficients for all direct owners. Panel C presents the correlation coefficients for the direct owners with only one stock. All correlation coefficients are significantly different from zero.

Figure 2 gives a visual representation of the notion that households generally do not diversify enough. The cumulative distributions of direct investors and the cumulative value of directly held stocks are portrayed, as a function of the numbers of stocks owned by a household. Across all five waves a nearly identical (and time invariable) pattern emerges. About 95% of the households own stocks of five or less companies. Almost 90% of all wealth is owned by this group of households. Between 40% and 50% of all wealth invested in equity is held by households with only one stock.

Similar to the findings of Polkovnichenko (2005) amongst households in the U.S., it seems that households in the Netherlands hold undiversified portfolios of individual stocks and at the same time invest in well-diversified mutual funds. To analyze more in-depth if households deliberately apply a preference-based approach to their portfolio diversification I turn to regression analysis.

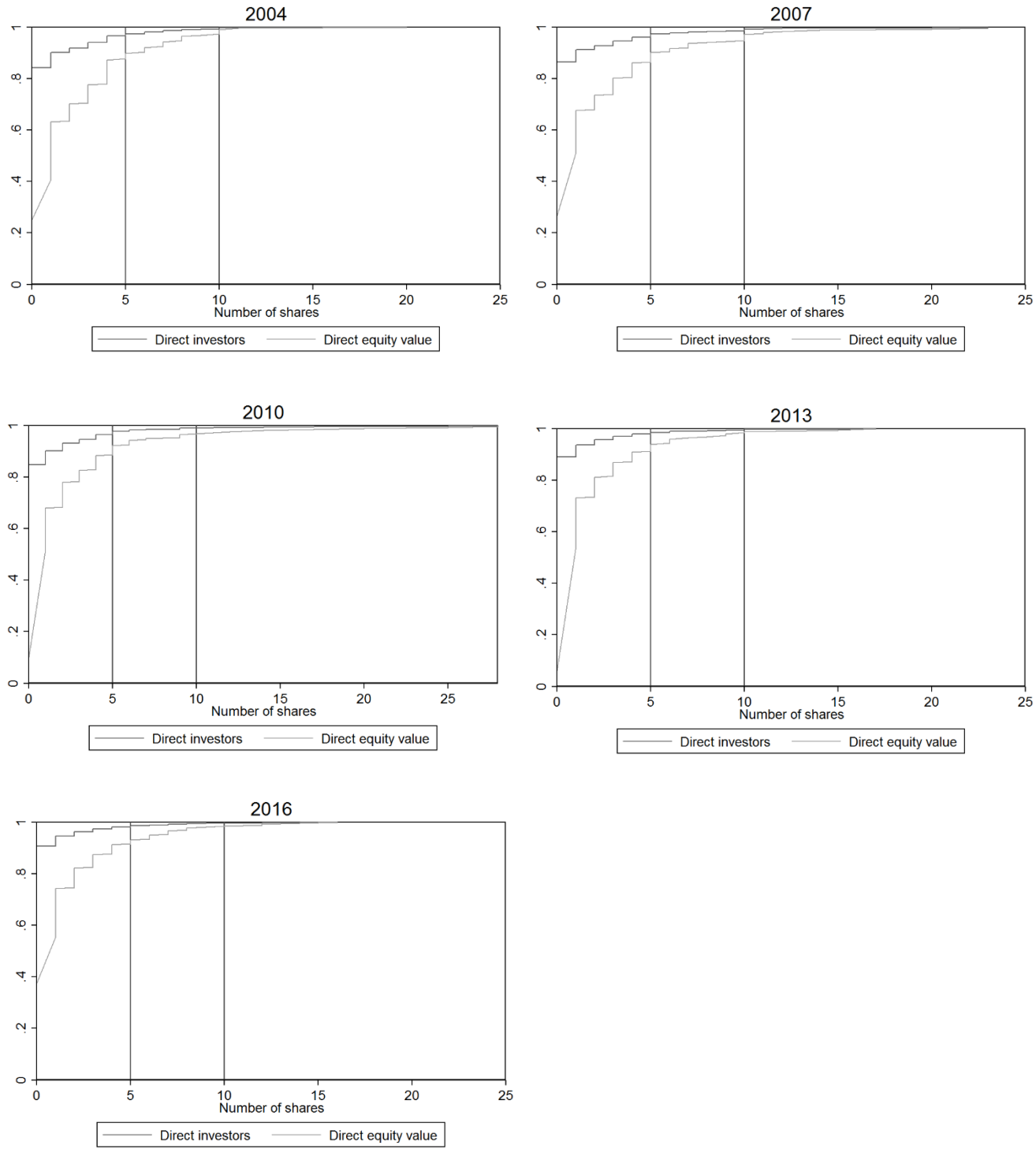


Figure 2 Cumulative distributions of direct investors and direct equity value as functions of the number of stocks in direct portfolios from the DNB Household Survey (DHS)

3.4 Awareness of riskiness portfolios

In the academic literature there are several studies that explain that biases, such as familiarity bias (French & Poterba, 1991) and overconfidence (Barber & Odean, 2000), play a role in holding undiversified portfolios. This type of explanation assumes that investors are not aware of the risks of under-diversification (Polkovnichenko, 2005). Following the approach of Polkovnichenko (2005) I use OLS regression analysis to test if households that directly invest realize that undiversified portfolios are riskier. If I can relate the level of diversification with the risk attitudes of household investors this would be consistent with a preference-based explanation and inconsistent with a bias-based explanation.

To this end, two different dependent variables are constructed. Both variables are measures of diversification. Firstly, I take the proportion of total financial assets directly invested in stocks ($\frac{DIR}{FA}$). Additionally, as a second measure of diversification and as a second dependent variable, I take the proportion of risky assets invested in directly held stocks $\left(\frac{DIR}{RIS}\right)$. The accounts that make up the variable risky assets have been outlined in Table 1. The rationale behind the first dependent variable is that it tests whether investors are aware of the overall risk level of individual stocks. Regarding the second variable, this construct tests whether households are aware of the relative risk level of this type of investments relative to other investments. Unlike Polkovnichenko (2005) the DHS does not allow to restrict the sample to households who (reportedly) traded stocks minimally three times in the year before the survey. Unfortunately, the DHS does not contain questions on the trading frequency of households. Therefore, I cannot distinguish between households trading actively and inactively (long-term investment in certain stocks). The reason for Polkovnichenko (2005) to introduce this restriction is to exclude households that unintendedly hold undiversified portfolios. Consequently, my results might be more conservative (underestimation of the effects) and should be treated with caution.

Both regressions consist of the same four independent variables. Initially, I use the same regressors as selected by Polkovnichenko (2005). *Log(FA)* is included to measure the effect of wealth and is measured by the natural logarithm of total financial assets. *Edu* is used to account for the effect of education on portfolio allocation. Education is measured by the highest education completed by the respondent. The variable is coded as 1 for no/elementary/secondary education, 2 for higher vocational education and 3 for academic education. Both wealth and education might

be related and or indicative to how prone individuals are to (psychological) biases (Polkovnichenko, 2005). Thirdly, the number of persons in the households, also referred to as 'dependents' is included in the regression (*Dep).* Lastly, I try to relate risk attitude to portfolio selection. I construct a measure of risk attitude based on questions in the survey and apply the methodology of Kapteyn and Teppa (2009). According to Kapteyn and Teppa (2008) this method captures risk aversion more directly than any other subjective measure explored in their study. Eight survey questions are selected, six related to risk aversion and two to precaution, to develop the risk attitude variable. Regarding the risk aversion questions, respondents have to answer to what extent they agree or disagree with statements related to money and investments. The two remaining questions ask respondents how they rate the importance of saving money for precautionary reasons. Because of the wording of some of the questions they are reversed for consistency. Appendix 3 shows the (reversed) questions. A factor analysis is applied to investigate the underlying structure of the eight indicators. The factor analysis is a principal components analysis (PCA) with varimax rotation (Kapteyn and Teppa, 2009). In this method, extraction is based on eigenvalues greater than 1. As a result of the analysis, it appears that three factors can explain most of the variation in the answers. Table 10 in Appendix 3 shows the respective factor loadings. I use the same interpretation as Kapteyn and Teppa (2009) in that Factor 1 and Factor 2 best correspond with the six questions on investments and that Factor 3 relates to the questions on precautions. Likewise, the first two factors are interpreted as measures of risk aversion and the third factor as a measure of precaution. Consequently and most interestingly with respect to this study, Factor 1 and Factor 2 are expected to play a role in portfolio choice. From now on, I will refer to the first two factors as "RA1" and "RA2". Both the variable *Dep* as well as the *RA* factors function as proxies for the willingness to take (financial) risk.

This results in the two following regression equations:

$$\frac{DIR}{FA} = \beta_o + \beta_1 log(FA) + B_2 Edu + \beta_3 Dep + \beta_4 RA1 + \beta_5 RA2 \quad (1)$$

$$\frac{DIR}{RIS} = \beta_o + \beta_1 log(FA) + B_2 Edu + \beta_3 Dep + \beta_4 RA1 + \beta_5 RA2 \quad (2)$$

As previously mentioned, *log(FA)* and *Edu* are expected to have a significant effect on both dependent variables as wealthier and more educated households are more likely to be not

susceptible to biases and this should impact portfolio allocation. Moreover, if education has an effect on the choice of portfolios, this would suggest that households know the differences between types of investment options and make deliberate choices (Polkovnichenko, 2005). This would be understood as having more knowledge of portfolio (under)diversification.

The proxies for risk-taking willingness, *Dep* and *RA*, measure if households are aware of the higher risk level of individual (directly held) stocks. If households are aware, both variables should be significant. However, if investors are not aware, the variables are expected to be not significant. Negative (and significant) coefficients would indicate that households are both aware and allocate their portfolio accordingly. More dependents (members in a household) and higher risk aversion expectedly reduce the willingness to take risk and thus, invest less in directly held stocks.

3.5 Preference for skewness

Based on the first descriptive statistics it appears that Dutch households hold a limited amount of stocks. This is in line with other studies that consistently, across different countries and time periods, find that investors are often (heavily) under-diversified (e.g. Friend & Blume, 1975; Goetzmann & Kumar, 2008). From a mean-variance framework perspective this is often referred to as a puzzle. According to standard portfolio theory an under-diversified investor unnecessarily accepts a higher risk without being compensated accordingly with higher expected returns (Mitton & Vorkink, 2007).

Following the approach of Mitton and Vorkink (2007) I investigate, as an alternative to rank-dependent preferences, whether the under-diversification of households can be explained by a preference for skewness. It might be that in a mean-variance-skewness framework the level of diversification makes (perfect) sense. The sample I use for this additional analysis contains the respondents of DHS wave 2016, plus the respondents of wave 2013 that were not part of the 2016 wave, in order to increase the sample size. To stay close to Mitton and Vorkink (2007), I focus throughout this additional analysis on direct investments in stocks, thereby leaving out indirect investments.

First step in this process is to gain access to detailed information about households' portfolios. Upon request, I receive information on the specific stocks owned by the DHS respondents. Subsequently, I obtain monthly return data from Datastream to measure the performance of all stocks in the household database. The data is obtained for a period of 60 months.

Since the DHS survey data of a specific year refers to the preceding year this implies monthly data from 2008 up to and including 2012 for the 2013 wave and from 2011 up to and including 2015 for the 2016 wave. The monthly return data of each stock is matched with the portfolios of the respective household(s). Due to the fact that the survey data is self-reported the quality of the data is imperfect. Not all households report the name of a stock and/or do not provide the value of their holdings whereas others specify their holdings ambiguously or incorrectly. Moreover, not from all stocks return data can be retrieved. As a result of these issues, the sample reduces significantly. Besides return data of stocks, data on the MSCI Europe Index and the one-month EURIBOR are obtained for the same time periods. The MSCI Europe Index proxies the return on the market portfolio. As the Netherlands is a member of the Eurozone and households' investments are mostly done in the same area, this is a natural choice (Von Gaudecker, 2015). The one-month EURIBOR rate is based on the average interest rates at which a (large panel) of European banks borrow from one another and approximates the risk-free rate.

Crucial in this additional analysis is the level of (under)diversification of households. Throughout this paper I will make use of three different measures of diversification (Mitton & Vorkink, 2007). All three measures are calculated at the latest point in time possible. This translates to data from 01-01-2013 for the 2013 wave and data from 01-01-2016 for the 2016 wave.
The first diversification measure, D_1, is the least complicated. It is defined as

$$D_{1,j} = \frac{1}{n} \tag{3}$$

where n is het number of stocks in household j's portfolio. The advantage of this measure is its simplicity and its consistency with the two other measures (all three associate higher values of D_j with lower levels of diversification). The disadvantage of this measure is its inability to capture the weight of the individual stocks in a portfolio. This measure assumes an equal allocation of money over all stocks. The second diversification measure, D_2, accounts for the weight of individual stocks and is comparable in its approach to a Herfindahl index. It is defined as

$$D_{2,j} = \sum_{i=1}^{n} w_{ij}^2 \tag{4}$$

where w_{ij} is the weight of stock i in household j's portfolio. Again, higher values of this diversification measure are associated with lower levels of diversification. Although D_2 is advantageous over D_1 it is still flawed. Both the first and second measure do not account for the covariance between stocks in a portfolio. A portfolio with two equally weighted stocks in industry A is treated in the same way as a portfolio with two equally weighted stocks of which one in industry A and one in industry B. Therefore, D_3, is constructed. It accounts for covariance between stocks and is defined as

$$D_{3,j} = \sum_{i=1}^{n} w_{ij}^2 + \left(1 - \sum_{i=1}^{n} w_{ij}^2\right) \bar{\rho}_j \tag{5}$$

where $\bar{\rho}_j = \sum_{h=1}^{n} \sum_{i=1}^{n} w_{h,j}\, w_{i,j} \rho_{hi} \,/\, \sum_{h=1}^{n} \sum_{i=1}^{n} w_{h,j}\, w_{i,j}$. The interpretation of this measure is directionally consistent with the previous measures. D_3 takes into account both the weights of stocks in a portfolio as well as the correlation between stocks in a portfolio. A closer look at the measure shows that higher levels of diversification can be achieved by either reducing the concentration of stocks in a portfolio or by reducing the correlation of the stocks in a portfolio. Although this measure is more sophisticated than the previous two it is not without shortcoming. As indicated by Mitton and Vorkink (2007) the weight that Equation 5 places on both sources of diversification, concentration and correlation, is rather arbitrary. Moreover, to assess the correlation between stocks data from Datastream is required on the specific stocks. This data and/or the relevant survey data is not available in all cases, which results in a (slightly) reduced sample size for analyses that use measure D_3.

In addition to three measures for diversification I construct a measure for skewness based on Mitton and Vorkink (2007). The skewness of portfolio returns is measured by the skewness coefficient, which is defined as

$$\hat{\mu}^3 = \frac{\frac{1}{60} \sum_{t=1}^{60} (r_t - \mu)^3}{\hat{\sigma}^3} \tag{6}$$

where the skewness coefficient is calculated from a 60-month window of monthly returns. Moreover, $\hat{\sigma}^3$ is the cube of the estimated portfolio return standard deviation (Mitton & Vorkink, 2007). Not in all cases it was possible to use a 60-month window. The threshold that is used to

calculate the skewness coefficient is a minimum availability of data for the 12 preceding months. Consistent with the diversification measures the skewness coefficient is calculated at the latest point in time possible for both waves. Important to note is that in constructing this measure the portfolio composition of a household is held constant over time. Hence, it is assumed that households have exactly the same portfolio (constant concentration and constant correlation of stocks) for the five years preceding the measurement date. This is (naturally) unrealistic, however taking into account the constraints of the survey data I deem this the best alternative option.
Equation 6 accounts for the (positive) correlation between variance and skewness. The equation measures in essence the incremental skewness on top of what would be expected given the variance in portfolio returns (Mitton & Vorkink, 2007).

After constructing these measures I can sort the household portfolios according to their level of diversification. I sort based on measures D_2 and D_3. Preferably I would sort the households in deciles or quintiles, however the data and sample size do not allow this. Due to the fact that the majority of households has either 1, 2 or 3 stocks the diversification measures are clustered around certain values. Therefore, the households can only be sorted into tertiles. Still, this is imperfect since observations are not equally sorted in one of the three groups. Tertile 1 corresponds to the most diversified portfolios and tertile 3 to the least diversified portfolios. After sorting the households into the tertiles the portfolio returns are calculated for the 60 subsequent months. I then calculate the average mean, variance and skewness coefficient of the monthly return for each household portfolio. Table 7 shows average statistics per tertile. Panel A and Panel B are sorted on diversification measure D_2 and D_3 respectively. The mean monthly return is slightly lower for less diversified portfolios, whereas the variances is slightly higher for undiversified portfolios. In panel A, with portfolios sorted on D_2, there seems to be clear pattern for the skewness coefficient. Less diversified portfolios have a substantially higher skewness coefficient. For portfolios sorted on D_3 this is not the case.

Table 7 Monthly return statistics of household portfolios sorted by level of diversification

Panel A: Portfolios sorted on D_2						
Tertile	N	D^2 mean	Mean	Variance	Skewness	Sharpe ratio
1 (High diversification)	29	0.25	0.26%	0.002	-0.139	0.113
2	29	0.48	0.33%	0.003	0.065	0.118
3 (Low diversification)	85	0.93	0.21%	0.005	0.189	0.129

Panel B: Portfolios sorted on D_3

Tertile	N	D^3 mean	Mean	Variance	Skewness	Sharpe ratio
1 (High diversification)	28	0.319	0.26%	0.002	0.012	0.111
2	28	0.557	0.28%	0.002	-0.041	0.128
3 (Low diversification)	83	0.954	0.24%	0.005	0.004	0.126

Note: This table presents household portfolio return statistics for the 60 months preceding portfolio formation. Household portfolios are sorted into tertiles based on their level of diversification, respectively measured by D_2 and D_3. Please refer to Equation 4 and 5 for more details. The number of portfolios in a tertile is given by *N*. The mean return is the average monthly return for portfolios in the tertile over the subsequent 60 months. The variance (skewness) of returns is the variance (skewness) of the 60 monthly returns, calculated for each household portfolio and averaged over all portfolios in the tertile.

To check for mean-variance efficiency I construct for each household its Sharpe ratio, which is a performance measure that adjusts returns for its level of risk. Performance measurement of an investment (portfolio) should take into account both return and risk. The Sharpe ratio is the most suitable measure in my case since I focus on a part of households' investment portfolio and not on the total portfolio. After constructing this measure, it allows to compare the performance of different portfolios. The Sharpe ratio is defined as

$$\frac{E(r_j) - r_f}{\sigma_j} \tag{7}$$

where $E(r_j)$ is the expected return of the portfolio of household j, r_f is the risk-free rate (measured by the one-month Euribor) and σ_j the standard deviation of the portfolio's returns. The expected return of a portfolio is based on the monthly returns for a period of 60 months prior to the measurement date, hereby calculating the *ex-ante* Sharpe ratio. The intuition of the ratio is that it gives the level of return per unit of risk. It divides excess return of a portfolio by the volatility of a portfolio. A higher (lower) Sharpe ratio is associated with superior (inferior) performance. The last column of Table 6 displays the average Sharpe ratio per tertile. It seems that the ratio is stable across the tertiles.

In light of this paper I am particularly interested in the relationship between the Sharpe ratio and the level of diversification and level of skewness of a households' portfolio. In their work, Mitton and Vorkink (2007) find a negative and statistically significant relationship between the Sharpe ratio and both D_3 and the skewness coefficient. Higher levels of diversification, which

corresponds with a lower D_3 value, are associated with higher Sharpe ratios. Moreover, higher skewness is related to a lower Sharpe ratio.

Using the Sharpe ratio, mean-variance efficiency can be related to both the portfolio diversification level and the skewness of a portfolio. For this purpose, I follow Mitton and Vorkink (2007) in estimating a non-parametric relationship between the variables of interest. I use a kernel-weighted local polynomial regression, also known as local polynomial smoothing. Firstly, non-parametric regression estimation is less rigid than its parametrical counterpart. It imposes fewer restrictions, or structure, on the relationship between the variables and covariates. It allows 'data to speak for themselves'. A local polynomial regression involves fitting the response to a polynomial form of the independent variable using locally weighted least squares (Fan and Gijbels, 1996). I use the Nadaraya-Watson kernel estimator[6], which is defined as

$$sr(D^i) = \frac{\sum_{j=1}^{n} k\left(\frac{D_j^i - D^i}{h}\right) sr_j}{\sum_{j=1}^{n} k\left(\frac{D_j^i - D^i}{h}\right)} \tag{8}$$

where h is a bandwidth parameter, $k(.)$ is a kernel function and n is the number of households. This estimator allows to quickly obtain, without assuming any functional form, the relationship between two variables. The y (dependent) variable in my case is the Sharpe ratio and the x (independent) variable is respectively D_3 and the *skewness coefficient*. Equation 8 displays the estimator for the relationship between the Sharpe ratio and a diversification measure. D_i is replaced by the relevant explanatory variable when estimating the relationship.

The exploration of a possible preference for skewness among investors in this section requires to construct two additional hypotheses. The first hypothesis is based on the finding of Goetzmann and Kumar (2004) that under-diversified portfolios are less efficient than diversified portfolios and the similar finding of Mitton and Vorkink (2007) that Sharpe ratio are lower for less-diversified portfolios. This leads to hypothesis 4A.

[6] The Nadaraya-Watson estimator was first proposed by Nadaraya (1964) and Watson (1964).

Hypothesis 4A: Mean-variance efficiency, measured by the Sharpe ratio, is positively (negatively) related to the diversification level of household's portfolios (diversification measure D_3).

Based on Mitton and Vorkink (2007) I hypothesize that households with investment portfolios having a higher skewness coefficient will have a lower Sharpe ratio and thus, a lower mean-variance efficiency. Hypothesis 4B is defined accordingly.

Hypothesis 4B: Mean-variance efficiency, measured by the Sharpe ratio, is negatively related to portfolio skewness as measured by the skewness coefficient.

3.6 Participation in the stock market

As can be seen in Table 1 household participation in the stock market is low in the Netherlands. The Dutch participation rate is substantially lower than in the U.S. as found by Polkovnichenko (2005). Moreover, those who do participate in the stock market hold a low number of stocks. Therefore, I turn to regression analysis and analyze what determinants are for participation in the stock market and for holding (only a few) individual stocks. The models are estimated using data from DHS wave 2016. As dependent variables I use two dummy variables indicating whether a household is either a *stockholder* or a *direct_owner*. Following Guiso, Sapienza, and Zingales (2008) these models are estimated using a probit model. A probit regression is a suitable model to estimate relationships with a binary outcome variable as the relationship with the independent variables is in this case non-linear. A probit model estimates the probability that a value will fall into one of the two possible outcomes. Hereby it assumes a normal distribution of errors. Applying a probit model results in a change in interpretation of the coefficients (as compared to OLS), but I will elaborate on this in paragraph 4.3.

Moreover, I use the number of stocks (*number_stocks)* of a household as a dependent variable to test what influences households to hold more stocks (hereby assuming that holding more stocks is associated with better diversification). This model is estimated using standard OLS.

Additionally, I estimate a regression with *investing_abroad* as a dummy dependent variable. This analysis also makes use of a probit model. Generally, investing in foreign markets (on top of the home market) should be beneficial for the diversification level of households.

The independent variables that I include in the regressions are *log(FA), log(income), female, age, Edu, Dep, RA(1)* and *RA(2). Log(FA)* and *log(income)* are the logarithmic

transformations of the variables financial assets (*FA*) and gross income (*income)*. Female is a dummy for gender and age is divided in three groups coded as (1) ≤ 40, (2) 41 ≤ X ≤ 64, (3) ≥ 65. *Edu, Dep, RA(1)* and *RA(2)* are as defined earlier.

In equation terms the regressions are estimated as follows:

$$P(Stockholder = 1) = \beta_o + \beta_1 log(FA) + \beta_2 log(income) + B_3 Female + B_4 Age + B_5 Edu + \beta_6 Dep + \beta_7 RA(1) + \beta_8 RA(2) \quad (9)$$

$$P(Direct\ owner = 1) = \beta_o + \beta_1 log(FA) + \beta_2 log(income) + B_3 Female + B_4 Age + B_5 Edu + \beta_6 Dep + \beta_7 RA(1) + \beta_8 RA(2) \quad (10)$$

$$Number\ shares = \beta_o + \beta_1 \log(FA) + \beta_2 \log(income) + B_3 Female + B_4 Age + B_5 Edu + \beta_6 Dep + \beta_7 RA(1) + \beta_8 RA(2) \quad (11)$$

$$P(Investing\ abroad = 1) = \beta_o + \beta_1 \log(FA) + \beta_2 \log(income) + B_3 Female + B_4 Age + B_5 Edu + \beta_6 Dep + \beta_7 RA(1) + \beta_8 RA(2) \quad (12)$$

Higher income is expected to be positively related to owning stocks (Vissing-Jorgensen, 2002). Women are generally more risk averse than men in financial decision-making (Byrnes, Miller & Schafer, 1999) and hence, most likely the coefficient for the variable *female* will be negative. For *Age*, Fagereng, Gottlieb and Guiso (2017) find that individuals rebalance their portfolio share as they get older and approach retirement. Slowly investors start to invest more indirectly and to refrain from investing in the stock market. This implies that *Age* should be negatively related to (direct) stock ownership. Similar to the models in paragraph 3.4 education is predicted to be positively related to *stockholder*, negatively related to holding stock directly and positively related to both the number of stocks owned and investing abroad. In addition, the proxies for willingness to take up risk most likely impact all dependent variables negatively.

4. Empirical results

4.1 Rank-dependent preferences

Table 8 shows the results of the two main regressions. Both regressions are estimated in order to test whether households realize that their way of investing is associated with higher levels of risk. Therefore, the main variables of interest in both regressions are the risk aversion factors. Both variables for risk aversion have alternately a positive and negative coefficient sign. Moreover, in approximately half of the models the variables are not statistically significant. Thus, it seems there is no clear and consistent relationship between the risk aversion factors and the level of diversification.

A negative sign for education would show that households are aware of the risk associated with holding stocks directly. The results show that this is only the case in more recent years, however this effect is not statistically significant for both dependent variables. Hence, education appears not to impact the fraction that households invest in directly held stocks. Regarding the number of dependents the coefficients have both a positive and negative sign. The magnitude of for example an increase of one unit for the number of dependents ranges from -0.6% to 8.4% on the fraction of direct equity relative to risky financial assets. These effects are economically significant taking into account the median direct equity allocations as reported in Table 5. However, the effects are not consistently statistically significant. To summarize this (regression) analysis, the results do not provide conclusive evidence on whether households are aware (or unaware) of the riskiness of holding stocks directly. Thus, this is not in line with hypothesis 2. In other words, we cannot reject the null hypothesis that households are unaware of the riskiness of under-diversification.

4.2 Preference for skewness

Panel A of Figure 3 shows and plots the estimated relationship between the households' Sharpe ratios and diversification measure D_3 using Equation 8. The graph shows a fairly stable, but slowly upward sloping trend. A higher value for the diversification measure (higher concentration and/or correlation of stocks) seems related to a higher Sharpe ratio. This contradicts the findings of Mitton and Vorkink (2007) who find a clear downward trending relationship. However, it should be noted that the (extremely) wide standard error bands indicate that the

relationship has to be interpreted with caution. Panel A provides by no means conclusive evidence about the relationship between the Sharpe ratio and the portfolio level of diversification. Replacing D_3 by the other diversification measures yields the same results.

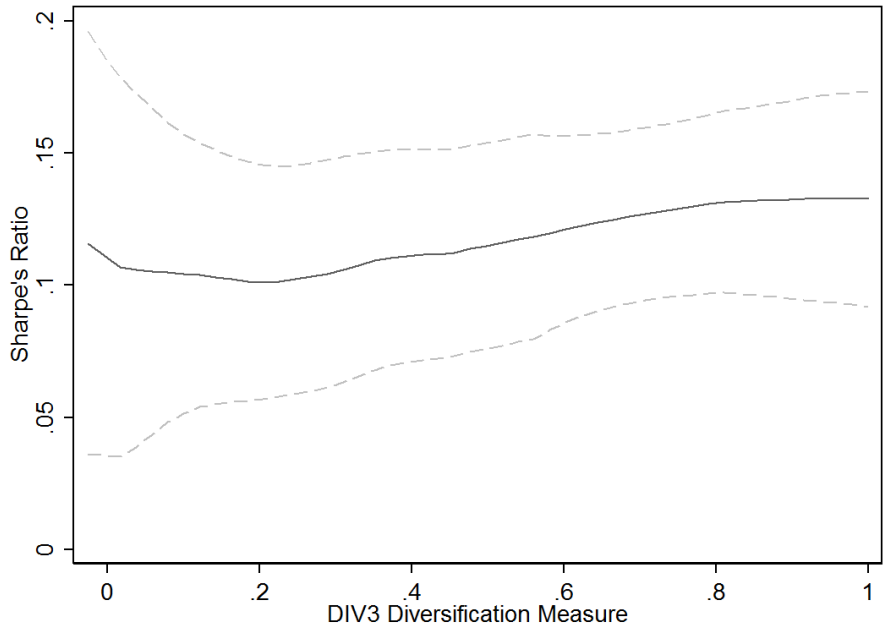
Panel B plots the estimated relationship between the Sharpe ratio and portfolio skewness of households. It seems that a negative non-linear relationship can be observed. The graph shows a downward sloping trend for skewness coefficients with a value higher than 0.5. High levels of skewness are associated with lower levels of the Sharpe ratio. This can also be framed as that investors get rewarded in skewness for the lack of mean-variance efficiency in their under-diversified portfolios. These portfolios are not efficient from a mean-variance perspective but might be efficient when additionally skewness is considered. Higher skewness comes at the expense of mean-variance efficiency, which lowers the Sharpe ratio, explaining the downward sloping trend. This does not fully explain why this seems to happen only after a certain threshold (a skewness coefficient of 0.5 in this case). It could be that particularly highly skewed investments (investment opportunities with an extreme upside potential) cannot be accommodated by a mean-variance (return and risk) framework.

The overall finding is in line with Mitton and Vorkink (2007). However, whereas Mitton and Vorkink (2007) claim a statistically significant difference between highly and negatively skewed household portfolios, I cannot conclude the same. The confidence interval is again extremely wide and requires to interpret the results with caution. Both plots in Panel A and Panel B are not in line with hypotheses 4A and 4B. Both the portfolio diversification measure as well as portfolio skewness are not evidently negatively related to mean-variance efficiency.

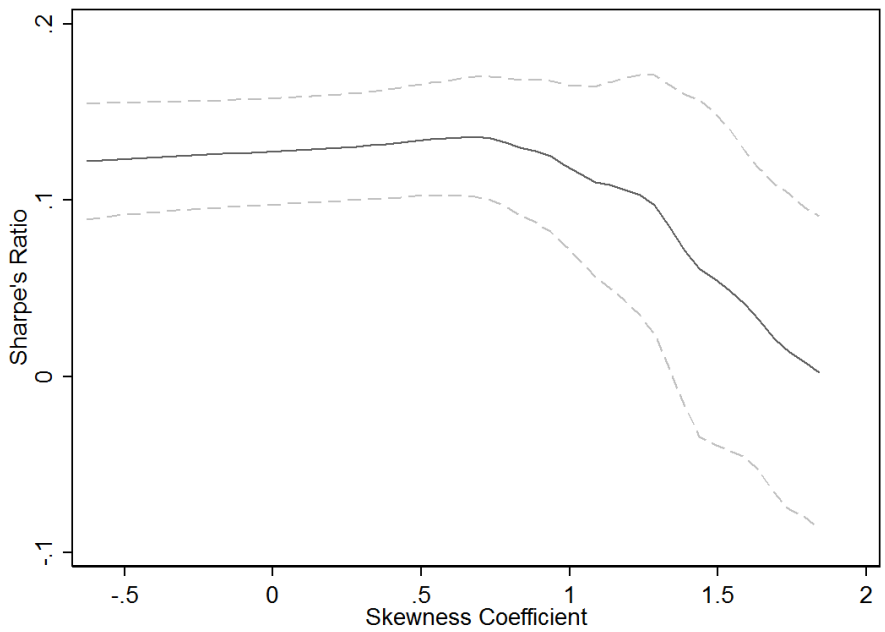
Table 8 OLS regression on the cross-sectional variation in direct equity relative to (risky) financial assets

	Log(FA)	Education	Number of dependents	RA(1)	RA(2)	*N*	Adj. R2
Panel A: 2004							
Total FA	0.0026	0.0362	0.0072	0.0062	-0.0404*	150	0.01
	(0.18)	(1.34)	(0.50)	(0.29)	(-1.96)		
Risky FA	-0.0214	0.0809**	0.0455**	0.0739**	0.0407	132	0.05
	(-0.89)	(2.00)	(2.04)	(2.14)	(1.25)		
Panel B: 2007							
Total FA	0.0162	0.0048	-0.0318*	-0.0513**	-0.0152	144	0.03
	(1.16)	(0.17)	(-1.85)	(-2.11)	(-0.63)		
Risky FA	-0.0619***	0.0162	-0.0058	-0.0688*	0.0446	125	0.06
	(-2.71)	(0.41)	(-0.23)	(-1.87)	(1.25)		
Panel C: 2010							
Total FA	0.0270*	0.0240	-0.0375**	-0.0619***	-0.0161	147	0.08
	(1.70)	(0.90)	(-2.12)	(-2.92)	(-0.65)		
Risky FA	-0.0513*	0.0417	-0.0271	-0.0505	0.0472	137	0.01
	(-1.88)	(0.89)	(-0.87)	(-1.38)	(1.09)		
Panel D: 2013							
Total FA	-0.0034	-0.0003	0.0129	-0.0403	-0.0554**	105	0.03
	(-0.22)	(-0.01)	(0.59)	(-1.48)	(-2.07)		
Risky FA	-0.0440	0.0326	0.0840**	-0.0394	0.0640	93	0.06
	(-1.55)	(0.62)	(2.25)	(-0.86)	(1.40)		
Panel E: 2016							
Total FA	-0.0007	-0.0172	-0.0075	-0.0246	-0.0208	146	-0.02
	(-0.05)	(-0.62)	(-0.42)	(-0.98)	(-0.86)		
Risky FA	-0.0881***	-0.0629	-0.0151	0.0122	0.0808**	129	0.09
	(-3.26)	(-1.35)	(-0.48)	(0.28)	(2.03)		

Note: *t*-statistics in parentheses, * p<.10, ** p<.05, *** p<.01. The dependent variables are fractions of directly owned equity relative to total financial assets and risky financial assets. The independent variables are as defined in the text. The sample is restricted to households owning direct equity.



Panel A



Panel B

Figure 3 Local polynomial estimation of the relationship between household's Sharpe ratio and the diversification measure D_3 (Panel A) and portfolio skewness coefficient (Panel B). Estimates are in solid lines, while the dashed lines show the 95% confidence interval.

4.3 Stock market participation

Table 11 to Table 14 in Appendix 4 show the results of additional analyses on the determinants for households being either a stockholder and/or direct owner of equity, the number of stocks owned and whether a household invests abroad. Table 11, Table 12 and Table 14 show the results of a probit regression which inherently changes the interpretation of the coefficients relative to the outputs of a standard OLS model. The interpretation is limited to assessing the sign of a coefficient and the statistical significance of a variable. The coefficient itself does not have a direct interpretation. One way to interpret the coefficients is to consider the marginal effect of a change in the independent variable. This is the change in y* for a unit change in x, where y* is the underlying latent propensity that the dummy variable equals 1.

Starting with the results of Table 11, a higher amount of financial assets, and to a lesser extent income, is associated with holding stocks. The variable *log(FA)* is in all cases positive and statistically significant. This confirms the observations based on Table 3 and hence, it confirms hypothesis 3. In addition to the coefficients, I report the marginal effects of the variables in Table 15 in Appendix 4 for a more intuitive interpretation of the effects of the independent variables in

this model. For the probit model with all independent variables included (model 8), the marginal effect of *log(FA)* is 0.0916. The interpretation of this number is that a difference of 1 in *log(FA)* corresponds to an increase of 0.0916 in y=1. In other words, the baseline probability changes by 0.0916 to a new level of expected probability. It should be noted that a difference of 1 in *log(FA)* is not simply an increase of €1 in financial assets.[7] The dummy variable *female* has a negative sign and is statistically significant at the 1% level for the models excluding the risks aversion factors. Including these factors takes away the explanatory power of the gender variable. A similar effects can be observed for the variable *edu.* Relative to the baseline of no or lower education, academic education has a positive and statistically significant on participation in the stock market. The statistical significance of this coefficients disappears after including the risk aversion factors. *Age* is positively related to the dependent variable *stockholder* and this effect is statistically significant for the group older than 65 in the full model. This would imply that the older an investor gets, the more inclined he/she is to participate in the stock market. The number of dependents in a household is negatively and statistically significantly related to *stockholder*. The marginal effect of a one unit increase in *Dep* (one more person in the household) decreases the likelihood of participation with 0.0185. Both risk aversion factors are negatively related to stock market participation and these effects are statistically significant at the 1% level.

Table 12 reports the results of the analysis of direct equity ownership. Overall, the results and relationships are very similar to the previous analysis with *stockholder* as a dependent variable. However, this analysis does not provide evidence that *age* is negatively related to owning equity directly. Life-cycle theory predicts that investors invest less in risky assets while they get older. *Edu* has a positive and predominantly statistically significant effect on owning equity directly. The variables that proxy for the willingness to take risk are the number of dependents in the household and the two variables for risk aversion. The variable *Dep* is both positively and negatively related to *direct_owner,* however this effect is all cases not statistically significant. Self-reported risk aversion is clearly negatively related to holding equity directly. Across all years this effect is statistically significant at the 1% level.

Table 13 reports the results of the OLS regression with the number of stocks owned as dependent variable. The interpretation of the logarithmically transformed variables *log(FA)* and

[7] To be more specific: an increase of 1 in *log(FA)* equals to multiplying *FA* by e (e=2.7183). This implies that the absolute change in *FA* is 2.7183, which simplifies to 2.7183**FA.* It percentage terms, this is a 172% change in financial assets.

log(income) is that a one unit increase in for example *log(FA)* will increase the number of stocks by the coefficient of *log(FA).* Please refer to footnote 7 for the interpretation of a one unit increase of a logarithmic independent variable. The value of financial assets has a clear positive and statistically significant effect on the number of stocks owned. Having more financial assets is associated with a higher number of stocks. A one-unit increase in *log(FA)* increases the amount of stocks owned by 0.12 to 0.14 (across the models). The coefficients of *log(income)* have positive signs as well, however this variable is not consistently statistically significant. The dummy variable *female* is negatively related to the amount of stocks owned and is statistically significant in the reduced models. Education is consistently positively related to the dependent variable and this effect is statistically significant. The same holds for variable *number of dependents*, although this relationship is negative. *Age* has a positive, but not statistically significant effect on the number of stocks owned. In line with the previous regression models, self-reported risk aversion is clearly negatively related to the number of stocks owned.

Lastly, Table 14 presents the results of a probit model that investigates the determinants of investing in stocks of foreign companies. The effect of financial assets on the decision to invest abroad is positive and statistically significant. Households that have more assets available to invest are more likely to invest in foreign stocks. The variable *Age* shows that investors above the age of 65 are less likely to invest abroad as compared to investors under 40. This effect is statistically significant. Additionally, the number of dependents in a household is negatively (and statistically significantly) related to *investing abroad.* All other variables do not have a statistically significant effect on the likelihood of owning stocks of foreign companies.

5. Conclusion

5.1 Discussion of the results

The analysis of financial assets reveals how Dutch households invest their assets and how their portfolios are constructed. Table 3 shows that stockholders have higher financial assets than non-stockholders and that the median values for financial assets do not change significantly over the years. Toward 2016 it seems that a lower proportion of households invests their assets in equity. Simultaneously, more and more of the equity value is held indirectly rather than directly. Another observation is that only a subset of the sample of households invests in equity, or in other words, stock market participation is generally low. Striking is that even households with substantial amounts of financial assets do not invest their money either directly or indirectly in equity. This finding is consistent with the finding of Polkovnichenko (2005) among American households and inconsistent with expected utility theory.

More in-depth analysis shows that the least wealthy groups in terms of financial assets mainly invest their money directly. This direct way of owning equity coincides with owning a very limited amount of stocks (in most cases only 1 or 2 stocks). Hence, the least wealthy households seem to be badly diversified implying that this (more vulnerable) group most likely bears a high level of portfolio risk. Groups with higher levels of financial assets own more stocks, however the amount of different stocks owned is still fairly low and the median value of stocks owned seems to decrease over time. To provide some insightful numbers, approximately 95% of the households own five or less stocks and the amount of total financial assets owned by this group is around 90%. Thus, a large amount of financial assets of the Dutch population are invested in limited-stock portfolios.

Several explanations exist that could explain why households hold undiversified portfolios of individual stocks. Firstly, I investigate whether rank-dependent preferences can explain the diversification level of investors. Polkovnichenko (2005) shows theoretically and empirically that from a rank-dependent framework owning only a few stocks makes sense. I use regression analysis to test whether this explanation is applicable for my data as well. In order to rule out that biases drive the under-diversification of households I analyze whether households are aware of the higher levels of risk that are associated with undiversified portfolios. The analysis shows that the level of education is not (statistically significantly) negatively related to the fraction of directly owned

equity relative to financial assets. This (most likely) indicates that households are unaware of the associated risk. It is hypothesized that negative coefficients for the variables risk aversion and the number of dependents in a household indicate that households consider the riskiness of individual stocks (in their portfolio). The results in Table 8 do not confirm this hypothesis. These variables are not consistently negatively related to the dependent variables and the effects are not in all cases statistically significant. Thus, it seems that rank-dependent preferences cannot explain the observed investment behavior of Dutch households.

Therefore, I turn to another explanation for under-diversification of households. Mitton and Vorkink (2007) provide evidence for the fact that investors have a preference for skewness and that a trade-off exists between capturing skewness and the level of diversification. This would imply that investors willingly hold a limited amount of stocks. Unlike Mitton and Vorkink (2007), I do not find that undiversified households have substantially higher levels of skewness. Moreover, both the portfolio diversification measures as well as portfolio skewness are not conclusively negatively related to mean-variance efficiency (as measured by the Sharpe ratio of a household's portfolio). Thus, a preference for skewness can also not explain the under-diversification of Dutch households.

In addition, I analyze the determinants of owning equity, directly owning equity, the number of stocks owned and owning foreign stocks. In almost all cases, I find that a higher amount of financial assets positively affects the ownership-related dependent variables. On the contrary, risk aversion is consistently negatively and statistically significantly related to owning equity (directly), the number of stocks owned and investing abroad. Education positively affects owning stocks (directly) which does show that the initial decision to invest in equity is influenced by the level of education. Apparently, based on rank-dependent preferences analysis, this effect remains limited to that initial stage and does not continue to impact *how* to invest in equity (in terms of allocation). Age is positively related to participation in the stock market and is negatively related to investing in foreign companies. The effects of the other explanatory variables alternate (both positive as well as negative signs, but inconsistent) and are often not statistically significant.

This study does not provide evidence for the fact that, universally, investors do not have rank-dependent or skewness preferences. It only shows that for this specific sample and time period these preferences do not seem to explain investment behavior. This does not deny the

existence of both types of preferences in empirical research. For that to claim more (empirical) research into these types of preferences should be done.

This study shows that households hold undiversified portfolios of individual stocks and that a substantial group of these direct investors invest in well-diversified funds simultaneously. Since both analyses on preferences cannot rule out that investors are simply unaware of the higher risk that is associated with the first component of this investment strategy, this gives ground to for example biased-based explanations for the under-diversification of households. This would not necessarily rules out a preference-based explanation. Both theories for the under-diversification of household portfolios could coexist.

The implications of the findings of this study could be interpreted as slightly disturbing. Non-participation in the stock market is widespread among the Dutch population at times that demand independence and individual responsibility for wealth accumulation. Moreover, under-diversification of household portfolios is commonplace, and it cannot be ruled out that households are unaware of the riskiness of this behavior. The more financially vulnerable groups in society are more likely to be negatively affected as they particularly display this behavior, which provides a strong(er) argument for policy-makers to act.

5.2 Limitations and future research

This study has several limitations that have both a major and minor impact on the results. First, as already mentioned, the DNB survey does not allow to determine households' frequency of trading. This is an issue for several analyses because I cannot exclude individuals that (un)willingly hold a specific stock or portfolio. The results of the main analyses on risk awareness are possibly too conservative.

Second, survey data in general is imperfect. Due to the fact that several, important, questions remain blank, are answered ambiguously and/or incorrectly, I had to clean, impute or even eliminate some observations. In particular, the analysis on the preferences for skewness suffers from imperfect data and associated constraints. In constructing household investment portfolios a substantial amount of observations are deleted from my sample. The quality of the information provided by households regarding their own stocks (name and quantity) was often poor. Moreover, other stocks could not be matched. It might be the case that the respondents I could not include in the additional portfolio analysis are inherently different than the group I am able to include. For example, investors that are able to provide detailed and correct information

about their portfolio might be more sophisticated investors than their counterparts. However, if this would be the case, I am left over with the more sophisticated investors in my sample. Since I do not find conclusive evidence for the effects of skewness, I am also not able to possibly overstate or overestimate an effect.

Another limitation related to portfolio construction is that for my analysis I assume that households hold the exact same portfolio (concentration and correlation) for a period of five years. As a result of imperfect survey data and too little detail in the survey questions, and taking into account my sample size, I deem this assumption the best alternative option. Consequently, I cannot perform a time-series regression analysis to investigate more formally whether the level of skewness is the result of a sophisticated decision-making process or plainly unintended.

In the preference for skewness part of this study I limit the scope of my research to direct equity ownership, hereby neglecting indirect ownership of equity through for example mutual funds. It might be the case that households (to a certain extent) compensate for their portfolio under-diversification by the means of investing in these type of funds. The claim I make throughout this paper is that a large body of households are undiversified investors. As a result of the scope of my analysis this might overstate the actual situation.

Future research should try to incorporate (even) more detailed information about household's portfolios. It would be beneficial for analyzing investment behavior to know more precisely the portfolio composition of households (frequency of trading, monthly updated portfolios, improved data quality). This would enable researchers to analyze data that reflect actual investment behavior more accurately. Additionally, future research should investigate empirically the role of biases or other explanations such as ambiguity aversion as has been done by Dimmock, Kouwenberg, Mitchell and Peijnenburg (2016) on household (equity) portfolios. Moreover, these studies should focus on the coexistence and/or interaction between types of explanations.

5.3 Concluding remarks

Based on data from the DHS survey I find that Dutch households have a low participation rate in the stock market and that they often hold undiversified and limited-amount-of-stock portfolios. This behavior can neither be fully explained by rank-dependent preferences nor by a preference for skewness. Future research should try to make use of more detailed household portfolio information to improve the quality of data and to enable the application of more

sophisticated analyses. These analyses should not be limited to preference-based theories of under-diversification, but should focus on alternative theories or the interaction between these theories.

6. References

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, *55*(2), 773-806.

Barberis, N., Huang, M., & Thaler, R. H. (2006). Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing. *American Economic Review*, *96*(4), 1069-1090.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*(1), 79-98.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2009). Fight or flight? Portfolio rebalancing by individual investors. *The Quarterly journal of economics*, *124*(1), 301-348.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, *115*(5), 707-747.

Cooper, I., & Kaplanis, E. (1994). Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. *The Review of Financial Studies*, *7*(1), 45-60.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American economic review*, 79-98.

Campbell, J. Y. (2006). Household finance. *The Journal of Finance*, *61*(4), 1553-1604.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics*, *119*(3), 559-577.

Dutch Banking Association (2017). Fact Sheet Loans, Savings and Investments. Retrieved from www.nvb.nl

Fagereng, A., Gottlieb, C., & Guiso, L. (2017). Asset Market Participation and Portfolio Choice over the Life- Cycle. *The Journal of Finance*, *72*(2), 705-750.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66* (Vol. 66). CRC Press.

Friedman, M., & Savage, L. J. (1948). The Utility Analysis of Choices Involving Risk. *Journal of Political Economy*, *56*(4), 279-304.

Friend, I., & Blume, M. E. (1975). The demand for risky assets. *The American Economic Review*, *65*(5), 900-922.

French, K. R., & Poterba, J. M. (1991). *Investor diversification and international equity markets (No. w3609).* National Bureau of Economic Research.

Gaudecker, H.M. von (2015). How Does Household Portfolio Diversification Vary with Financial Sophistication and Advice? *Journal of Finance* 70(2), 489-507.

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 433-463.

Gomes, F., & Michaelides, A. (2005). Optimal Life- Cycle Asset Allocation: Understanding the Empirical Evidence. *The Journal of Finance*, *60*(2), 869-904.

Graham, J. R., Harvey, C. R., & Huang, H. (2009). Investor competence, trading frequency, and home bias. *Management Science*, *55*(7), 1094-1106.

Guiso, L., & Sodini, P. (2012). Household finance: An emerging field.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *Journal of Finance*, *63*(6), 2557-2600.

Guiso, L., & Jappelli, T. (2006). Information acquisition and portfolio performance.

Haliassos, M. (2003). Stockholding: recent lessons from theory and computations. In *Stockholding in Europe* (pp. 30-49). Palgrave Macmillan UK.

Haliassos, M., & Bertaut, C. C. (1995). Why Do So Few Hold Stocks?. *Economic Journal*, *105*(432), 1110-29.

He, X. D., Kouwenberg, R., & Zhou, X. Y. (2017). Rank-Dependent Utility and Risk Taking in Complete Markets. *SIAM Journal on Financial Mathematics*, *8*(1), 214-239.

Hurd, M., Van Rooij, M., & Winter, J. (2011). Stock market expectations of Dutch households. *Journal of Applied Econometrics*, *26*(3), 416-436.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-292.

Kapteyn, A., & Teppa, F. (2011). Subjective measures of risk aversion, fixed costs, and portfolio choice. *Journal of Economic Psychology*, *32*(4), 564-580.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. *Advances in experimental social psychology*, *20*, 255-295.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77-91.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, 247-257.

Neumann, J. V., & Morgenstern, O. (1944). Theory of games and economic behavior.

Polkovnichenko, V. (2005). Household portfolio diversification: A case for rank-dependent preferences. *Review of Financial studies*, *18*(4), 1467-1502.

Rabin, M. (2000). Risk aversion and expected- utility theory: A calibration theorem. *Econometrica*, *68*(5), 1281-1292.

Roy, A. D. (1952). Safety First and the Holding of Assets. *Econometrica, 20*(3), 431-448.

Schroder Investment Barometer (2017).

Shefrin, H., & Statman, M. (2000). Behavioral portfolio theory. *Journal of financial and quantitative analysis*, *35*(02), 127-151.

Vissing-Jorgensen, A. (2002). *Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures* (No. w8884). National Bureau of Economic Research.

7. Appendix

Appendix 1 DNB household survey

The six modules of the DNB household survey are the following:

1. General Information on the Household
2. Household and Work
3. Accommodation and Mortgages
4. Health and Income
5. Assets and Liabilities
6. Economic and Psychological Concepts

This study mainly makes use of the questions asked in module 5. For additional analyses, module 1 and 6 are consulted.

Appendix 2 Demographics of households

Table 9 Descriptive statistics of households per group and year

DHS year and cohort	Age *(% of population)*			Gender *(% of population)*		Education *(% of population)*			Number of dependents (household size)
	<40	41-64	65+	Male	Female	No/elementary/secondary	Higher vocational	Academic	
2004									
1	35.6	48.1	16.4	64.2	35.8	43.5	46.4	10.1	2.4
2	21.9	53.9	24.2	77.2	22.8	38.4	46.0	15.6	2.3
3	8.3	54.8	36.9	82.1	17.9	34.5	34.5	31.0	2.1
4	20.0	20.0	60.0	100.0	0.0	60.0	20.0	20.0	1.6
All	26.7	51.2	22.1	72.0	28.0	40.5	45.0	14.5	2.3
2007									
1	32.9	45.5	18.6	64.0	36.0	40.9	51.1	8.0	2.4
2	17.4	52.8	29.8	72.5	27.5	39.6	43.0	17.4	2.2
3	6.1	57.1	36.7	80.6	19.4	30.9	39.2	29.9	2.1
4	16.7	33.3	50.0	100.0	0.0	33.3	41.7	25.0	2.2
All	23.0	51.1	25.9	69.9	30.1	39.3	46.2	14.6	2.3
2010									
1	20.7	56.1	23.2	61.6	38.4	46.5	45.2	8.4	2.3
2	11.4	52.6	36.0	71.4	28.6	40.3	41.4	18.4	2.1
3	4.2	52.1	43.8	85.4	14.6	25.0	45.8	29.2	2.1
4	0	30.0	70.0	90.0	10.0	10.0	20.0	70.0	1.8
All	13.7	56.5	32.8	69.7	30.3	40.5	42.9	16.6	2.2
2013									
1	25.1	45.5	29.4	64.0	36.0	41.8	48.4	9.8	2.2
2	13.4	40.7	45.8	75.0	25.0	35.0	44.4	20.6	2.1
3	3.5	41.9	54.7	83.7	16.3	27.9	39.5	32.6	2.1
4	0.0	80.0	20.0	80.0	20.0	0.0	60.0	40.0	1.6
All	17.3	43.1	39.6	71.3	28.8	37.0	45.7	17.3	2.1
2016									
1	28.5	43.4	28.1	58.7	41.3	36.0	51.5	12.5	2.4
2	17.5	38.5	44.0	67.4	32.6	30.6	49.7	19.7	2.2
3	2.9	42.7	54.4	82.5	17.5	25.2	41.8	33.0	2.1
4	0.0	33.3	66.7	66.7	33.3	16.7	50.0	33.3	1.7
All	21.3	41.0	37.7	64.6	35.4	32.6	49.9	17.5	2.3

Note: Reported values are percentages of total population. The values reported for *Number of Dependents* are means.

Appendix 3 Risk aversion proxy

The risk aversion variables are based on six questions related to taking risks and two questions related to (precautionary) savings. The following six risk-related statements have to be answered on a seven point scale ranging from totally disagree to totally agree:

1. I think it is more important to have safe investments and guaranteed returns, than to take a risk to have a chance to get the highest possible returns.
2. I do not invest in shares, because I find this too risky.
3. If I think an investment will be profitable, I am prepared to borrow money to make this investment. (R)
4. I want to be certain that my investments are safe.
5. If I want to improve my financial position, I should take financial risks. (R)
6. I am prepared to take the risk to lose money, when there is also a chance to gain money. (R)

The (R) indicates that the answer-scale has been reversed in order to account for the reversed phrasing of the questions.
The two (precautionary) savings questions have to be answered on a seven point scale ranging from very unimportant to very important:

7. to have some savings to cover unexpected expenses
8. to have enough money in your bank account to be sure you will be able to meet your financial liabilities

Table 10 Factor analysis of risk aversion and precautionary motives using PCA with varimax rotation

Factors	Factor 1 (Riskav1)	Factor 2 (Riskav2)	Factor 3 (Precaution)	Uniqueness
Risk aversion				
1	0.0642	**0.8224**	0.0998	0.3096
2	**0.6189**	0.3426	0.0034	0.4995
3	**0.5583**	-0.0761	0.1627	0.6560
4	0.0204	**0.8441**	0.0778	0.2810
5	**0.7921**	-0.0952	-0.0817	0.3568
6	**0.8532**	0.1093	-0.0269	0.2594
Precautionary motives				
7	-0.0537	0.1263	**0.8264**	0.2982
8	0.0036	0.0468	**0.8448**	0.2841

Note: Data is from DHS wave 2016. The analysis yields similar results for other waves. The factor numbers refer to the statements above. The largest factor loading for each statement is in bold.

Appendix 4 Determinants of stock market participation.

Table 11 to Table 15 show the results of additional analyses on the determinants of stock market participation.

Table 11 Probit regression stockholder[8]

	(1) Stockholder	(2) Stockholder	(3) Stockholder	(4) Stockholder	(5) Stockholder	(6) Stockholder	(7) Stockholder	(8) Stockholder
Log(FA)	0.4745***	0.4689***	0.4638***	0.4610***	0.4521***	0.4507***	0.4457***	0.4750***
	(13.48)	(12.44)	(12.40)	(12.21)	(11.60)	(11.56)	(10.95)	(11.21)
Log(income)		0.3266***	0.2799***	0.2237***	0.2268***	0.2486***	0.1704**	0.1480*
		(4.26)	(3.66)	(2.89)	(2.93)	(3.15)	(2.03)	(1.81)
Female			-0.2806***	-0.3127***	-0.2962***	-0.3440***	-0.0689	-0.0696
			(-2.65)	(-2.93)	(-2.69)	(-3.12)	(-0.57)	(-0.56)
Edu (Base group: No/elementary/secondary edu)								
Higher vocational education				0.1179	0.1385	0.1478	0.1303	0.1340
				(1.08)	(1.21)	(1.28)	(1.03)	(1.05)
Academic education				0.2904**	0.3264**	0.3175**	0.1742	0.1579
				(2.06)	(2.22)	(2.15)	(1.11)	(0.99)
Age (Base group: <40)								
41-64					0.1159	0.0832	0.1724	0.1656
					(0.84)	(0.60)	(1.18)	(1.13)
65>					0.1289	0.0110	0.2895*	0.3096*
					(0.87)	(0.07)	(1.78)	(1.88)
Number of dependents						-0.1254***	-0.0988**	-0.0959**
						(-2.80)	(-2.17)	(-2.10)
RA(1)							-0.6393***	-0.6051***
							(-11.37)	(-10.75)
RA(2)								-0.3277***
								(-6.05)
N	1380	1225	1225	1225	1225	1225	1178	1178
Pseudo R^2	0.22	0.23	0.24	0.24	0.24	0.25	0.34	0.37

Note: *z*-statistics in parentheses, * p<.10, ** p<.05, *** p<.01. The dependent variable is a dummy variables indicating 1 if a household participates in the stock market by investing either directly or indirectly and 0 if otherwise. The independent variables are as defined in the text.

[8] This table as well table 11 reports the Pseudo R^2 of the estimated regressions. Unlike the standard R^2, which is a measure of explained variation or goodness of fit, the Pseudo R^2 cannot be interpreted in this way. In general, the Pseudo R^2 can be used to evaluate different specifications of the same model. Therefore, it should be seen as a measure of relative fit. Measures with a value between 0.2 and 0.4 indicate a good model fit McFadden (1978). Thus, the models in Table 9 have a good model fit.

Table 12 Probit regression direct owner

	(1) direct_owner	(2) direct_owner	(3) direct_owner	(4) direct_owner	(5) direct_owner	(6) direct_owner	(7) direct_owner	(8) direct_owner
Log(FA)	0.3228***	0.3111***	0.3054***	0.3017***	0.3047***	0.3036***	0.2842***	0.2999***
	(9.17)	(8.12)	(8.07)	(7.88)	(7.43)	(7.39)	(6.74)	(7.05)
Log(income)		0.3810***	0.3274***	0.2599***	0.2638***	0.2724***	0.1952*	0.1738
		(3.88)	(3.39)	(2.67)	(2.67)	(2.73)	(1.87)	(1.63)
Female			-0.3011**	-0.3361***	-0.3404**	-0.3535***	-0.0683	-0.0500
			(-2.40)	(-2.63)	(-2.53)	(-2.61)	(-0.46)	(-0.33)
Edu (Base group: No/elementary/secondary edu)								
Higher vocational education				0.1819	0.1773	0.1797	0.1388	0.1294
				(1.43)	(1.32)	(1.33)	(0.94)	(0.87)
Academic education				0.3301**	0.3183*	0.3122*	0.1661	0.1460
				(2.07)	(1.93)	(1.89)	(0.94)	(0.81)
Age (Base group: <40)								
41-64					-0.0726	-0.0858	-0.0386	-0.0594
					(-0.48)	(-0.56)	(-0.24)	(-0.36)
65>					-0.0460	-0.0883	0.1092	0.1274
					(-0.27)	(-0.51)	(0.59)	(0.68)
Number of dependents						-0.0430	-0.0056	0.0019
						(-0.89)	(-0.11)	(0.04)
RA(1)							-0.6200***	-0.5916***
							(-9.90)	(-9.61)
RA(2)								-0.2698***
								(-4.53)
N	1380	1225	1225	1225	1225	1225	1178	1178
Pseudo R^2	0.12	0.15	0.16	0.16	0.16	0.16	0.27	0.29

Note: *z*-statistics in parentheses, * p<.10, ** p<.05, *** p<.01. The dependent variable is a dummy variables indicating 1 if a household participates in the stock market by investing either directly or indirectly and 0 if otherwise. The independent variables are as defined in the text.

Table 13 Cross-sectional analysis of number of stocks owned by a household

	(1) Number of stocks	(2) Number of stocks	(3) Number of stocks	(4) Number of stocks	(5) Number of stocks	(6) Number of stocks	(7) Number of stocks	(8) Number of stocks
Log(FA)	0.1424***	0.1478***	0.1438***	0.1409***	0.1336***	0.1317***	0.1162***	0.1292***
	(8.43)	(7.25)	(7.05)	(6.91)	(6.28)	(6.20)	(5.31)	(5.91)
Log(income)		0.1390***	0.1106**	0.0781	0.0859	0.0925*	0.0809	0.0899
		(2.74)	(2.12)	(1.45)	(1.59)	(1.71)	(1.42)	(1.59)
Female			-0.1807**	-0.2046**	-0.1854**	-0.2114***	-0.0990	-0.0890
			(-2.29)	(-2.58)	(-2.29)	(-2.59)	(-1.17)	(-1.06)
Education				0.1331**	0.1542***	0.1553***	0.0968*	0.0954*
				(2.47)	(2.73)	(2.75)	(1.66)	(1.65)
Age					0.0645	0.0263	0.0757	0.0817
					(1.21)	(0.47)	(1.32)	(1.44)
Number of dependents						-0.0752**	-0.0741**	-0.0732**
						(-2.37)	(-2.28)	(-2.28)
RA(1)							-0.2890***	-0.2787***
							(-7.48)	(-7.26)
RA(2)								-0.1766***
								(-4.66)
N	1380	1225	1225	1225	1225	1225	1178	1178
adj. R^2	0.05	0.06	0.06	0.06	0.06	0.07	0.11	0.13

Note: *t*-statistics in parentheses, * p<.10, ** p<.05, *** p<.01. The dependent variable is a continuous variable that equals the amount of stocks owned by a household. The independent variables are as defined in the text.

Table 14 Probit regression investing abroad

	(1) Investing abroad	(2) Investing abroad	(3) Investing abroad	(4) Investing abroad	(5) Investing abroad	(6) Investing abroad	(7) Investing abroad	(8) Investing abroad
Log(FA)	0.1644*	0.1887**	0.1856*	0.1865*	0.2436**	0.2492**	0.2444**	0.2487**
	(1.83)	(1.99)	(1.96)	(1.95)	(2.25)	(2.20)	(2.17)	(2.22)
Log(income)		-0.0811	-0.1081	-0.0870	-0.0946	-0.0230	0.0110	0.0047
		(-0.40)	(-0.52)	(-0.42)	(-0.42)	(-0.10)	(0.05)	(0.02)
Female			-0.1574	-0.1503	-0.1726	-0.1665	-0.1707	-0.1627
			(-0.48)	(-0.46)	(-0.53)	(-0.51)	(-0.51)	(-0.48)
Edu (Base group: No/elementary/secondary edu)								
Higher vocational education				-0.0838	-0.1478	-0.1951	-0.1861	-0.1910
				(-0.28)	(-0.49)	(-0.65)	(-0.62)	(-0.64)
Academic education				-0.1067	-0.2136	-0.3733	-0.3925	-0.4027
				(-0.32)	(-0.61)	(-1.02)	(-1.07)	(-1.10)
Age (Base group: <40)								
41-64					-0.3759	-0.5540	-0.5483	-0.5805
					(-1.01)	(-1.41)	(-1.39)	(-1.49)
65>					-0.4922	-0.8017*	-0.8152*	-0.8244**
					(-1.30)	(-1.91)	(-1.93)	(-1.97)
Number of dependents						-0.2632**	-0.2791**	-0.2773**
						(-1.99)	(-2.04)	(-2.03)
RA(1)							0.0158	0.0212
							(0.10)	(0.14)
RA(2)								-0.0484
								(-0.30)
N	148	138	138	138	138	138	137	137
Pseudo R^2	0.02	0.03	0.03	0.03	0.04	0.06	0.07	0.07

Note: z-statistics in parentheses, * p<.10, ** p<.05, *** p<.01. The dependent variable is a dummy variable indicating 1 if a household invests in stocks abroad and 0 if otherwise. The independent variables are as defined in the text

Table 15 Marginal effects probit regressions

Variable	Stockholder	Direct_owner
	dy/dx	*dy/dx*
Log(FA)	0.0916	0.0309
Log(Income)	0.0285	0.0179
Female	-0.0134	-0.0051
Education		
Higher vocational	0.0250	0.0128
Academic	0.0299	0.0146
Age		
41-64	0.0281	-0.0056
65>	0.0576	0.0140
Number of dependents	-0.0185	0.0002
RA(1)	-0.1167	-0.0610
RA(2)	-0.0632	-0.0278

Note: This table reports the marginal effects of a one unit change in the independent variables on either *stockholder* or *direct_owner*. The results correspond to the data from the 2016 wave. Dy/dx for factor variables *Education* and *Age* is the discrete change from the base level.

Table 16 Pearson correlation for independent variables

	Log(FA)	Log(income)	Female	Education	Age	Number of dependents	RA(1)	RA(2)
Log(FA)	1							
Log(income)	0.2623	1						
Female	-0.1477	-0.2651	1					
Education	0.1136	0.2418	0.043	1				
Age	0.2106	-0.07	-0.2207	-0.2941	1			
Number of dependents	-0.0685	0.1041	-0.0625	0.1154	-0.3055	1		
RA(1)	-0.1659	-0.1733	0.1812	-0.2014	0.1095	-0.0449	1	
RA(2)	0.1454	0.0484	0.0239	0.0251	0.0218	0.0084	0	1

Note: The results correspond to data from the 2016 wave. Riskav1 and riskav2 are orthogonal to one another by construction, therefore the variables are uncorrelated.

Appendix 5 Variables

Table 17 Description of the variables in this paper

Variable	Description
Age	Age of the respondent
Female	Dummy variable indicating whether respondent is female
Edu	Highest completed education: 1) No/elementary/secondary education 2) Higher vocational education 3) Academic education
Dep	Number of dependents in a household (=household size)
Income	Total gross annual income of a household
Log(income)	Logarithmic distribution of total gross annual income of a household
Financial assets (FA)	Total financial assets of a household
Log(FA)	Logarithmic distribution of total financial asset of a household
Financial advice	Type of financial advice: 1) Professional 2) Family/friends 3) Own judgment/no advice)
Riskav1 (RA1)	First measure of risk aversion (extract from factor analysis)
Riskav2 (RA2)	Second measure of risk aversion (extract from factor analysis)
Z_risk_agg	Alternative measure of risk aversion
Stockholder	Dummy = 1 if household holds equity
Direct_owner	Dummy = 1 if household invests directly in the stock market
Indirect_owner	Dummy = 1 if household invest indirectly in the stock market (mutual fund)
DIRFA	Direct equity investment as a fraction of total financial assets
DIRRIS	Direct equity investment as a fraction of total risky investments
Number_stocks	Amount of companies from which a household has stocks
Investing_abroad	Dummy = 1 if household invests in foreign company
D_1 (DIV1)	Diversification measure based on the amount stocks owned by a household
D_2 (DIV2)	Diversification measure based on a Herfindahl index
D_3 (DIV3)	Diversification measure based on a Herfindahl index and correlation between stocks
Skewness coefficient	Coefficient measuring the level of skewness captured by portfolio returns
Sharpe ratio	Measure for calculating risk-adjusted return based on the return of a portfolio, the risk-free rate and the standard deviation of a portfolio's return.