EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

University of Tübingen
Working Papers in
Economics and Finance

No. 74

# The Portfolio Structure of German Households: A Multinomial Fractional Response Approach with unobserved Heterogeneity

by

Gideon Becker

Faculty of Economics and Social Sciences
www.wiwi.uni-tuebingen.de

# The Portfolio Structure of German Households: A Multinomial Fractional Response Approach with unobserved Heterogeneity *

Gideon Becker**

This draft: September 1, 2014

**Abstract**

What determines the risk structure of financial portfolios of German households? In this paper we estimate the determinants of the share of financial wealth invested in three broad risk classes. We employ a new econometric approach - the so called fractional multinomial logit model - which allows for joint estimation of shares while accounting for their fractional nature. We extend the model to allow for unobserved heterogeneity across households via maximum simulated likelihood. We find that self-assessed appetite for risk as well as the level of wealth have strong positive effects on the riskiness of the average household's portfolio. These findings largely stay true even after we control for the potential confounding effects of unobserved differences across households via correlated random effects.

*Key words:* household finance, portfolio composition, non-linear panel data model, fractional response model, unobserved heterogeneity

*JEL:* C15, C33, C35, C51, C58, D14, G11

*The research for this paper was supported by the German Science Foundation (DFG-project BI 767/2-1). The author would like to thank Martin Biewen, Joachim Grammig, Jeffrey Wooldridge and seminar participants at the Universities of Tuebingen and Hohenheim, the summer school "New Developments in Panel Data Methods" at the University of Mainz and the young academics workshop at the German Statistical Week 2013 for helpful comments and suggestions. All errors remain my own.

**University of Tübingen. Contact: University of Tübingen, Department of Economics and Social Sciences, Mohlstr. 36, D-72074 Tübingen, Germany. Phone: 07071-29-75438 E-mail: `gideon.becker@uni-tuebingen.de`

Contents

## List of Figures



## List of Tables

# 1 Introduction

In this paper we look into the composition of the financial portfolios of German households. How private households invest their wealth is an important research area as it determines the financial well-being of individual households as well as the performance of the overall economy. This became evident during the financial crisis of 2007 which was caused, among other things, by mistaken investments by parts of the US population. In addition, the financial portfolio of the average household has become more and more complex in recent years partly due to the need to complement waning public pension systems in many industrialized countries. It is, thus, crucial to investigate the driving forces behind these decision processes. Germany as the leading European economy is an interesting case as it exhibits one of the highest saving rates of rich developed countries. At the same time German households shun high-risk, high-return, investments such as stocks and instead opt for more conservative investment strategies (see Börsch-Supan and Essig, 2002; Börsch-Supan and Eymann, 2002).

The theoretical finance literature mainly makes statements about the share of a household's portfolio allocated to risky assets. The rest of the portfolio is thought of as being ”safe” without further distinction (Gollier, 2002). However, Carroll (2002) notes, that this division is not easily applicable to empirical research as most financial assets are neither completely safe nor clearly risky. Consequently, most empirical studies only focus on the share of wealth invested in risky assets, usually equities.[1] By contrast, we analyze the composition of a household's financial wealth in a joint framework. Specifically, we focus on the risk structure of a household's portfolio by dividing financial wealth into three risk classes (”clearly safe”, ”fairly safe” and ”clearly risky”) similar to other empirical papers like the ones in Guiso et al. (2002). In this fashion, one gets a better idea about the overall structure of household portfolios.

An appropriate econometric model for this context has to take into account the bounded nature of asset shares which lie between zero and one. Several models used in the literature (such as linear regression or Heckman selection models) are not suited for this situation. Here we follow the approach of Papke and Wooldridge (1996) who show that one can model the

[1] A good overview of several empirical papers is given in Table 1 in Cardak and Wilkins (2009).

non-linear conditional expectation of a fractional dependent variable via a non-linear function in the spirit of a binary response model. This approach has been extended to the chase of multiple shares by Mullahy (2011) and Murteira and Ramalho (2013) who also account for the fact that shares have to sum up to one in such a framework. We use their extension in order to model the shares of the three aforementioned risk classes jointly. Furthermore, we adapt their model to a panel data in the spirit of Train (2003) to control for unobserved heterogeneity across households.

Our analysis contributes to the field in several ways: We examine household portfolios for Germany which despite its importance has not been studied as thoroughly as other countries like the United States. This relatively low level of research activity is mainly due to a lack of appropriate data. Here we employ the SAVE survey - a rich micro survey on saving and investment decisions of German households which is still underexploited in our opinion. Furthermore, we do not only analyze the share of financial wealth invested in stocks as many previous studies but also the proportions held in savings accounts and "fairly safe" assets. In this way we obtain a more comprehensive picture of the portfolio structure and its determinants. On the methodological side we contribute by modeling shares via fractional response models. This approach is arguably more appropriate than other often used models in this context. What is more, by modeling shares in a joint framework we incorporate the interdependences between asset classes compared to a situation where one investigates each share separately. Finally, we extend multivariate fractional response models to the panel data case. Specifically we show how such data can be used to control for unobserved time-invariant household characteristics which might be related to covariates and thus bias the regression results.

The rest of the paper is organized as follows: Section 2 gives an overview of the current state of empirical research on the portfolio composition of private households. In Section 3 we motivate the econometric models which are employed in the subsequent estimations. Description of the data-set and summary statistics for our sample are presented in Section 4 before we report the empirical results in Section 5. Section 6 sums up the analysis and proposes potential future work.

## 2 Literature Review

Our research belongs to the new field of research known as household finance. This research area is mainly concerned with the portfolio choice and asset allocation of private households. A comprehensive introduction to the field is given by Campbell (2006). More extensive overviews of the growing body of empirical and theoretical studies on this topic can be found in Guiso et al. (2002) and Guiso and Sodini (2013). Here we review some of the main findings in the household finance literature. To explain the portfolio composition, empirical analyses typically control for demographic factors (i.e. age and education) and the financial resources of a household (i.e. income and wealth). More recently studies have focused on the effects of behavioral factors (i.e. risk-aversion and preferences) as well as other risk factors (i.e. health risk).

In the following we present results from the previous literature on some of the key variables for our analysis. A very comprehensive overview of the implications of theoretical household finance models on asset allocation can be found in Gollier (2002). One of his main conclusions is that higher wealth levels should be associated with more risky investment behavior due to declining relative risk aversion over wealth. In a similar fashion, King and Leape (1998) conclude that risky financial assets can be seen as a type of luxury good with high wealth-elasticities. Another explanation for a positive wealth effect is given by Cocco (2005) who argues that higher levels of wealth thwart the deterrent effect of fixed participation costs. In general, empirical research on portfolio choice strongly supports this alleged positive relationship between wealth and risky asset share (see Guiso et al., 2002; Campbell, 2006; Wachter and Yogo, 2010). Looking at German households in the 1990's Börsch-Supan and Eymann (2002) report positive wealth effects for the share of wealth invested in risky assets. Carroll (2002) finds the same pattern by analyzing the portfolio composition of rich households in the United States. He suggests that this relationship might be explained by capital market imperfections or bequest motives.

Many studies (e.g. Carroll, 2002; King and Leape, 1998) also find a positive effect of household income on the share invested in risky assets. One might argue that the mechanism of action in this case is that a higher monthly income provides a better cushion against losses realized on the risky part of one's portfolio. However, the relationship is not as clear as for wealth as there are studies that do not find a significant effect of income (Cardak and Wilkins, 2009) or do not include it in the regression to begin with (Börsch-Supan and Eymann, 2002).

Another important determinant of the risk structure of a portfolio is the tolerance toward financial risk. It should be self-evident that households with more risk-tolerant members will hold riskier portfolios compared to otherwise comparable households that are composed of more risk averse individuals. Thus, risk preference is an important component of any theoretical model for portfolio choice (see Gollier, 2002). This aspect has been scrutinized and confirmed by several studies (Campbell, 2006; Guiso and Sodini, 2013) in recent years. Guiso and Paiella (2008) give a detailed account on this aspect and conclude, among other things, that more risk averse actors choose outcomes that expose them to fewer risk.

When it comes to the effect of investor age on the portfolio structure the prevailing view is that older investors shy away from riskier and less liquid investments due to their shorter time horizon - see Gollier (2002) for a theoretical explanation and Campbell (2006) and Guiso and Sodini (2013) for a general overview. However, King and Leape (1998) argue that older investors might hold riskier and more complex portfolios because they were able to gather more investment experience over the years. Therefore, they are better able to asses information regarding the risk-return tradeoff of an investment. Additionally, both Ameriks and Zeldes (2004) and Wachter and Yogo (2010) look into this issue in detail and do not find evidence for a decrease of the share of financial wealth invested in equities. Thus, the effect of age on an investor's portfolio is not as clear as one might think. As Campbell (2006) notes, it is impossible to disentangle age, time and cohort effects. Usually one excludes cohort effects as an identifying assumption (Heaton and Lucas, 2000a).

Education is generally thought to be associated with more risky portfolios. King and Leape (1998) reason that the cost of obtaining and understanding information regarding assets can be expected to be lower for people with higher levels of education. Campbell (2006) argues in the same fashion that more educated households can process information more easily and thus avoid investment mistakes.

One recurring finding in the empirical literature is that men exhibit riskier investment strategies compared to women (Hinz et al., 1997; Halko et al., 2012). Felton et al. (2003) note that this gender gap could be due to higher levels of optimism for men. Optimism towards the economic future might influence an investor's behavior in two ways: if an investor expects a positive economic development, it is rational to participate in this upswing by investing in equities. On the other hand, self-assessed positive expectations could be a sign for a generally higher level of optimism in that agent as suggested by Puri and Robinson (2007).

Several studies (Guiso et al., 1996; Cardak and Wilkins, 2009; Campbell, 2006; Guiso and Paiella, 2008) find that households which cannot easily participate in the credit market hold less risky portfolios. This is in line with Gollier (2002) who predicts that liquidity constrains will likely decrease the share of risky assets as it inhibits consumption smoothing in the face of negative return shocks.

Recently, the literature has also focused on background risk such as health risk which is expected to move the portfolio towards more safe assets as it increases the overall risk exposure of a household (see Heaton and Lucas, 2000b). Rosen and Wu (2004) look into the effect of poor health on the portfolio allocation of American households. They find that having poor health is associated with more conservative portfolios.

## 3 Econometric Models

In this section we examine modeling strategies to jointly estimate the conditional means of a set of asset shares. A suitable model needs to take into account the intrinsic properties of such multivariate fractional data - for instance, the bounded nature of each individual share. In addition, we want to apply such a model to a panel data framework. This allows one to control for unobservable time-constant household characteristics that otherwise might bias the regression results.

We start by reviewing fractional models for the univariate cross-sectional case as introduced by Papke and Wooldridge (1996) and expanded to panel data by Papke and Wooldridge (2008). Next, we show how this approach can be extended to the multivariate case, as suggested by

Mullahy (2011) and Murteira and Ramalho (2013). Finally, we combine these two extensions of the original model to allow for the estimation of mean shares for several assets jointly in a panel data context via simulation methods as described in Haan and Uhlendorff (2006).

### 3.1 Univariate Fractional Response Models

In economics and finance the variable of interest is often a proportion or fraction, meaning it is bounded between 0 and 1 and can take on any value in between: $0 \leq y_i \leq 1$ where $i$ is the index for agents such as persons, households or firms. Examples of dependent variables with fractional nature in financial context are, for instance, the share of portfolio invest in stocks or the debt to asset ratio - see Cook et al. (2008) for an overview of several applications. In the context of household finance the share of a household's wealth invested in risky assets is the most studied quantity that fits this description.

Usually one is interested in the effect of a vector of explanatory variables $\boldsymbol{x}_i$ on the conditional mean of the fractional response variable. A popular estimation strategy is to employ OLS and estimate the conditional mean as a linear combination of the covariates:

$$E[y_i|\boldsymbol{x}_i] = \boldsymbol{x}_i\boldsymbol{\beta} \tag{1}$$

The advantages of this estimation strategy are its simplicity and the fact that the parameter vector $\boldsymbol{\beta}$ can be readily interpreted as marginal effects. However, the linear model does not take into account the non-linearity of the conditional expectation due to the bounded nature of the dependent variable. Thus, the coefficients obtained in this fashion can only be seen as linear approximation of the true marginal effects. Moreover, predicted values obtained from this method cannot be expected to lie within the [0,1] boundaries. Papke and Wooldridge (1996) note that this problem is analogous to the employment of a linear probability model for binary dependent variables.

One possible approach to account for the bounded nature of $y_i$ is to employ a parametric model for the density of $y_i$ conditional on $\boldsymbol{x}_i$. Popular choices for this variant are the beta regression model or the inflated beta regression.[2] However, as Papke and Wooldridge (1996) point out, this approach is often sensitive to misspecifications of the distributional assumptions and thus likely to yield inconsistent estimations.

Other models, applied in the empirical household finance literature, are used to accommodate the nonlinear nature of the conditional mean but often entail their own disadvantages[3]: A censored regression approach, as used by Wachter and Yogo (2010), is unlikely to be appropriate for shares as fractional data is not censored at the boundaries, but rather defined over this range (see Cook et al., 2008). Heckman selection models are likewise not suitable as they are intended for situations where one observes the dependent variable only conditionally on the outcome of a selection process. The double-truncated Tobit model is conceptually more appropriate as it is meant to handle corner solutions. However, Stavrunova and Yerokhin (2012) caution against this approach because of model-sensitivity. Miniaci and Weber (2002) give a very comprehensive overview of issues encountered in empirical studies in household finance as well as a survey of appropriate microeconometric models in this context.

Due to these potential shortcomings, Papke and Wooldridge (1996) propose another estimation strategy which has become increasingly popular in recent years due to its computational simplicity and intuitive appeal. They argue that a straightforward way to impose the necessary constraints on the conditional mean is to model it via a nonlinear link function: $0 < G(\cdot) < 1$. In this fashion, the conditional expectation as well as predicted values are ensured to lie between the boundary values:

$$E[y_i|\boldsymbol{x}_i] = G(\boldsymbol{x}_i\boldsymbol{\beta}) \in (0, 1) \tag{2}$$

The authors note that $G(\cdot)$ will often be a cumulative density function (CDF) as it naturally fulfills the requirement but in principle any type of function can be used. Analogous to the binary case, the link function will usually be given by either the standard normal CDF $G(\cdot) = \Phi(\cdot)$

[2] Both Ramalho et al. (2011) and Cook et al. (2008) provide an excellent overview of different models used to estimate the conditional mean of a fractional variable in this fashion.

[3] Here we give a non-exhaustive list of models employed in the household finance literature to estimate the risky asset share: (i) OLS: Heaton and Lucas (2000a); (ii) Tobit: Cardak and Wilkins (2009), Poterba and Samwick (2003) and Rosen and Wu (2004); (iii) Heckman: Bertaut and Starr-McCluer (2002) and Börsch-Supan and Eymann (2002).

or the logistic function $G(\cdot) = \Lambda(\cdot) = exp(\cdot)/[1 + exp(\cdot)]$. These specifications respond to the fractional probit model and the fractional logit model (Flogit), respectively. Similar to Papke and Wooldridge (1996) we will focus on the Flogit specification where the conditional density for the $i$th individual is given by $f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = [G(\boldsymbol{x}_i\boldsymbol{\beta})]^{y_i}[1 - G(\boldsymbol{x}_i\boldsymbol{\beta})]^{1-y_i}$ and the conditional mean is defined as:

$$E[y_i|\boldsymbol{x}_i] = \frac{exp(\boldsymbol{x}_i\boldsymbol{\beta})}{[1 + exp\,(\boldsymbol{x}_i\boldsymbol{\beta})]} \tag{3}$$

Papke and Wooldridge (1996) propose a quasi maximum likelihood estimator (QMLE) of the coefficient vector $\boldsymbol{\beta}$. Here the sum of individual Bernoulli likelihood contributions

$$\mathcal{L}_i(\boldsymbol{\beta}) = f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = [G(\boldsymbol{x}_i\boldsymbol{\beta})]^{y_i}[1 - G(\boldsymbol{x}_i\boldsymbol{\beta})]^{1-y_i} \tag{4}$$

is maximized to obtain the QML estimator:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{L}_i(\boldsymbol{\beta}) \tag{5}$$

Note that as the Bernoulli distribution is a member of the linear exponential family, $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal irrespective of the true conditional distribution of $y_i$ given $\boldsymbol{x}_i$ as long as the conditional mean is correctly specified. This allows for the possibility that $y_i$ can be binary or continuous - for instance, it can take on corner values with positive probability and values in between with probability zero. In particular, the model accommodates corner solutions, i.e. situations in which there is a large amount of corner values. In such a situation the model still yields consistent estimates. For instance, in their own application on the participation rates of employees in 401(k) pensions plans, Papke and Wooldridge (1996) note that for about 43 % of the firms in their sample all employees participate in a pension scheme.

In a non-linear framework, the $\boldsymbol{\beta}$ coefficients are no longer equal to the partial effects. Instead, the partial effects are non-linear functions of the coefficients and exhibit the same sign as the betas. These are given by $PE_{ik} = \frac{\partial E[y_i|\boldsymbol{x}_i]}{\partial x_{ik}} = g(\boldsymbol{x}_i\boldsymbol{\beta})\beta_k$ where $g(\cdot)$ is the derivative of the link function with respect to its argument.[4] As the partial effects do depend on the values in $\boldsymbol{x}_i$ one is usually interested in the average partial effects (APE) given by $APE_k = E[PE_{ik}] = E[g(\boldsymbol{x}_i\boldsymbol{\beta})\beta_k]$. These are estimated by their sample analogs: $\widehat{APE}_k = \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i\hat{\boldsymbol{\beta}})\hat{\beta}_k$.

### 3.2 Fractional Response Models with Unobserved Heterogeneity

So far we have looked at the cross sectional case where one observes $N$ individuals at a given moment in time. Now we turn to the situation where one has access to panel data, i.e. where we observe the same agents repeatedly over time.[5] In this case $t$ denotes the index for the observed time periods from 1 to T. We denote the share for person $i$ in period $t$ as $y_{it}$. Here, we define the vector of shares for the $i$th individual over time as $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ and for the covariates as $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})'$.

An important issue in this context is that the variation over the two dimensions $i$ and $t$ is not the same. It rather stands to reason that observations will be correlated over time as the cross-sectional units exhibit unobserved time constant characteristics. The panel data literature is mainly concerned with how to deal with this unobserved heterogeneity which we denote as $\alpha_i$.

One way to proceed in this situation is simply to ignore the time dimension and estimate the coefficient vector as before by pooling over all observations. Hereby, one obtains the pooled (or partial) ML estimator by maximizing the pooled likelihood $\sum_{i=1}^{N}\sum_{t=1}^{T} log f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. In the Flogit specification the conditional density for the $i$th household in time-period $t$ is given by $f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\beta}) = [\Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta})]^{y_{it}}[1 - \Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta})]^{1-y_{it}}$. The resulting estimator is as robust as before since we do not restrict the relationship of the variables over time. In effect, one treats the unobserved time-invariant characteristics $\alpha_i$ as a nuisance term. The only practical difference to the cross-sectional case is that one has to account for the serial correlation over time induced by $\alpha_i$. This can be easily done in most statistical packages by clustering with respect to the cross-sectional unit. Papke and Wooldridge (2008) refer to this approach as pooled fractional

[4] For the logit specification $g(\cdot)$ is given by $\Lambda(\cdot)[1 - \Lambda(\cdot)]$.

[5] Generally panel data methods are still underused in household finance - one of the few exceptions is Alessie et al. (2004) who model the joint decision to hold stocks and bonds over time.

response model. For Germany, Eickelpasch and Vogel (2011) use this approach to estimate the effect of firm characteristics on export intensity.

Even though this approach is straightforward, it does not utilize the main potential benefits provided by panel data. For a start, as they do not account for the serial dependence over time, pooled models are generally less efficient than panel data models, that do account for the error structure directly. More importantly, panel data models allow one to account for potential bias due to correlation of covariates with unobservable time-constant individual characteristics. A well-known example comes from labor economics where one expects an upward bias for the return to education in a standard Mincer regression due to positive correlation of education level with workers unobserved ability. In the context of household finance one can think of unobserved household characteristics such as frugality or time preferences in financial matters. These are presumably constant over time and correlated with important determinants of portfolio composition such as level of wealth or risk aversion. Carroll (2002) hypothesizes that richer households hold more risky portfolios due to heterogeneous risk preferences across households. This in turn might lead to a situation where risk takers end up much richer than the rest of the population. This hypothesis implies that parts of this positive relationship are spurious and thus should vanish once one controls for unobserved household characteristics.

In the following we show how Papke and Wooldridge (2008) extend their original fractional response model to account for unobserved effects in a panel data framework. They let the time-constant unobserved heterogeneity term $\alpha_i$ enter the link function additively in the fashion of a single index model. Assuming that the covariates are strictly exogenous, given $\alpha_i$, they write the conditional mean as $E[y_{it}|\boldsymbol{X}_i, \alpha_i] = E[y_{it}|\boldsymbol{x}_{it}, \alpha_i] = G(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)$, $t = 1, \ldots, T$.[6] We continue to use the logistic link function which leads to the random effects Flogit model for which the conditional mean given the unobserved heterogeneity is written as:

$$E[y_{it}|\boldsymbol{x}_{it}, \alpha_i] = \Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i) = \frac{exp(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)}{[1 + exp(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]} \in (0, 1) \tag{6}$$

[6] Strict exogeneity is a standard assumption of static panel data models which rules out lagged dependent variables.

Note that the random effects Flogit model can be seen as a special case of the random parameters model where, instead of letting all parameters be random, one allows only for a random intercept. Cameron and Trivedi (2005) refers to this as neglected heterogeneity model.

If the individual characteristics $\alpha_i$ were observed one could condition on them and estimation would be straight forward. For instance, one could compute partial effects simply by plugging in the true values $\alpha_i$ into $PE_{ik} = g(\boldsymbol{x}_i\boldsymbol{\beta} + \alpha_i)\beta_k$. However, as this is clearly not feasible, the main challenge is to find an expression of $y_{it}$ that does not depend on $\alpha_i$ directly.

One approach to deal with this issue is to employ fixed-effects models. Thereby one conditions on a sufficient statistic which renders it unnecessary to deal with the unobserved heterogeneity. Yet only for few non-linear models a suitable sufficient statistic can be found. A well known example for the binary case is the fixed effects logit model where one conditions on all past successes. However, Papke and Wooldridge (2008) note that this model is not applicable to fractional dependent variables. In addition, they stress that even if it was possible to estimate such a model, it is not necessarily desirable to do so as marginal effects - the main interest of most analyses - are not identified (see also Wooldridge, 2010).

Thus the standard approach in the panel data literature on non-linear models is to define $\alpha_i$ as a random variable with a given distribution. At the same time one assumes a suitable distribution for $y_{it}$, conditional on the unobserved heterogeneity $\alpha_i$. In this fashion one can integrate out the individual-specific effects to obtain the unconditional joint distribution of $y_{it}$.

In addition to strict exogeneity usually independence over time conditional on $\boldsymbol{X}_i$ and $\alpha_i$ is assumed, i.e. $f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}) = \prod_{t=1}^{T} f(y_{it}|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta})$. For the Flogit specification the conditional density of household $i$ in time-period $t$, given the individual-specific effect, reads as $f(y_{it}|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta}) = [\Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]^{y_{it}}[1 - \Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]^{1-y_{it}}$. Then, the likelihood contribution of the $i$th household, i.e. the joint distribution conditional on $\alpha_i$, becomes:

$$\mathcal{L}_i(\boldsymbol{\beta}) = f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}) = \prod_{t=1}^{T}[\Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]^{y_{it}}[1 - \Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]^{1-y_{it}} \tag{7}$$

The unconditional joint distribution of shares for the $i$th household which no longer depends on $\alpha_i$ is obtained by integrating out the unobserved heterogeneity:

$$f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \boldsymbol{\beta}) = \int_{-\infty}^{\infty} \left[ \prod_{t=1}^{T} f(y_{it}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}) \right] h(\alpha_i) d\alpha_i \tag{8}$$

This can also be seen as the expected likelihood contribution of a given agent: $E_\alpha[\mathcal{L}_i(\boldsymbol{\beta})]$. Usually a normal distribution is assumed for the unobserved heterogeneity but other distributions or a semi-parametric approach are equally possible. Here, we follow an approach similar to Papke and Wooldridge (2008) and assume that $\alpha_i$ is normally distributed : $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$.[7] The maximum likelihood estimator for the coefficient vector $\boldsymbol{\beta}$ is obtained by maximizing the sum of log-likelihoods with respect to $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \boldsymbol{\beta}) \tag{9}$$

It is important to note that, except for special cases, there will be no analytical solution for the univariate integral over the individual-specific effect $\alpha_i$ in Equation 8. Thus, some kind of numerical integration is needed - either via deterministic methods (Gaussian quadrature) or simulation methods (Monte Carlo Integration). Normally Gaussian quadrature methods are used as they are fairly reliable and easy to compute. Quadrature routines are implemented in most standard statistical packages.

The main drawback of the random effects approach is that the individual-specific effects $\alpha_i$ are assumed to be independent from the covariates. As mentioned before, we apply panel data models in large part to account for the potential confounding influence of $\alpha_i$. We have also seen that fixed effects models, if at all feasible, are quite restrictive in a non-linear context. Thus Papke and Wooldridge (2008) advocate an approach which can be seen as a middle ground between the rather unrealistic RE approach and the more restrictive FE models. They employ a correlated random effects (CRE) model as introduced by Mundlak (1978) and refined by Cham-

[7]The zero mean assumption is unproblematic as long as there is an intercept in the model as a non-zero mean of the unobserved heterogeneity is then absorbed by the intercept.

berlain (1984). In the CRE approach one models the time-constant unobserved heterogeneity $\alpha_i$ as a linear combination of the time averages of the time varying covariates:

$$\alpha_i = \bar{\boldsymbol{x}}_i \boldsymbol{\xi} + a_i \tag{10}$$

Where $\bar{\boldsymbol{x}}_i$ is the vector of time averages of the covariates for the $i$th household ($\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$). Here $a_i$ is assumed to be normally distributed similar as before: $a_i \sim \mathcal{N}(0, \sigma_a^2)$. Then, $\alpha_i$ is also normally distributed conditional on the covariates: $\alpha_i | \boldsymbol{X}_i \sim \mathcal{N}(\boldsymbol{x}_i \boldsymbol{\xi}, \sigma_a^2)$. In this manner one allows for a relationship between the covariates and the unobserved heterogeneity even though one has to restrict the relationship somewhat compared to the FE approach where it is completely unrestricted.[8]

Our choice of the logistic link function requires some discussion. In principle either a logit or probit specification is possible for the distribution of $y_{it}$ similar to the cross-sectional case. However, in the panel data literature the random effects probit model is typically preferred. Likewise, Papke and Wooldridge (2008) choose a probit specification in contrast to the logistic link used in Papke and Wooldridge (1996). The reason for this is that if a normal distribution is assumed for $\alpha_i$, the resulting random effects fractinal probit model allows for simple computation of the average partial effects from the scaled coefficients due to the mixing property of two normally distributed random variables (see also Wooldridge, 2010). In contrast, it is not as straight forward to obtain APE's in a random effects logit specification. Here we focus on the RE Flogit model due to the use of a multinomial logit specification in the multivariate case.

Even though the average partial effects cannot be obtained from the scaled coefficients in this case, it is still easy to come by the APEs if one uses simulation methods to integrate out the unobserved heterogeneity. As already mentioned, deterministic numerical integration is preferred for univariate integrals as simulation methods are computationally more intensive. Yet, simulation methods are better suited for multidimensional integrals as multivariate quadrature methods quickly become infeasible with higher dimensionality. Hence, we will be in need of

[8]Note that the distinction between the CRE and the FE approach exists only for non-linear models - in the linear case they result in the same estimator (see Wooldridge, 2010).

simulation methods for the multivariate case anyway. In the following we will give a short introduction to Monte Carlo Integration and the resulting Maximum Simulated Likelihood (MSL) estimator. The concepts for the univariate case easily generalize to higher dimension. For the rest of the discussion we borrow heavily from chapter 12 of Cameron and Trivedi (2005). For a more general exposition to Monte Carlo Integration and Maximum Simulated Likelihood please refer to this chapter or the extensive treatment in Train (2003).

Consider a situation, similar to the one in Equation 8, where we want to solve an intractable integral of the form

$$f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = \int_{-\infty}^{\infty} f(y_i|\boldsymbol{x}_i, \alpha_i, \boldsymbol{\beta}) h(\alpha_i) d\alpha_i \tag{11}$$

where $f(\cdot)$ is the function to be integrated and $h(\cdot)$ is a known pdf. The basic idea of Monte Carlo integration is to sample from the distribution $h(\alpha_i)$. One plugs the $R$ simulated values $\alpha_i^r$ into the function to be integrated. We obtain the Monte Carlo integral, which is the simulated likelihood contribution of agent $i$, by averaging over all resulting expressions:

$$\mathcal{SL}_i(\boldsymbol{\beta}) = \widehat{f}(y_i|\boldsymbol{x}_i, \alpha_{iR}, \boldsymbol{\beta}) = \frac{1}{R} \sum_{r=1}^{R} f(y_i|\boldsymbol{x}_i, \alpha_i^r, \boldsymbol{\beta}) \tag{12}$$

This Monte Carlo estimator of $f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta})$ in turn is used to perform the maximum simulated likelihood (MSL) estimation which by the law of large numbers yields a consistent estimator for the true coefficient vector $\boldsymbol{\beta}$ as $R \to \infty$:

$$\widehat{\boldsymbol{\beta}}_{MSL} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log \mathcal{SL}_i(\boldsymbol{\beta}) \tag{13}$$

From this it is easy to see how to to obtain the partial effect $PE_{ik} = E_\alpha \left[ g(\boldsymbol{x}_i\boldsymbol{\beta} + \alpha_i) \right] \beta_k$. We plug the simulated values of the unobserved heterogeneity into the formula for the partial effects and average them out. This leads to a consistent estimator for the true partial effects as the size of the simulated sample increases: $\widehat{E}_\alpha \left[g(\boldsymbol{x}_i\boldsymbol{\beta} + \alpha_i)\beta_k\right] = \frac{1}{R}\sum_{r=1}^{R} g(\boldsymbol{x}_i\boldsymbol{\beta} + \alpha_i^r)\beta_k \xrightarrow{p} E_\alpha \left[g(\boldsymbol{x}_i\boldsymbol{\beta} + \alpha_i)\beta_k\right]$ as $R \to \infty$. The average partial effects are then estimated as:

$$\widehat{APE}_k = \frac{1}{NR}\sum_{r=1}^{R}\sum_{i=1}^{N} g(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}} + \alpha_i^r)\widehat{\beta}_k$$

Compared to the conditional mean approach used for cross-sectional analysis, as described in Subsection 3.1, the fully parametric approach employed here requires stronger assumptions and is computationally more intensive. However, the named potential benefits - especially the ability to control for individual characteristics such as thriftiness or foresightedness in financial matters - likely more than outweigh these downsides.

### 3.3 Fractional Multinomial Response Models

The aim of our analysis is not to estimate the conditional mean for a single share alone but rather for several shares that together comprise the underlying total. As mentioned in the introduction most empirical studies focus on the share of wealth invested in risky assets. Yet, some studies also examine other aspects of households' portfolios besides the proportion allotted to equities. However, most of these papers employ univariate models for each individual share and thus cannot capture the relationship between asset classes. For instance, Börsch-Supan and Eymann (2002) separately examine the determinants of the shares of fairly safe and risky assets for Germany in the 1990's. Rosen and Wu (2004) follow a similar approach for four financial asset types for the United States. These approaches ignore the fact that share levels depend upon each other. Here we follow Mullahy (2011) who models the shares of several financial assets in a joint framework via a multivariate fractional response model by means of the Survey of Consumer Finance (SCF) for the United States.[9]

For modeling a multivariate framework with $J$ different assets, we return to the cross-sectional case. We denote the share of the $j$th asset held by the $i$th individual as $y_{ij}$. A suitable model for this situation must reflect the bounded nature of each individual share (i.e. $0 \leq y_{ij} \leq 1$

[9] For the participation decisions Bertaut and Starr-McCluer (2002) and Alessie et al. (2004) provide joint estimations.

for $j = 1 \dots J$) as well as the fact that shares have to add up to unity (i.e. $\sum_{j=1}^{J} y_{ij} = 1$). This implies that the resulting predicted shares from such a model should also lie between zero and one (i.e. $E[y_{ij}|\boldsymbol{x}_i] \in (0,1)$ for $j = 1 \dots J$) and add up to one (i.e. $\sum_{j=1}^{J} E[y_{ij}|\boldsymbol{x}_i] = 1$). The latter condition also implies that the marginal effects for a system of equations with the same covariates in each equation have to sum up to zero. Such a behavior is also expected in the context of asset shares as the increase in the share of one asset has to come at the expense of other assets. Overall the changes induced by the change in a covariate should sum up to zero.

In general, estimating the conditional mean for each share individually (as done by Rosen and Wu, 2004; Wachter and Yogo, 2010) does not guarantee to fulfill these necessary properties. For instance, Rosen and Wu (2004) estimate several asset shares individually via univariate Tobit models. They note that it is not ensured that the predicted shares will add up to one without imposing constraints on the Tobit estimations.[10] It therefore stands to reason that the most straightforward way to estimate several shares is in a joint framework. This is the approach taken by two recent papers, Mullahy (2011) and Murteira and Ramalho (2013). They both concentrate on multivariate fractional dependent data where the main focus is on modeling the conditional mean of shares jointly.

A natural way to approach this is by proceeding analogously to the discrete choice setting. There, binary choice models, which are used in situations where an agent has to choose between two different possibilities, are generalized to model the decision between several unordered alternatives via multinomial choice models. In the same fashion one can extend the fractional response models by Papke and Wooldridge (1996) to fractional multinomial response models in order to estimate several shares at once. In principle several link functions are possible in this context but often a multinomial logit specification is employed. The reason being that this choice drastically simplifies the computational burden compared to, for instance, a multinomial probit specification because no correlations across alternatives are assumed (see chapter 15 in Cameron and Trivedi, 2005). Extending the Flogit model from Subsection 3.1 in this fashion is straight forward. Using a multinomial logit specification as link function results in the so called

[10] However, they assert that the sum of the marginal effects of the individual equations is close enough to zero to conclude that this is a minor problem in their application.

fractional multinomial logit model (to which we will refer to as FMlogit). This is the main model specification in both Mullahy (2011) and Murteira and Ramalho (2013) and is given by:[11]

$$E[y_{ij}|\boldsymbol{x}_i] = \Lambda(\boldsymbol{x}_i\boldsymbol{\beta}_j) = \frac{exp(\boldsymbol{x}_i\boldsymbol{\beta}_j)}{\left[\sum_{h=1}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}, \qquad j = 1\ldots J \tag{14}$$

Mullahy (2011) mentions several applications of this model. For instance, Koch (2010) uses a multinomial fractional response model to estimate expenditure shares in South Africa. It is easy to see that this specification naturally enforces the constraints outlined above. Estimating the fractional multinomial logit model, as in the discrete case, requires some normalization - usually by setting the coefficients of the first equation to zero: $\boldsymbol{\beta}_1 = \mathbf{0}$. Thus, the conditional expectations for all the equations can be written as:

$$E[y_{ij}|\boldsymbol{x}_i] = \frac{1}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}, \qquad j = 1 \tag{15}$$

$$E[y_{ij}|\boldsymbol{x}_i] = \frac{exp(\boldsymbol{x}_i\boldsymbol{\beta}_j)}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}, \qquad j = 2, \ldots, J \tag{16}$$

It is important to point out that in this case the betas give even less information regarding the partial effect of a variable on the conditional mean compared to the univariate case where one could at least infer the direction and significance of an effect. This lack of information is due to the fact that the weighted sum of all other betas is needed to calculate the partial effects. This can be seen by writing out the partial effect of the $k$th regressor on the $j$th share:

$$PE_{ijk} = \frac{\partial E[y_{ij}|\boldsymbol{x}_i]}{\partial x_{ik}} = E[y_{ij}|\boldsymbol{x}_i] \cdot \left[\beta_{jk} - \frac{\sum_{h=2}^{J} \beta_{hk} exp(\boldsymbol{x}_i\boldsymbol{\beta}_j)}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}\right] \tag{17}$$

For this reason we will mainly report the estimated average marginal effects[12] when presenting our results as these can be readily interpreted in the usual way. Compared to a situation where

[11] An implementation for Stata® is provided by Buis (2008) in an ado file named `fmlogit`.

[12] $\widehat{APE}_{jk} = \frac{1}{N}\sum_{i=1}^{N} \widehat{PE}_{ijk}$

one estimates each share individually, it is an advantage of this joint framework that the marginal effects are bound to cancel each other out.

Analogously to the univariate case one can define the quasi maximum likelihood estimator for the multinomial logit specification by writing the likelihood contribution of a single agent:

$$\mathcal{L}_i(\boldsymbol{\beta}) = \prod_{j=1}^{J} E[y_{ij}|\boldsymbol{x}_i]^{y_{ij}} \tag{18}$$

Again, the sum of the individual log-likelihoods is maximized to obtain the estimator for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{L}_i(\boldsymbol{\beta}) \tag{19}$$

Murteira and Ramalho (2013) note that the multinomial fractional logit model exhibits the well known independence of irrelevant alternatives (IIA) property which implies a very restrictive substitution pattern over shares. Namely, the ratio between two shares will not depend on the characteristics of other shares, i.e. the substitution patterns are reduced to pairwise comparisons. This is due to the simplifying assumption of independence over equations in the model which is unlikely to hold in the application to asset shares. Murteira and Ramalho (2013) suggest alternative models, such as the nested logit or the mixed logit which are not afflicted with this issue. In the latter model parameters are assumed to be random, i.e. different for each agent. Allowing these random parameters to be correlated across equations leads to unrestricted substitution patterns so that the ratio of two shares is no longer independent of the other alternatives. Therefore, in the next subsection we will look at a special case of this model in more detail.

Besides the conditional mean models presented above, both Mullahy (2011) and Murteira and Ramalho (2013) consider fully parametric models which model the entire joint conditional distribution of shares. The main candidate for this approach is the Dirichlet-Multinomial (DM) model which is the multivariate extension of the beta-binomial model in the univariate case. Both papers note that this model is potentially attractive as it allows one to model other features of the distribution in addition to its mean, such as the probability of corner outcomes. Moreover, it is potentially more efficient if the true underlying distribution follows a DM density. However, the main disadvantage of this modeling strategy is that one has to make assumptions about the entire distribution of shares which might easily be violated in practice. This is particularly severe as the DM distribution is not robust to misspecifications as the fractional multinomial logit. Furthermore, in a situation where the underlying total is not the same for every individual one has to transform the data in order to make it suitable for a DM regression model. This transformation is arbitrary and potentially leads to inconsistent estimations.

Both papers compare these two approaches to assess their validity. Murteira and Ramalho (2013) conduct Monte Carlo studies for both types of models and find that the DM model at best yields only modest advantages in terms of efficiency compared to the fractional multinomial logit. At the same time inconsistencies seem to be a problem in the fully parametric approach. Mullahy (2011) applies both the fractional multinomial logit model and the DM model for shares of financial assets to the SCF data set. The average partial effects for both models are roughly similar and there are no clear indications of an efficiency gain of the DM model compared to the conditional mean model. Overall, the results of his application give little support for the DM model especially with regard to the non-robustness of the method. All in all, this evidence does not speak in favor of the fully parametric approach.

### 3.4 Fractional Multinomial Response Models with Unobserved Heterogeneity

Murteira and Ramalho (2013) suggest that an application of the fractional multinomial logit model in a panel data context looks very promising. A combination of the two extensions of the basic Flogit model presented in Subsection 3.2 and Subsection 3.3 does indeed seem to suggest itself. However, to the best of our knowledge, ours is the first study to extend the fractional multinomial logit model in this fashion. We follow the approach by Haan and Uhlendorff (2006) who implement a multinomial logit model with random intercepts for panel data via maximum simulated likelihood in Stata®. In this context the share of asset $j$ for household $i$ in time-period $t$ is written as $y_{ijt}$. As before the shares $y_{ijt}$ lie between zero and one and add up to unity over all $J$ categories. We write the vector of shares for household $i$ and asset $j$ over time as $\boldsymbol{y}_{ij} = (y_{ij1} \dots y_{ijT})'$ and the vector of all shares for household $i$ as $\boldsymbol{y}_i = (\boldsymbol{y}_{i1} \dots \boldsymbol{y}_{iJ})'$. Correspondingly to the authors' proceeding for categorical dependent variables, we define the mean share of asset $j$ for household $i$ in period $t$, conditional on the covariates $\boldsymbol{x}_{it}$ and a $J \times 1$ vector of unobserved heterogeneity $\boldsymbol{\alpha}_i$, as:

$$E[y_{ijt}|\boldsymbol{x}_{it}, \boldsymbol{\alpha}_i] = \frac{exp(\boldsymbol{x}_{it}\boldsymbol{\beta}_j + \alpha_{ij})}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}_h + \alpha_{ih}\right)\right]} \tag{20}$$

We allow the unobserved heterogeneity $\boldsymbol{\alpha}_i = (\alpha_{i1}\ \alpha_{i2}\ \dots\ \alpha_{iJ})'$ to affect each share differently. In addition, we allow the individual-specific effect to be correlated over equations. One usually assumes that $\boldsymbol{\alpha}_i$ follows a multivariate normal distribution with unrestricted variance-covariance matrix.

Haan and Uhlendorff (2006) remark that such a model can be seen as a a special type of a mixed logit model. Mixed logit models are a generalization of the multinomial logit model where one allows the parameter vector $\boldsymbol{\beta}$ to be different for each individual agent and assumes a distribution of these random coefficients $\boldsymbol{\beta}_i \sim g(\boldsymbol{\beta})$. The model presented here is a special case insofar as we only allow for random intercepts $\alpha_{ij}$ and use the same covariates $\boldsymbol{x}_{it}$ in each equation instead of letting them vary over shares.

Mixed logit models have become increasingly popular in recent years. Hole (2007) demonstrates a simple implementation of mixed logit models in Stata®. Revelt and Train (1998) use mixed logit models for repeated choices in a panel data context to estimate the determinants of buyer's choice of refrigerator efficiency. Train (2003) offers an excellent exposition on mixed logit models. With regard to the advantages of mixed logit models over multinomial logit models he notes that mixed logit models allow "for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time" (Train, 2003, p.138). Thus, the aforementioned IIA property that afflicts the FMlogit model can be avoided by such a random effects model as suggested by Murteira and Ramalho (2013). Furthermore, accounting for unobserved heterogeneity in a panel data context potentially leads to a more efficient estimator compared to a pooled model with clustered standard errors. Most importantly, this approach allows to control for time-invariant unobserved heterogeneity in a CRE framework analogous to the univariate case.

The estimation of the model via maximum simulated likelihood corresponds to the procedures in Subsection 3.2 and Subsection 3.3. We write the likelihood contribution of a single household $i$ for all shares and time periods conditional on the unobserved effects as:

$$\mathcal{L}_i(\boldsymbol{\beta}) = f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\alpha}_i, \boldsymbol{\beta}) = \prod_{j=1}^{J}\prod_{t=1}^{T} E[y_{ijt}|\boldsymbol{x}_{it}, \boldsymbol{\alpha}_i]^{y_{ijt}} \tag{21}$$

As we do not observe the household-specific effects we can only write the expectation over the multivariate distribution of $\boldsymbol{\alpha}_i$ for which we have to solve the corresponding multidimensional integral:

$$E_{\boldsymbol{\alpha}}[\mathcal{L}_i(\boldsymbol{\beta})] = \int_{-\infty}^{\infty} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\alpha}_i, \boldsymbol{\beta}) f(\boldsymbol{\alpha}_i) d\boldsymbol{\alpha}_i \tag{22}$$

One approximates this expression by drawing $R$ values $\boldsymbol{\alpha}^r$ from the corresponding multivariate distribution and sum over the draws which leads to the simulated likelihood contribution for each agent:

$$\mathcal{SL}_i(\boldsymbol{\beta}) = \frac{1}{R}\sum_{r=1}^{R} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\alpha}_i^r, \boldsymbol{\beta}) \tag{23}$$

The resulting estimator maximizes the sum of log simulated likelihoods:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{SL}_i(\boldsymbol{\beta}) \tag{24}$$

Then the average partial effects can be estimated as explained before by plugging in the realizations of $\boldsymbol{X}_i$ and $\boldsymbol{\alpha}_i^r$ and average over all observations and the simulated heterogeneity.

The main disadvantage of this approach is that one has to solve multivariate integrals in the computation of the likelihood function. In principle it is possible to approximate them via multivariate quadrature methods. However, this approach is computationally expensive which is why we implement the model via maximum simulated likelihood estimation.[13] When we described the MSL approach in Subsection 3.2 we did not say anything about how to actually

[13] We implemented the model in Mata® due to speed gains.

draw from the distribution of unobserved heterogeneity. The standard Monte Carlo simulation suggests itself but often exhibits a bad coverage over the domain of integration. Instead, one often uses quasi-random sampling methods such as Halton sequences where draws are no longer independent from one another. As a consequence, the coverage achieved by Halton sequences is much better compared to independent random sampling. Train (2003) notes that for mixed logit model convergence is achieved much faster with Halton sequences compared to standard simulation methods - the required number of draws is about an order of magnitude lower. For the implementation of our model we thus use the `mdraws` command by Cappellari and Jenkins (2006) which allows to generate Halton sequences in Stata®. Haan and Uhlendorff (2006) also use `mdraws` and compare the simulation method to deterministic quadrature. While finding no advantage of MSL for univariate integrals, they note that it is much faster compared to quadrature for higher dimensional integrals. In addition, they state that after 100 Halton draws the estimation results are stable. We come to the same conclusion in our application. Drukker and Gates (2006) provide another way to draw from Halton sequences in Mata®. Bhat (2001) discusses the application of Halton sequences to estimate mixed logit models.

To illustrate the structure of the unobserved heterogeneity, consider a situation where J = 3. For the purpose of identification the coefficients for the first equation, $\boldsymbol{\beta}_1$ and $\alpha_1$ have been normalized to zero and the unobserved heterogeneity $\boldsymbol{\alpha}_i$ is assumed to follow a multivariate normal distribution. Allowing for correlation across equations, we define the household-specific effects as $\boldsymbol{\alpha}_i = \boldsymbol{L}\boldsymbol{\epsilon}_i$ where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\boldsymbol{L}$ is a lower-triangle matrix resulting from the cholesky-decomposition of the variance-covariance matrix $\boldsymbol{\Sigma_a}$: $\boldsymbol{LL'} = \boldsymbol{\Sigma_a}$. The elements of $\boldsymbol{L}$ have to be estimated along with $\boldsymbol{\beta}$.

$$\boldsymbol{\alpha}_i = \begin{pmatrix} \alpha_{i2} \\ \alpha_{i3} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{i2} \\ \epsilon_{i3} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma_\alpha}) = \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha_2} & \sigma_{\alpha_2\alpha_3} \\ \sigma_{\alpha_2\alpha_3} & \sigma^2_{\alpha_3} \end{pmatrix} \right) \tag{25}$$

In generating $\boldsymbol{\alpha}_i^r$ one simulates $\boldsymbol{\epsilon}_i$ by drawing via Halton sequences from a standard multivariate normal distribution: $\boldsymbol{\alpha}_i^r = \boldsymbol{L}\boldsymbol{\epsilon}_i^r$. To accommodate correlated random effects in this framework we write the random effects for each equation similar to the univariate case as $\alpha_{ij} = \bar{\boldsymbol{x}}_i\boldsymbol{\xi}_j + a_{ij}$.

Then $\boldsymbol{a}_i$ is assumed to be multivariate normal with variance-covariance matrix $\boldsymbol{\Sigma_a}$ and $\boldsymbol{\alpha}_i$ is conditionally multivariate normal distributed as before:

$$\boldsymbol{\alpha}_i|\boldsymbol{X}_i \sim \mathcal{N}(\boldsymbol{f}(\bar{\boldsymbol{x}}_i), \boldsymbol{\Sigma_a}) = \mathcal{N}\left(\begin{pmatrix}\bar{\boldsymbol{x}}_i\boldsymbol{\xi}_2\\ \bar{\boldsymbol{x}}_i\boldsymbol{\xi}_3\end{pmatrix}, \begin{pmatrix}\sigma^2_{a_2} & \sigma_{a_2a_3}\\ \sigma_{a_2a_3} & \sigma^2_{a_3}\end{pmatrix}\right) \tag{26}$$

# 4 Data

## 4.1 SAVE Survey

The data for our empirical analysis stems from the SAVE survey (”Sparen und Altersvorsorge in Deutschland”), a representative panel study on some 2,000[14] private German households.[15] The survey was established in 2001 by the Mannheim Research Institute for the Economics of Aging (MEA) with a focus on saving and investment decisions on the household level in Germany, a research area for which there was little micro data available previously. Therefore, the SAVE survey is the first micro-level panel dataset for Germany which provides detailed information on the financial situation of households. The responding family member is asked about the wealth composition of the household as a whole. Thus, the subsequent analysis is carried out on the household level. Specifically, households report the amount of money invested in a wide range of financial assets such as stocks, bonds or saving accounts, which can be used to compute the share of financial wealth invested in each asset class. These data are supplemented by various questions about the sociodemographic make-up of a household including characteristics like employment status, education level or household income.[16] Furthermore, a plurality of potential determinants of a household’s investment behavior such as self-assessed risk aversion, attitudes toward life or expectations of the economy are provided. This wealth of information makes the SAVE an ideal survey for the analysis of portfolio composition.

Two other micro panels that are potentially viable for panel data analyses of the financial portfolio of German households are the German Socio Economic Panel (GSOEP) and the Panel

[14] As of 2010 for its ninth wave.

[15] See Börsch-Supan et al. (2008) for a more detailed account on the survey.

[16] The interviewed household member who answers questions regarding the personal level is not necessarily the household head.

of Household Finance (PHF) by the German Bundesbank.[17] The GSOEP is the longest running panel survey in Germany (currently in its 28th wave) and has a vast array of different covariates to offer. In addition, the number of households asked in each year is significantly higher than in the SAVE survey. The main disadvantage of the GSOEP for our analysis is that it offers very little information on the financial situation of households. Usually it only asks for participation in certain financial assets such as stocks or savings accounts. The annual GSOEP waves do not provide information on the exact amount of money invested in these assets. Some papers, such as Barasinska et al. (2012) and Dierkes et al. (2011), use the GSOEP to estimate participation rates in asset classes. For 2002 and 2007 households where asked about their wealth situation but in a much less detailed way compared to the SAVE survey. In contrast, the PHF asks for financial wealth in even greater detail than the SAVE survey. Otherwise it is comparable to the SAVE survey in the number of households participating and the type of questions asked. However, currently it is not possible to conduct panel data analysis with the PHF as only its first wave has been released as yet. Furthermore, the survey is conducted only every three years. Hence, the SAVE survey is better suited for our analysis than the aforementioned surveys.

### 4.2 Sample

After the first wave in 2001, a main random sample was established for the year pair 2003 and 2004. Hereafter consecutive surveys where conducted for every year between the years 2005 and 2010. For the year pair 2011 and 2012, only a single reduced questionnaire was issued which solely covers core variables and does not provide information on household asset holdings.

We employ a sample consisting of six consecutive waves of the SAVE survey from 2005 to 2010. Our selection is due to the fact that the resulting sample offers the highest level of consistency in terms of sample composition and questions asked. Moreover, this time span is particularly interesting in the context of this paper as it includes the financial crisis of 2007 to 2008 which potentially had a huge impact on the behavior of retail investors.

[17] For more information on these surveys see Wagner et al. (2008) for the GSOEP and von Kalckreuth et al. (2012) for the PHF.

Overall there are 15,587 observations for 3,941 different households in the sample. Out of these 15,587 observations 13,475 exhibit positive levels of gross wealth (86.5 % of the sample). Furthermore, financial wealth holding is observed for 12,420 observations (80 % of the sample). For the purpose of our analysis these households constitute the target population as only they allow us to construct financial asset shares.

In addition, we exclude observations with very low levels of financial wealth - namely those with less than 100 € invested in financial assets. The reason being that asset shares computed for such low levels are notoriously unreliable.[18] We also exclude households which own business assets or are headed by a self-employed person. This is because these households are facing an enormous background risk which one cannot control for easily within a regression framework. Thus, the analysis is limited to the part of the population that is not exposed to this kind of background risk. Finally, observations with missing values for key variables such as income are excluded. This leaves us with 10,995 observations for 3,232 unique households. If not otherwise indicated, the subsequent summary statistics and regression results are based on these observations and have to be interpreted as being conditional on positive amounts of financial wealth as well as a lack of business activities.[19]

### 4.3 Financial Asset Classes

The most interesting aspect of the SAVE survey within the context of this paper is the section on financial asset holding. Participants have to state whether they have invested money in a number of different assets and the amount of money invested in the given class. Financial wealth is subdivided into two different categories in the questionnaire - monetary assets (”Geldvermögen”) and retirement provision. The first category is made up of five different asset classes: (i) savings accounts, (ii) building savings contracts, (iii) fixed income securities/bonds, (iv) common forms of equity such as direct stock-holding or traditional funds and (v) less common forms of equities

[18]One could argue for a higher threshold in the vicinity of 500 € or 1000 € as only then serious portfolio decisions can be made. However, restricting the sample any further does not alter the sample composition or the regression results much. Thus, we stick to the lower threshold.

[19]As certain groups, such as wealthy households, are oversampled in the SAVE survey we also use appropriate weights for all summary statistics to ensure their representativeness.

such as hedge-funds or financial innovations. The section on old-age provision consists of the four categories (i) whole life insurance policies, (ii) company pension plans, (iii) private retirement schemes and state-subsidized retirement plans (”Riester-Rente”).

It is a well established fact in the literature on the investment behavior of private households that the typical financial portfolio is poorly diversified - see for instance Barasinska et al. (2012). This phenomenon can also be seen in our data where more than 50 % of households in the sample have invested their financial wealth in only one or two out of the nine existing classes and 72 % hold three assets or less. Another way in which this under-diversification becomes apparent is through the fact that on average 74 % of financial wealth is held in a single asset class. Furthermore, we can see in Table 1 that the share conditional on ownership for every asset (except for Riester pensions) lies above 30 %. This underscores that most people allocate their money very roughly.

*Insert Table 1 about here.*

Looking more closely at the composition of household portfolios Table 1 list the make up of the different assets in our sample. As one would expect due to their conceptual simplicity and their lack of entrance costs, savings accounts constitute the most common asset type in the sample. On average 35 % of a household’s financial portfolio is made up of money held in one or several saving accounts. Almost three quarters of the households in the sample own at least one savings account which is the highest fraction for any one asset by far.

Two other very popular investment vehicles in Germany are building saving contracts and whole life insurance policies both of which are long-term illiquid investments with fixed rates. Consequently, the participation rate for both assets lies above 40 % in the sample which makes them the most widely used forms of investment after savings accounts. They constitute 13.5 % and 19.5 % of the average portfolio, respectively.

The bond class consists of a wide range of fixed-income securities such as government bonds or corporate bonds. Yet, only few households in our sample (12 %) incorporate bonds into their portfolio and the overall share of this asset class lies below 5%. The three categories for pension schemes in the survey - private pensions, company pensions and ”Riester” pensions each feature share rates around 5%. On average, households have invested about 17 % of their financial wealth in these three pension schemes.

A glance at the two equity classes reveals that about 29 % of the sample members hold stocks or traditional funds which account for about 10 % of their financial wealth on average. Less common forms of equity play virtually no role in most people's portfolio. However, for those households that have actually invested in this asset type, it constitutes almost a third of their financial wealth. Overall 11 % of financial wealth is invested in equities and slightly less than a third of the sample participates in these assets. This phenomena is well known as the "stock holding puzzle": only a minority of households hold any form of equity even though historically these assets have achieved abnormal risk premia compared to other investments.[20] In their seminal paper Haliassos and Bertaut (1995) analyze this phenomenon for the United States and conclude that inertia and departure from expected-utility maximization can partly explain this apparent underinvestment. This puzzle was scrutinized many times in subsequent research. For instance, the studies in Guiso et al. (2002) analyze, stock holding propensity for different industrialized countries. Generally one finds that this puzzle is more pronounced for continental European countries compared to Anglo-Saxon countries. For Germany Börsch-Supan and Essig (2002) note that during the 1990's the number of households that hold stocks have increased markedly, albeit from a very low level.

### 4.4 Construction and Composition of Risk Classes

Our main interest is to get a more complete picture of the risk profile of households' financial portfolios. Hence, we are not interested in individual assets that are similar in nature to each other.[21] In order to divide financial products into different categories we need to differentiate them by their degree of inherent riskiness. However, in the data one can only observe the amount of money invested in a given class but not the actual return on an asset. Thus, an

[20] This finding is closely related to the so called "equity premium puzzle" as first coined by Mehra and Prescott (1985) (See also Gollier, 2002, pp.34).

[21] For instance, the SAVE survey separately asks for investments in so called Riester pension products - a type of government subsidized retirement accounts introduced in 2002 - even though these products are very similar to other private pension schemes. This facilitates the analysis of adoption rates of this new scheme in the population as done by Börsch-Supan et al. (2012). In the context of this paper, however, such a distinction is of no interest.

allocation of assets by means of a mean-variance approach is not feasible here. Instead, we opt for a classification strategy which is widely used in the household finance literature[22] - we divide financial assets into the three categories "clearly safe" assets, "fairly safe" assets and "clearly risky" assets. As Carroll (2002) notes, the distinction between safe and risky assets, as used in theoretical analysis, is not easily applicable to empirical research. The main caveat is that most assets are neither completely safe nor can they be considered to be clearly risky. The general consensus among empirical researchers in this field is therefore to regard equities as clearly risky while saving accounts are considered to be clearly safe. The remainder of the assets are then subsumed into a middling category which we denote as "fairly safe".

However, for these assets some ambiguity remains regarding whether they better belong into the "clearly safe" or the "clearly risky" category instead. This uncertainty is due to the relatively high level of aggregation in the SAVE survey and the corresponding lack of insight into the composition and risk exposure of each class. One example in this regard is the "fixed income securities" category, which encompasses bonds with very different risk structures. For instance, short term German government bonds are generally considered to be almost risk-free. Compared to this, other bonds face higher levels of risk: a longer maturity introduces inflation risks, a lower rating grade leads to a higher default risk and bonds issued in other currency areas are exposed to currency risk. Thus to take another example, long-term, low-quality, corporate bonds can represent a rather risky investment class. The key problem in this context is that we cannot distinguish between the different types of bonds and assign them to the appropriate category.

*Insert Figure 1 and Table 2, Table 3 about here.*

The categorization of asset classes according to their inherent risk is summarized in Table 2. Table 3 and Figure 1 give an impression of the distributions of the share of financial wealth invested in each of the three risk categories. Table 3 provides information about the mean and median share invested in each risk category along with the percentage of households that have invested no money, parts of their monetary assets or all their monetary assets in a given class.

[22] This approach is employed throughout Guiso et al. (2002) (in particular by Bertaut and Starr-McCluer (2002), Börsch-Supan and Eymann (2002) and Carroll (2002)) and adopted by subsequent paper, for instance Atella et al. (2012) or Barasinska et al. (2012).

The information in Table 3 complement those given in Table 1 which provides summary statistics for the risk classes along with those for the financial assets themselves. We can see that both an average household and a household at the median hold the majority of their financial wealth in the "fairly risky" category. This is to be expected as most assets cannot be categorized as safe or risky with certainty. In addition, it is apparent that distributions for "clearly safe" and "clearly risky" assets are skewed to the right as their median shares are much lower than their mean shares - the median share for risky assets is even 0 %. The exact shape of the distributions can be seen in Figure 1 where we plot histograms for each category next to each other. It is evident that there is a significant number of corner values for each share. This is especially pronounced for risky assets for which 67 % of the sample population does not participate in. As we have emphasized in Section 3 the prevalence of such corner values does not affect the validity of our modeling strategy as we are mainly interested in making statements about the conditional mean of shares.

### 4.5 Explanatory Variables and Descriptive Analysis

At this point we give a descriptive overview over the explanatory variables for the regressions. We use a variety of well-established covariates like the ones used in the studies in Guiso et al. (2002). We control for the wealth and income levels of a household as well as self-assessed risk aversion. In addition, we include standard demographic characteristics of the household reference person such as age, gender, marriage status, number of kids and whether a household is situated in the former GDR. Furthermore, we account for education level and employment status. Finally, we include whether a household has limited access to financing, the expectation towards the economic future and potential health problems. The reasons for including these variables was given in Section 2. In Table 4 we provide the definition of each covariate while Table 5 gives the descriptive statistics for the covariates. From this we can see, for instance, the average household in the sample holds about 40,000 € in financial assets. Financial wealth is very unevenly distributed which is illustrated by that fact that the richest household in the sample owns more than three million Euros in monetary assets - more than 175 times the median household's financial wealth. Furthermore, there is a large difference between the financial wealth of the average household and its overall and net wealth level. The latter two measures of a household's financial well-being are on average much larger than financial wealth (about 179,000 € and 150,000 €, respectively) due to high levels of real estate wealth for many households. Note also that the net household wealth is even negative for some households whose overall debt exceeds their combined asset positions. Moreover, it can be seen that average age in the sample is 52 years, about 60 % in the sample are married and 27 % of the sample households are situated in East Germany. Furthermore 12 % hold a college degree and 67 % are willing to take at least some risk in financial matters.

*Insert Table 4 and Table 5 about here.*

Before we turn to the estimation results it is worthwhile to explore the relationship between the portfolio composition and the main explanatory variables descriptively by plotting share levels against different values of the determinants. Of course these pattern cannot be interpreted as causal as they are likely distorted by other factors that are correlated with both the variable and the makeup of the portfolio. Nonetheless, this visual inspection can give us a good first impression of the relationships in the data.

*Insert Figure 2 about here.*

Figure 2 plots the average shares for each risk class for each decile of household financial wealth.[23] Evidently, the average portfolio composition becomes riskier the higher one gets in the distribution of financial wealth. Notably, this substitution seems to be happening directly between "clearly safe" and "clearly risky" assets while the fraction invested in "fairly safe" assets remains stable. The change is quite pronounced - households in the lowest decile hold on average only 1 % of their financial wealth in equities but 43 % in savings accounts. In contrast, the top ten percent of the households with respect to financial wealth hold an equal fraction in both assets (25 %).

[23] One could also use other measures such as gross total wealth or net wealth but the overall pattern is much the same. Additionally, financial wealth is most directly linked to the shares of financial wealth and thus, in our opinion, the most plausible option.

From the impression in Figure 2 we conclude that using a log transformation as is often done in empirical research on risky asset shares is probably not appropriate here. Instead, we allow for a more flexible relationship between financial wealth and each share variable by including dummies for the quartiles of the distribution of wealth as in Alessie et al. (2002) and Banks and Tanner (2002).[24]

*Insert Figure 3 about here.*

Looking at the evolution of the portfolios over income deciles in Figure 3 it can be seen that the pattern is similar to that observed for financial wealth. As before, the shares of ”clearly risky” assets and ”clearly safe” assets start of at very different levels and then converge over the course of the income distribution. Unlike for wealth, the proportion allotted to”fairly safe” assets does increase with household income. Due to the similar pattern we use the same modeling strategy for household income as for wealth.

*Insert Figure 4 about here.*

The change of portfolio shares across age is given in Figure 4. We can see that the share invested in risky financial assets is largely constant over age groups. This is not necessarily what one might expect given the shorter investment horizon of older participants but is in line with the findings of Wachter and Yogo (2010). On the other hand one can observe considerable differences for the remainder of the portfolio for different ages. The share invested in ”fairly safe” assets exhibits a broad bulge for the ages 30 to 50 and a subsequent sharp decline. Conversely, we observe low fractions of money deposited in savings accounts for prime age adults and an increasing proportion for senior households. This pattern is not unexpected keeping in mind that many investment products in the ”fairly safe” category feature a provident nature. These tend to be acquired during one’s prime and then used up once one approaches retirement age. Savings accounts on the other hand are relatively liquid which makes them more attractive for senior citizens. To capture the obvious pattern for the share invested in ”clearly safe” and ”fairly

[24]One could also use other measures such as gross total wealth or net wealth but the overall pattern is much the same. Additionally, financial wealth is most directly linked to the shares of financial wealth and thus, in our opinion, the most plausible option. We also find that using finer quantiles such as deciles does not add to the explanatory power of the model.

safe assets" we introduce age with quadratic terms in our model even though this is probably less suitable for the share of "clearly risky" assets.

*Insert Figure 5 about here.*

Another important determinant of the portfolio risk structure is the tolerance toward financial risk. The SAVE questionnaire asks participants to rate their own appetite for financial risks on a scale from 0 to 10. In Figure 5 we plot the average shares for each value of the self-assessed financial risk attitude.[25] Households with lower levels of appetite for financial risk hold on average lower portions of their wealth in equities and higher shares of safe assets while the share invested in "fairly safe" assets largely remains unchanged. However, this relationship holds only up to a score of 7 points for the risk tolerance. At this point the shares for "clearly safe" and "clearly risky" assets are about the same (about 25 % each). After this point the gap between the two shares opens up again until the ratio for observations with a risk attitude equal to 10 is nearly the same as for those who score 0 points. This pattern clearly makes no sense if one assumes that the stated preference is equal to the actual risk attitude. A potential explanation for this phenomena is indicated by the distribution of the risk aversion indicator. The sample population is highly risk averse when it comes to financial investments: 33 % state that they would not take any financial risk at all and 66 % state a score of no more than 2. Only 5 % of the sample feature a score of 8 or higher which results in a very small sample size for this group. It could also be argued that the self-assessment of these 5 % does not accurately reflect their true risk attitude. In any case, we use a dummy for a score higher than 0 in the regressions as we think it gives more reliable information regarding the true risk aversion of the sample.

*Insert Figure 6 about here.*

Finally, before we turn to the regression results we examine the evolution of the portfolio structure over time. As mentioned before one could expect a change in the composition of assets over the course of the financial crisis - either trough active reallocation or due to losses realized on the risky part of the portfolio. However, as one can see in Figure 6, we do not find evidence for this

[25] Here a 0 indicates no willingness to take on risk in financial matters. 10 indicates maximum willingness to take on risk.

in the descriptive statistics. Rather, the share levels remain relative constant over time. One can reason that due to the low level of equity holdings German household where less exposed to financial losses during the financial crisis to begin with.

## 5 Empirical Results

In the following, we present the regression results for the multivariate regression models introduced in Section 3. As noted there, the estimated coefficients in these models do not give us much information about the effect of a variable on the means of shares. Thus, we mainly report the computed average partial effects as described in Section 3.[26] Note that these marginal effects do sum up to zero over equations as they should when covariates are associated with a reallocation across assets in a portfolio. The covariates used in each model are those introduced in Section 2 and described in Subsection 4.5. Year dummies are included in each regression but the corresponding effects are not reported to conserve space.

### 5.1 Pooled Model

Table 6 shows the marginal effects for the fractional multinomial logit model pooled over years. The reported z-statistics and p-values are based on fully robust standard errors clustered at the household level to account for serial correlation due to unobserved heterogeneity. The effects for most covariates are largely as expected either from theoretical consideration, prior empirical work, or what one might anticipate intuitively.

*Insert Table 6 about here.*

The effect of age on investor behavior is highly significant for "clearly safe" and "fairly safe" assets. The age of an investor plays an inverse role in the share of wealth invested in savings accounts and assets in the middle risk class. Age in levels negatively affects the share invested in "clearly safe" assets while squared age has a positive effect. This positive effect prevails up to a turning point around the age of 42 years. For the share of wealth invested in "fairly safe"

[26]The full regression results are available on request.

assets the turning point is the same but the relationship is reversed - for older investors the negative effect dominates. These findings correspond to the shape of the age curves for these two asset classes that can be seen in Figure 4. As hypothesized in Subsection 4.5 this pattern probably reflects the provision nature of many assets in the "fairly safe" category which becomes less important as one approaches retirement. For the risky asset category we see much smaller effects. We observe a positive effect of higher age for the share invested in equities in line with the hypothesis by King and Leape (1998).

Whether a household is situated in the former GDR does not seem to play a role in the portfolio formation. The marginal effects for each share are very small and distinctively not significant. This is an interesting finding as one might expect different investment patterns between the two regions due to differences in the socialization of households. Either enough Germans have moved from one part of the country to the other to diminish these differences or the differences might not have been very pronounced to begin with.

Gender exhibits the anticipated effect on the share of risky assets. We find that households with a male head allocate on average 1.6 % more of their funds to equities even as we control for risk-attitudes and expectations of investors.[27] For the remainder of the portfolio we do not see significant effects.

Similar outcomes are observed with respect to the effect of marital status. The share invested in savings accounts or the middle category is largely unaffected by the marriage status while it is associated with a 1.7 % lower share invested in equities.

The number of kids does influence the allocation between "clearly safe" and "fairly safe" assets but does not have an effect on the expected share invested in "clearly risky" assets. Here each additional child in the household is associated with a 3.5 % lower share of money held in savings accounts and an almost equal rise in the share of money held in life insurance policies and the like. The provision character of the middle category could be the determining factor of this effect. These findings are largely in line with other studies such as Guiso et al. (2002) or Rosen and Wu (2004).

[27]This finding should, however, not be overstated as we cannot be sure that the person answering the questionnaire is also the one making the financial decisions in all households.

For education we observe the well established finding (Campbell, 2006) that a higher level of education is associated with a higher share of wealth invested in stocks and similarly risky financial assets. We find that on average a person with a college degree holds almost 5 % more of her financial wealth in risky assets compared to an otherwise comparable individual in the lowest education group. The marginal effects for both education dummies are highly significant. Investors with the highest level of education hold a significantly lower share of their financial wealth in the middle category. Meaning, on average they hold 3.8 % less shares in these assets compared to a household at the lowest education level. Presumably this is because their education lets them process information on this topic more easily so that they can take advantage of the potentially higher returns of risky assets. For the safe asset class there is no clear effect.

If a household has at least one member who is working full-time this affects the portfolio structure significantly. Such a household invests on average 3.3 % less in "clearly safe" assets but 5.9 % more in "fairly safe" assets. These findings are quite plausible as a full-time employment offers more security to invest money in longer-term assets rather than to put it into a savings account. It is less plausible though that such households also hold 2.5 % less of their financial wealth in "clearly risky" assets as the labor market status should make it more bearable to endure negative equity shocks.

Retired households hold about 3.2 % less of their financial wealth in building savings contracts and the like and 3.6 % more in savings accounts. Equities are not affected. This probably reflects the shift from long-term illiquid investments (potentially with a provisioning nature) towards more liquid and safer assets at the end of one's working life in accordance with the findings for an investor's age.

The effects of unemployment are less easy to interpret as they imply a higher share of "fairly risky" assets (4.7 %) and a lower share held in savings accounts (-3.8 %) while there is no visible effect on the share invested in stocks. It is not clear how to interpret these effects because it seems more plausible to expect a reallocation towards safer assets.

The results for financial wealth confirm the picture one gets from Figure 2 in Section 4 as well as the findings in the literature. Generally speaking, a household that is situated higher in the distribution of financial wealth holds a significantly higher share of risky assets. At the same time the share of safe assets is reduced by almost the same amount. For instance, a household exhibiting a level of financial wealth that puts it in the upper quartile of the wealth distribution holds on average 15.2 % less of that wealth in savings accounts. Simultaneously such a household invests 13.64 % more in equities, compared to an otherwise comparable household in the lower quartile of the wealth distribution. Both effects are highly statistically significant. This relationship becomes gradually weaker for lower quartiles but remains highly significant. Again, this is in line with the descriptive evidence presented in Subsection 4.5. The ”fairly safe” category, on the other hand, is largely unaffected by the level of financial wealth - the associated marginal effects are much smaller compared to those for the other two categories and never significant.

The household income does not seem to affect the risky asset share in our data. The estimates for the quantiles feature relatively small marginal effects and are not significant at all. This is somewhat surprising because many previous studies find a positive effect of income on the share of risky assets as we have noted in Section 2. However, our findings are robust to different model specifications and measures of household income. Additionally, other studies also do not find an effect of income (Cardak and Wilkins, 2009) or do not even include it in their estimation (Börsch-Supan and Eymann, 2002). Thus, it seems that household income is not nearly as important in this regard as household wealth. For the two other asset classes we find a division between the upper and the lower half of the income distribution with respect to income. On average households in the upper half of the distribution invest about 3.6 % more in ”fairly risky” assets and about 4.5 % less in ”clearly safe” assets. One explanation for this behavior could be that households with higher incomes can more easily afford to make regular payments on defined contribution plans or the like. Another reason could be that these households are better able to afford initial payment for certain investments in this class.

Being a risk taker in financial matters has a major influence on the portfolio composition of a household. As expected a large positive effect is observed for the risky asset share which increases by 5.8 %. This increase in " clearly risky" assets goes hand in hand with a decrease in the share of "clearly safe" assets of -3.6 % as well as "fairly safe" assets (-2.2 %). All effects are highly statistically significant. These results are quite intuitive as more risk tolerant households can be expected to shift their portfolio toward more risky assets and away from safer investments. The same is true for the fact that this shift is stronger for more secure assets. These findings are again in line with previous research as noted in Section 2.

In accordance with the reasoning in Section 2 we find that households that express an optimistic view of the economic future hold more risky portfolios. On average such households hold a 1.2 % higher share of stocks. At the same time we observe a reduced investment in "fairly risky" assets of 1.8 %. However, the latter effect is only barely significant.

A household that has trouble to obtain a loan exhibits a 3.5 % lower share of equities in its portfolio. This is in accordance with the line of argument of Gollier (2002) presented earlier. However, the effects of such a liquidity constraint are even more pronounced for the other two risk classes: being constrained in ones ability to borrow money leads to a 10.4 % lower share invested in savings accounts and an increase of 13.9 % for the share held in "fairly safe" assets. Both effects are highly statistically significant. Why there is such a strong effect for these assets is not apparent and should thus be interpreted with caution.

As Rosen and Wu (2004) note, uncertainty due to bad health should let investors opt for a less risky portfolio composition. Our findings support this argument. Households with members considering their state of health as poor hold on average a 3 % lower share of equities in their portfolio. This shift in the portfolio goes along with a one-to-one decrease (-3.1 %) of the share of money held in the middle category. The share invested in "clearly safe" assets remains unaffected. This result is reasonable as such a household should have an increased interest in insuring itself against uncertain future outcomes which can probably be best achieved by the investment vehicles in the middling group.

Summing up our results we find that most variables in the estimation exhibit the expected effects. Among the strongest determinants of the asset structure are the wealth level of a household as well as self-assessed risk-tolerance, both of which shift the portfolio towards more risky portfolio compositions. These effects are in correspondence with the reasoning in the financial literature which attributes major influence on the portfolio formation to these variables. It is also evident that if one only concentrates on the share invested in stocks, one misses major evolutions in the financial portfolios as the variables affect each share very differently. For some covariates such as the number of kids in a household or whether the head is retired we see a shift between funds held in savings accounts and in building saving contracts and the like while risky assets are unchanged. One would not see these effects in a univariate framework. As the marginal effects naturally sum up in this model specification, it is easy to see how the change in a given variable changes the composition of a portfolio over risk classes. Usually the shift in the structure of portfolio is plausible. For instance, we observe that for risk-tolerant households the share of an asset in the portfolio is the higher the riskier that asset class is. On the other hand one should not overemphasize the substitution pattern - we see for instance that with rising wealth level a one-to-one substitution between savings accounts and equities occurs while the share of life insurances and the like does not change. It could however be that households reallocate funds from "clearly safe" assets to "fairly safe" ones while at the same time shifting money from the middle category to "clearly safe" assets. In this way, it would be possible that inflows and outflows of money into this category roughly cancel each other out.

### 5.2 Random Effects Model

In the following we present the results for the fractional multinomial logit model where we allow for unobserved heterogeneity via maximum simulated likelihood as previously described in Subsection 3.4. The statistical inference of the model is based on bootstrap resampling. Specifically, we compute fully robust cluster-corrected standard errors on the basis of 500 bootstrap samples. We test for the presence of unobserved heterogeneity in the data by computing a likelihood ratio test for a model where the random effects are equal to zero against a model with unrestricted effects. The resulting test statistic of 189.47 corresponds to a p-value of 0. We thus can reject the null hypothesis of no random effects on any conventional significance level.

*Insert Table 7 about here.*

The marginal effects for the random effects fractional multinomial logit model are presented in Table 7. Notably, the marginal effects are very similar in magnitude to those for the pooled model. This is unsurprising as the main difference between the two models is how the error structure is defined. The significance of the marginal effects for the two models is also quite similar. As noted before one would expect smaller standard errors for the random effects models but we observe this only for some of our regressors. This is likely due to the fact that we have also made the random effects model fully robust to any departures from the standard assumptions regarding the error term.

Due to these similarities we do not elaborate on the results of the random effects model here. In any case the random effects model is primarily a stepping stone in order to arrive at the correlated random effects model. As we have noted in Section 3 the main reason for introducing panel data methods is to account for correlations of the time-constant household-specific effects and the covariates. Hence, we rather focus on the CRE model in the next subsection.

### 5.3 Correlated Random Effects Model

Finally, we turn to the results of the correlated random effects framework. We use a Mundlak specification of the CRE model by including the time averages of the covariates as additional variables in the regression as outlined in Section 3. In doing so, one utilizes only the variation within households over time to identify the effects of the explanatory factors which then measure the deviation from their time average. In this framework one cannot include time averages for variables which do not vary over time such as gender. These would be dropped in the regression due to perfect multicollinearity with the original variables. Thus, when controlling for unobserved heterogeneity, only the effects for time-varying regressors can be interpreted meaningfully.[28] For this reason we report in Table 8 only the marginal effects for those variables

[28] A similar reasoning applies to variables that change deterministically with time such as age.

for which we also included their time-averages in the regression. We focus on households which have been observed for at least three years to ensure enough variation over time. However, our results are quite robust to different sub-samples.

*Insert Table 8 about here.*

Looking at Table 8, we find that most effects are no longer significant once we control for potential correlations with unobserved heterogeneity. This is not an uncommon finding for CRE applications as the variation over time is typically much lower than the cross-sectional variation.[29] Another potential reason for the lack of significance of effects is that many variables exhibit very small marginal effects once we control for household-specific characteristics. For instance, the magnitudes of the effects for the number of kids in a household are much closer to zero compared to the previous models. One conclusion that can be drawn from this is that the former results for these variables might be attributable to unobserved differences across households. On the other hand, some variables, like the retirement dummy, exhibit effects that are not much smaller in magnitude compared to the ones before. The effect of retirement on the share of wealth invested in savings accounts is even significant on the 5 % significance level. We find that the marginal effect (3.8 %) is almost identical to the ones found in Subsections 5.1 and 5.2. The effects for the other two categories are negative as before but insignificant. Therefore it is likely that retirement has at least some sort of genuine impact on a household's portfolio.

More importantly, we find that the wealth quartiles and the risk-tolerance indicator exhibit highly significant effects in this model with associated p-values being close to zero. The effect of risk-tolerance on the share of wealth invested in risky assets is still highly significant after we account for household-specific characteristics. It exhibits the anticipated positive sign while its magnitude is somewhat diminished compared to before. We find that on average risk-tolerant households hold 2.2 % more of their financial wealth in equities. Being somewhat tolerant towards financial risk is associated with an about 1 % lower share held in both "clearly safe" and "fairly safe" assets. Therefore, the directions of these effects remain the same while the observed

[29] Table 9 illustrates this by listing the overall, between and within variation of the dependent and independent variables.

reduction in magnitude is comparable to that seen for the "clearly risky" category. However, the effects for these two shares are no longer significant. In spite of this, it looks as though the portfolios of risk-tolerant investors are riskier even after controlling for the idiosyncrasies of these households.

We find that wealth quartiles exhibit by far the strongest effects within the CRE framework which is in accordance with our previous findings. The marginal effects of a higher position in the wealth distribution on the share invested in savings accounts is still negative and even somewhat higher in comparison to the results in Subsections 5.1 and 5.2. This difference in the magnitude of the effects is the stronger the higher quartile: for the second quartile the effects are quite similar while the effect of being in the highest wealth quartile is about 2.5 percentage points higher for the CRE model compared to the RE model. The significance of the effects for the CRE model are even slightly higher than for the previous models. The effects of wealth on the risky asset share continue to be positive and highly significant but their magnitudes are smaller than previously found. As for the "clearly safe" category, this difference in magnitude becomes larger for higher quartiles. Interestingly, one now observes significant positive wealth effects also for the share of "fairly safe" assets. This is in contrast to the results in Subsections 5.1 and 5.2 where we found no such impact. Even more interestingly, these effects are actually larger in magnitude than those observed for the risky asset share. For instance, a household in the highest wealth quartile holds on average an about 10 % higher share of "fairly safe" assets compared to a household at the bottom of the wealth distribution. These findings are also consistent with the reasoning presented in Section 2.

When looking at the effect of the means of wealth quartiles over time (given in Table 10) we find that they are negative for "fairly risky" assets while they are positive for the "clearly risky" category. Thus the effects potentially cancel each other out in the equation for "fairly risky" assets while they reenforce each other for the equity share. This explains the difference to the models which do not include time averages. One way to interpret these numbers is along the line of Carroll (2002), i.e. that the unobserved factors found in richer households are at the same time associated with a preference for riskier portfolios. Thus, at least a part of the positive effect of wealth on the risky asset share can be attributed to these confounding factors. Once we discount the effect of the household-specific characteristics we find that there is still a meaningful shift away from savings accounts associated with wealth. However, this shift is now allocated to the two remaining risk classes with a slightly higher increase for the "fairly risky" asset.

All things considered, we find that the effects of most variables become quite small and insignificant when we account for differences across households via correlated random effects. However, the effects for the level of wealth and the risk-attitude of households are still sensible and significant. This is quite plausible as these factors have been found to be among the most influential determinants of household portfolios. The main results for these variables remain relative stable compared to the findings in Subsections 5.1 and 5.2. Thus, one can be quite confident that the observed relationships are not altogether spurious. However, one cannot rule out that the effect of the other regression is spurios which is an important finding.

## 6 Conclusion

This paper explores the determinants of the risk structure of financial portfolios of German households. To this end we consider the share of financial wealth allocated to three broad risk classes as proposed by former studies (Guiso et al., 2002): "clearly safe" assets (savings accounts), "fairly safe" assets (life insurance, building savings contracts, etc.) and "clearly risky" assets (equities). This approach stands in contrast to many other papers which often consider only the share invested in risky assets. We account for the bounded nature of expected shares in a joint modeling framework by using a fractional multinomial logit model as suggested by Mullahy (2011) and Murteira and Ramalho (2013). Furthermore, we utilize the panel dimension of the SAVE survey to account for unobserved heterogeneity across households. In this way, we control for potential confounding effects on the explanatory variables which might otherwise influence the results. We consider a wide range of different covariates which have been found by previous research to affect the portfolio composition such as demographic characteristics and financial resources of a household.

Our findings suggest that among the most important influencing factors are the level of wealth of a household as well as its tolerance towards financial risk. Both factors are associated with significantly riskier portfolios. Considering ”clearly safe” and ”fairly safe” assets in addition to ”clearly risky” assets allows us to get a more complete picture of household portfolios. For instance, our random effects estimates suggest that portfolio reallocation associated with investor age occurs mainly between the two former asset classes: senior households hold higher shares in savings accounts while prime-age households invest more in assets with provisioning character as found in the middling category. By modeling the asset shares jointly one can easily track portfolio shifts across assets associated with certain explanatory variables. This can be exemplified by the indicator for risk-tolerance for which one can follow precisely the associated shift from more secure to more risky assets.

Once we control for household specific effects via a correlated random effects approach we find that most variables are no longer significant. This is not surprising as this method relies on the variation over time to identify effects which is typically lower than the variation across households. Nevertheless, we still find meaningful and highly significant effects for the level of wealth on all shares as well as for the risk-tolerance on the share of risky assets. As before both factors are associated with more risky portfolios. From this evidence we cautiously conclude that the observed effects for these variables are not mainly attributable to unobserved differences between households but rather represent some genuine relationship. A conclusion regarding the effects of other variables is more difficult. It could be that the effects of these variables are spurios but it could also be the case that their time-variation is to small to meaningfully separate their effects of the unobserved heterogeneity.

As we have outlined in Section 3, fractional response models can accommodate large numbers of corner values and still yield consistent estimation for the conditional mean of shares. Thus, the number of shares at the boundaries in our application can be regarded as unproblematic. Nevertheless, in the spirit of Mullahy (2011) and Murteira and Ramalho (2013) it might still be interesting to employ models that allow for more general features of the conditional share distribution in future research.

## References


Alessie, R., Hochguertel, S., and Van Soest, A. (2004). Ownership of stocks and mutual funds: A panel data analysis. *Review of Economics and Statistics*, 86(3):783–796.

Alessie, R., Hochgürtel, S., and van Soest, A. (2002). Household portfolios in the Netherlands. In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 341–388. The MIT Press.

Ameriks, J. and Zeldes, S. (2004). How do household portfolio shares vary with age? Working paper, Columbia Business School.

Atella, V., Brunetti, M., and Maestas, N. (2012). Household portfolio choices, health status and health care systems: A cross-country analysis based on share. *Journal of Banking & Finance*, 36(5):1320–1335.

Banks, J. and Tanner, S. (2002). Household portfolios in the United Kingdom. In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 219–250. The MIT Press.

Barasinska, N., Schäfer, D., and Stephan, A. (2012). Individual risk attitudes and the composition of financial portfolios: Evidence from german household portfolios. *The Quarterly Review of Economics and Finance*, 52(1):1–14.

Bertaut, C. and Starr-McCluer, M. (2002). Household portfolios in the United States. In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 181–217. The MIT Press.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7):677–693.

Börsch-Supan, A., Coppola, M., Essig, L., Eymann, A., and Schunk, D. (2008). The German SAVE study: Design and results. *meaStudies*.

Börsch-Supan, A., Coppola, M., and Reil-Held, A. (2012). Riester pensions in Germany: design, dynamics, targetting success and crowding-in. Technical report, National Bureau of Economic Research.

Börsch-Supan, A. and Essig, L. (2002). Stockholding in Germany. University of Mannheim (Germany).

Börsch-Supan, A. and Eymann, A. (2002). Household portfolios in Germany. In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 291–340. The MIT Press.

Buis, M. L. (2008). Fmlogit: Stata module fitting a fractional multinomial logit model by quasi maximum likelihood. Statistical Software Components, Boston College Department of Economics.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

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

Cappellari, L. and Jenkins, S. P. (2006). Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *The Stata Journal*, 6(2):156–189.

Cardak, B. A. and Wilkins, R. (2009). The determinants of household risky asset holdings: Australian evidence on background risk and other factors. *Journal of Banking & Finance*, 33(5):850–860.

Carroll, C. D. (2002). Portfolios of the rich. In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 389–429. The MIT Press.

Chamberlain, G. (1984). Panel data. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2 of *Handbook of Econometrics*, chapter 22, pages 1247–1318. Elsevier.

Cocco, J. F. (2005). Portfolio choice in the presence of housing. *Review of Financial studies*, 18(2):535–567.

Cook, D. O., Kieschnick, R., and McCullough, B. D. (2008). Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance*, 15(5):860–867.

Dierkes, M., Klos, A., and Langer, T. (2011). A note on representativeness and household finance. *Economics Letters*, 113(1):62–64.

Drukker, D. M. and Gates, R. (2006). Generating halton sequences using mata. *Stata Journal*, 6(2):214–228.

Eickelpasch, A. and Vogel, A. (2011). Determinants of the export behaviour of german business services companies. *The Service Industries Journal*, 31(4):513–526.

Felton, J., Gibson, B., and Sanbonmatsu, D. M. (2003). Preference for risk in investing as a function of trait optimism and gender. *The journal of behavioral finance*, 4(1):33–40.

Gollier, C. (2002). What does theory have to say about household portfolios? In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 27–54. The MIT Press.

Guiso, L., Haliassos, M., and Jappelli, T. (2002). *Household portfolios*. The MIT Press.

Guiso, L., Jappelli, T., and Terlizzese, D. (1996). Income risk, borrowing constraints, and portfolio choice. *The American Economic Review*, 86(1):158–172.

Guiso, L. and Paiella, M. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic Association*, 6(6):1109–1150.

Guiso, L. and Sodini, P. (2013). Household finance. an emerging field. In Constantinides, G. M., Harris, M., and Stulz, R. M., editors, *Handbook of the Economics of Finance*. Elsevier.

Haan, P. and Uhlendorff, A. (2006). Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *Stata Journal*, 6(2):229–245.

Haliassos, M. and Bertaut, C. (1995). Why do so few hold stocks? *The Economic Journal*, 105(432):1110–1129.

Halko, M.-L., Kaustia, M., and Alanko, E. (2012). The gender effect in risky asset holdings. *Journal of Economic Behavior & Organization*, 83(1):66–81.

Heaton, J. and Lucas, D. (2000a). Portfolio choice and asset prices: The importance of entrepreneurial risk. *The Journal of Finance*, 55(3):1163–1198.

Heaton, J. and Lucas, D. (2000b). Portfolio choice in the presence of background risk. *The Economic Journal*, 110(460):1–26.

Hinz, R. P., McCarthy, D. D., and Turner, J. A. (1997). Are women conservative investors? gender differences in participant-directed pension investments. *Positioning pensions for the twenty-first century*, 91:103.

Hole, A. (2007). Estimating mixed logit models using maximum simulated likelihood. *Stata Journal*, 7(3):388–401.

King, M. A. and Leape, J. I. (1998). Wealth and portfolio composition: Theory and evidence. *Journal of Public Economics*, 69(2):155–193.

Koch, S. F. (2010). Fractional multinomial response models with an application to expenditure shares. Technical report, University of Pretoria, Department of Economics.

Mehra, R. and Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15:145–161.

Miniaci, R. and Weber, G. (2002). Econometric issues in the estimation of household portfolio models. In Guiso, L., Haliassos, M., and Jappelli, T., editors, *Household portfolios*, pages 143–178. The MIT Press.

Mullahy, J. (2011). Multivariate fractional regression estimation of econometric share models. Working Paper 16354, University of Wisconsin-Madison.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Murteira, J. M. and Ramalho, J. J. (2013). Regression analysis of multivariate fractional data. *forthcoming in Econometric Reviews*.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–32.

Papke, L. E. and Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1-2):121–133.

Poterba, J. M. and Samwick, A. A. (2003). Taxation and household portfolio composition: Us evidence from the 1980s and 1990s. *Journal of Public Economics*, 87(1):5–38.

Puri, M. and Robinson, D. T. (2007). Optimism and economic choice. *Journal of Financial Economics*, 86(1):71–99.

Ramalho, E. A., Ramalho, J. J., and Murteira, J. M. (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys*, 25(1):19–68.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4):647–657.

Rosen, H. S. and Wu, S. (2004). Portfolio choice and health status. *Journal of Financial Economics*, 72(3):457–484.

Stavrunova, O. and Yerokhin, O. (2012). Two-part fractional regression model for the demand for risky assets. *Applied Economics*, 44(1):21–26.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Number emetr2 in Online economics textbooks. SUNY-Oswego, Department of Economics.

von Kalckreuth, U., Eisele, M., Blanc, J. L., Schmidt, T., and Zhu, J. (2012). The PHF: A comprehensive panel survey on household finances and wealth in Germany. Discussion Paper 13/2012, Deutsche Bundesbank.

Wachter, J. A. and Yogo, M. (2010). Why do household portfolio shares rise in wealth? *Review of Financial Studies*, 23(11):3929–3965.

Wagner, G. G., Göbel, J., Krause, P., Pischner, R., and Sieber, I. (2008). Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland: Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender). *AStA Wirtschafts-und Sozialstatistisches Archiv*, 2(4):301–328.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, volume 1. The MIT Press, 2 edition.

**All tables and graphs in the appendix are computed using GSOEP survey weights.**

## A Tables

**Table 1:** Participation Rates and Shares of Financial Wealth

| Asset Class | Uncond. Share | Part. Rate | Cond. Share |
|---:|---:|---:|---:|
| Savings Account | 34.9 % | 73.9 % | 47.2 % |
| Building Saving Contract | 13.5 % | 43.6 % | 30.9 % |
| Life Insurance | 19.5 % | 40.8 % | 47.8 % |
| Bonds | 3.8 % | 12.0 % | 31.1 % |
| Private Pension | 5.4 % | 16.5 % | 32.8 % |
| Company Pension | 6.8 % | 22.6 % | 30.1 % |
| Riester Pension | 5.1 % | 23.1 % | 22.1 % |
| Stocks | 9.7 % | 28.8 % | 33.2 % |
| Other Equities | 1.4 % | 4.2 % | 30.6 % |

**Table 2:** Classification of Financial Assets by Risk

| Safe | Fairly Safe | Risky |
|---|---|---|
| Savings accounts, money market accounts or fixed deposit accounts | Building savings contracts<br>Life insurance<br>Fixed income securities<br>Private pension shemes<br>Occupational pension shemes<br>Riester pension plans | Equities like directly held stocks, equity funds, real estate funds or other funds<br>Other equities such as hedge funds and financial innovations |

**Table 3:** Shares of Financial Wealth by Risk Class

| Category | Mean | Median | 0 | 1 | (0,1) |
|---|---:|---:|---:|---:|---:|
| Safe | 34.9 % | 19.4 % | 26.1 % | 15.5 % | 58.4 % |
| Fairly Safe | 54.0 % | 59.8 % | 20.1 % | 19.3 % | 60.6 % |
| Risky | 11.1 % | 0 % | 69.2 % | 1.3 % | 29.5 % |

**Table 4:** Regressor Description

| Variable | Description |
|---|---|
| Year # | Dummy for observation in year # |
| Wealth Quartile | Quartiles for household financial wealth |
| Income Quartile | Quartiles for monthly net household income |
| Age | Age of respondent |
| Male | Dummy for gender of respondent |
| Married | Dummy for married couple |
| Number of Kids | Number of kids in household |
| East | Dummy variable for household living in GDR |
| Fulltime | Dummy for respondent working fulltime |
| Unemployed | Dummy for respondent being unemployed |
| Retired | Dummy for retired respondent |
| High Education | College Degree |
| Low Education | No more than Hauptschule and vocational degree |
| Risk Taker | Dummy variable if respondent is not totally risk averse in financial matters (Riskattitude > 0 on a scale from 0 to 10) |
| Positive Outlook | Expectations for economic situation of household and Germany as a whole higher than average |
| Liquidity Constraint | Problems obtaining a loan in the past |
| Health Problems | Bad assessment of own health (Healthsituation<3 on a scale from 0 to 10) |

**Table 5:** Summary Statistics

| **Variable** | **Mean** | **Std. Dev.** | **Min.** | **Max.** |
|---|---:|---:|---:|---:|
| Wealth | 178,543.32 € | 314,451.19 € | 100 € | 11,435,000 € |
| Netwealth | 149,798.01 € | 302,572.71 € | -337,756 € | 11.245,000 € |
| Financial Wealth | 40,077.03 € | 88,366.30 € | 100 € | 3,215,000 € |
| Income | 2,252.38 € | 1,556.37 € | 100 € | 40,000 € |
| Age | 51.51 | 16.47 | 18 | 98 |
| Male | 0.48 | 0.50 | 0 | 1 |
| Married | 0.59 | 0.49 | 0 | 1 |
| Number of Kids | 0.63 | 0.95 | 0 | 8 |
| East | 0.27 | 0.45 | 0 | 1 |
| Fulltime | 0.36 | 0.48 | 0 | 1 |
| Unemployed | 0.07 | 0.25 | 0 | 1 |
| Retired | 0.34 | 0.48 | 0 | 1 |
| High Education. | 0.12 | 0.33 | 0 | 1 |
| Low Education | 0.36 | 0.48 | 0 | 1 |
| Risk Taker | 0.67 | 0.48 | 0 | 1 |
| Attitude toward Risk | 2.16 | 2.49 | 0 | 10 |
| Positive Outlook | 0.20 | 0.40 | 0 | 1 |
| Liquidity Constraint | 0.06 | 0.24 | 0 | 1 |
| Health Problems | 0.08 | 0.28 | 0 | 1 |
| Observations | 11056 | | | |

**Table 6:** Marginal Effects for Pooled Model

| Variables | Safe | Fairly Safe | Risky |
|---|---|---|---|
| Age | -0.018*** | 0.022*** | -0.003** |
| | (-8.93) | (9.75) | (-2.48) |
| $Age^2$ | 2.2e-04*** | -2.6e-04*** | 3.9e-05*** |
| | (11.07) | (-11.63) | (2.90) |
| Male | -0.005 | -0.012 | 0.016** |
| | (-0.44) | (-1.04) | (2.32) |
| Married | 0.003 | 0.014 | -0.017** |
| | (0.24) | (1.22) | (-2.27) |
| Number of Kids | -0.035*** | 0.034*** | 0.002 |
| | (-5.67) | (5.48) | (0.47) |
| East | 0.005 | -0.012 | 0.007 |
| | (0.46) | (-0.99) | (0.81) |
| High Education | 0.012 | -0.038** | 0.025*** |
| | (0.81) | (-2.43) | (2.99) |
| Low Education | 0.017* | 0.006 | -0.023*** |
| | (1.69) | (0.51) | (-2.98) |
| Working Fulltime | -0.033** | 0.059*** | -0.025*** |
| | (-2.53) | (4.54) | (-3.00) |
| Unemployed | -0.038* | 0.047** | -0.009 |
| | (-1.93) | (2.42) | (-0.62) |
| Retired | 0.036** | -0.032* | -0.004 |
| | (2.35) | (-1.91) | (-0.38) |
| Income 2nd | -0.018 | 0.014 | 0.004 |
| | (-1.48) | (1.09) | (0.44) |
| Income 3rd | -0.042*** | 0.035** | 0.006 |
| | (-3.11) | (2.5) | (0.62) |
| Income 4th | -0.049*** | 0.037** | 0.013 |
| | (-3.23) | (2.33) | (1.18) |
| Wealth 2nd | -0.062*** | 0.011 | 0.051*** |
| | (-4.86) | (0.77) | (4.46) |
| Wealth 3rd | -0.093*** | 0.016 | 0.077*** |
| | (-6.85) | (1.11) | (6.34) |
| Wealth 4th | -0.152*** | 0.016 | 0.136*** |
| | (-10.15) | (1.01) | (10.77) |
| Risk Taker | -0.036*** | -0.022*** | 0.058*** |
| | (-4.51) | (-2.62) | (9.23) |
| Positive Outlook | 0.006 | -0.018* | 0.012** |
| | (0.66) | (-1.88) | (2.09) |
| Liquidity Constraints | -0.104*** | 0.139*** | -0.035** |
| | (-4.8) | (6.24) | (-2.33) |
| Health Problems | -0.001 | 0.031* | -0.03*** |
| | (-0.04) | (1.85) | (-2.69) |
| Year Dummies | Yes | Yes | Yes |
| Log-likelihood: -9474.956 | | Observations: 10,989 | |

This table presents the calculated marginal effects based on the estimation results of the pooled multinomial fractional logit model. Fully robust clustered z-statistics (calculated via delta-method) are in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively.

**Table 7:** Marginal Effects for Random Effects Model

| **Variables** | **Safe** | **Fairly Safe** | **Risky** |
|---|---|---|---|
| Age | -0.017*** | 0.02*** | -0.003* |
| | (-8.97) | (9.33) | (-1.81) |
| $Age^2$ | 2.2e-04*** | -2.4e-04*** | 3.5e-05*** |
| | (10.81) | (-11.02) | (2.1) |
| Male | -0.007 | -0.009 | 0.016** |
| | (-0.56) | (-0.96) | (2.06) |
| Married | 0.003 | 0.013 | -0.015** |
| | (0.23) | (1.42) | (-2.3) |
| Number of Kids | -0.032*** | 0.03*** | 0.002 |
| | (-5.22) | (5.55) | (0.42) |
| East | 0.004 | -0.009 | 0.005 |
| | (0.55) | (-1.01) | (0.57) |
| High Education | 0.012 | -0.038*** | 0.026*** |
| | (0.92) | (-2.69) | (2.69) |
| Low Education | 0.017** | 0.008 | -0.025*** |
| | (1.98) | (0.93) | (-3.43) |
| Working Fulltime | -0.027** | 0.052*** | -0.025*** |
| | (-2.49) | (4.45) | (-2.64) |
| Unemployed | -0.03* | 0.037** | -0.007 |
| | (-1.75) | (2.29) | (-0.57) |
| Retired | 0.036*** | -0.032*** | -0.004 |
| | (3.05) | (-2.78) | (-0.1) |
| Income 2nd | -0.016 | 0.011 | 0.005 |
| | (-1.48) | (1.12) | (0.34) |
| Income 3rd | -0.034*** | 0.026** | 0.008 |
| | (-2.75) | (2.08) | (0.57) |
| Income 4th | -0.042*** | 0.026** | 0.016 |
| | (-3.11) | (1.98) | (1.25) |
| Wealth 2nd | -0.062*** | 0.009 | 0.053*** |
| | (-5.32) | (0.75) | (4.69) |
| Wealth 3rd | -0.099*** | 0.018 | 0.08*** |
| | (-6.83) | (0.77) | (6.61) |
| Wealth 4th | -0.159*** | 0.018 | 0.142*** |
| | (-10.72) | (0.89) | (11.47) |
| Risktaker | -0.036*** | -0.023*** | 0.059*** |
| | (-4.6) | (-2.87) | (8.86) |
| Positive Expectations | 0.004 | -0.016* | 0.012** |
| | (0.26) | (-1.93) | (2.51) |
| Liquidity Constraints | -0.082*** | 0.116*** | -0.034** |
| | (-4) | (5.92) | (-2.32) |
| Healthproblems | 0.006 | 0.024* | -0.03*** |
| | (0.63) | (1.84) | (-2.91) |
| Year Dummies | Yes | Yes | Yes |
| Log-likelihood: -9380.220 | | Observations: 10,989 | |

This table presents the calculated marginal effects based on the estimation results of the multinomial fractional logit model with unobserved heterogeneity. Clustered bootstrap z-statistics based on 500 bootstrap samples are given in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively.

**Table 8:** Marginal Effects for Correlated Random Effects Model

| **Variables** | **Safe** | **Fairly Safe** | **Risky** |
|---|---|---|---|
| Married | -0.015 | 0.017 | -0.003 |
| | (-1.05) | (1.17) | (-0.22) |
| Number of Kids | -0.003 | 0.001 | 0.002 |
| | (-0.26) | (-0.03) | (0.43) |
| East | -0.058 | -0.068 | 0.126 |
| | (-0.37) | (-0.6) | (1.6) |
| High Education | -0.025 | 0.026 | -0.001 |
| | (-1.18) | (1.23) | (-0.17) |
| Low Education | 0.01 | -0.004 | -0.006 |
| | (0.39) | (0.03) | (-0.53) |
| Working Fulltime | 0.004 | 0.021 | -0.024 |
| | (0.21) | (1.16) | (-1.91) |
| Unemployed | -0.002 | -0.001 | 0.003 |
| | (-0.12) | (-0.06) | (0.27) |
| Retired | 0.038** | -0.025 | -0.012 |
| | (2.17) | (-1.54) | (-0.85) |
| Income 2nd | 0.012 | 0.005 | -0.017 |
| | (0.81) | (0.58) | (-1.56) |
| Income 3rd | 0.012 | -0.001 | -0.011 |
| | (0.65) | (0.16) | (-1.07) |
| Income 4th | -0.004 | 0.006 | -0.002 |
| | (-0.39) | (0.5) | (-0.19) |
| Wealth 2nd | -0.068*** | 0.029** | 0.039*** |
| | (-5.89) | (2.55) | (3.16) |
| Wealth 3rd | -0.122*** | 0.074*** | 0.048*** |
| | (-9.38) | (5.58) | (3.5) |
| Wealth 4th | -0.184*** | 0.099*** | 0.085*** |
| | (-12.15) | (6.7) | (5.74) |
| Risktaker | -0.01 | -0.011 | 0.022*** |
| | (-0.99) | (-1.58) | (3.53) |
| Positive Expectations | 0.001 | -0.005 | 0.004 |
| | (0.11) | (-0.48) | (0.6) |
| Liquidity Constraints | -0.005 | 0.022 | -0.017 |
| | (-0.3) | (1.13) | (-1.01) |
| Healthproblems | 0.005 | 0.014 | -0.019* |
| | (0.07) | (1.25) | (-1.95) |
| Log-likelihood: -8424.58 | | Observations: 10,003 | |

This table presents the calculated marginal effects based on the estimation results of the multinomial fractional logit model with correlated random effects. Clustered bootstrap z-statistics based on 500 bootstrap samples are given in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively.

**Table 9:** Overall, Between and Within Variation

| Variable | $\sigma_{overall}$ | $\sigma_{between}$ | $\sigma_{within}$ |
|---|---|---|---|
| Lowrisk | 0.380 | 0.346 | 0.215 |
| Midrisk | 0.391 | 0.351 | 0.222 |
| Highrisk | 0.227 | 0.188 | 0.141 |
| Wealth 1st | 0.416 | 0.368 | 0.272 |
| Wealth 2nd | 0.456 | 0.336 | 0.338 |
| Wealth 3rd | 0.441 | 0.319 | 0.329 |
| Wealth 4th | 0.413 | 0.333 | 0.245 |
| Inc 1st | 0.468 | 0.426 | 0.230 |
| Inc 2nd | 0.428 | 0.348 | 0.280 |
| Inc 3rd | 0.433 | 0.347 | 0.280 |
| Inc 4th | 0.389 | 0.329 | 0.209 |
| Age | 16.473 | 17.039 | 1.459 |
| Male | 0.499 | 0.499 | 0.000 |
| Married | 0.492 | 0.474 | 0.163 |
| Number of Kids | 0.954 | 0.915 | 0.295 |
| East | 0.447 | 0.441 | 0.030 |
| Working Fulltime | 0.479 | 0.443 | 0.180 |
| Unemployed | 0.256 | 0.230 | 0.159 |
| Retired | 0.475 | 0.460 | 0.145 |
| High Education | 0.327 | 0.302 | 0.114 |
| Low Education | 0.479 | 0.455 | 0.182 |
| Risk Taker | 0.484 | 0.376 | 0.332 |
| Positive Outlook | 0.397 | 0.277 | 0.286 |
| Liquidity Constraints | 0.236 | 0.195 | 0.157 |
| Health Problems | 0.278 | 0.237 | 0.184 |

This table presents the standard deviations of the dependent and independent variables used in our main regressions. The overall variation is compared to the variation between and within households to give an impression of the relative variability of the variables over time.

**Table 10:** Marginal Effects for Time Averages

| **Variables** | **Safe** | **Fairly Safe** | **Risky** |
|---|---|---|---|
| Married | 0.012 | 0.005 | -0.017 |
| | (0.55) | (0.19) | (-1.21) |
| Number of Kids | -0.033 *** | 0.025 * | 0.009 |
| | (-2.77) | (1.85) | (1.06) |
| East | 0.040 | 0.034 | -0.074 * |
| | (0.38) | (0.31) | (-1.87) |
| High Education | 0.048 * | -0.076 *** | 0.028 * |
| | (1.70) | (-2.74) | (1.83) |
| Low Education | 0.021 | -0.008 | -0.013 |
| | (1.06) | (-0.38) | (-0.87) |
| Working Fulltime | -0.054 ** | 0.06 *** | -0.006 |
| | (-2.37) | (2.67) | (-0.37) |
| Unemployed | -0.028 | 0.05 | -0.023 |
| | (-0.75) | (1.35) | (-0.87) |
| Retired | -0.016 | -0.003 | 0.021 |
| | (-0.60) | (-0.12) | (1.01) |
| Inc 2nd | -0.031 | 0.011 | 0.020 |
| | (-1.34) | (0.43) | (1.12) |
| Inc 3rd | -0.057 ** | 0.056 ** | 0.001 |
| | (-2.43) | (2.24) | (0.07) |
| Inc 4th | -0.042 | 0.055 * | -0.014 |
| | (-1.54) | (1.91) | (-0.65) |
| Wealth 2nd | 0.005 | -0.017 | 0.012 |
| | (0.18) | (-0.57) | (0.53) |
| Wealth 3rd | 0.052 * | -0.097 *** | 0.045 * |
| | (1.86) | (-3.47) | (1.90) |
| Wealth 4th | 0.043 | -0.122 *** | 0.079 *** |
| | (1.40) | (-4.17) | (3.50) |
| Risk Taker | -0.060 *** | -0.025 | 0.084 *** |
| | (-3.49) | (-1.39) | (5.69) |
| Positive Outlook | 0.006 | -0.023 | 0.017 |
| | (0.28) | (-1.16) | (1.34) |
| Liquidity Constraints | -0.190 *** | 0.212 *** | -0.022 |
| | (-4.57) | (4.84) | (-0.78) |
| Health Problems | 0.007 | 0.02 | -0.026 |
| | (0.24) | (0.60) | (-1.13) |

This table presents the calculated marginal effects for the time averages based on the estimation results of the multinomial fractional logit model with correlated random effects. Clustered bootstrap z-statistics based on 500 bootstrap samples are given in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively.

# B Graphs

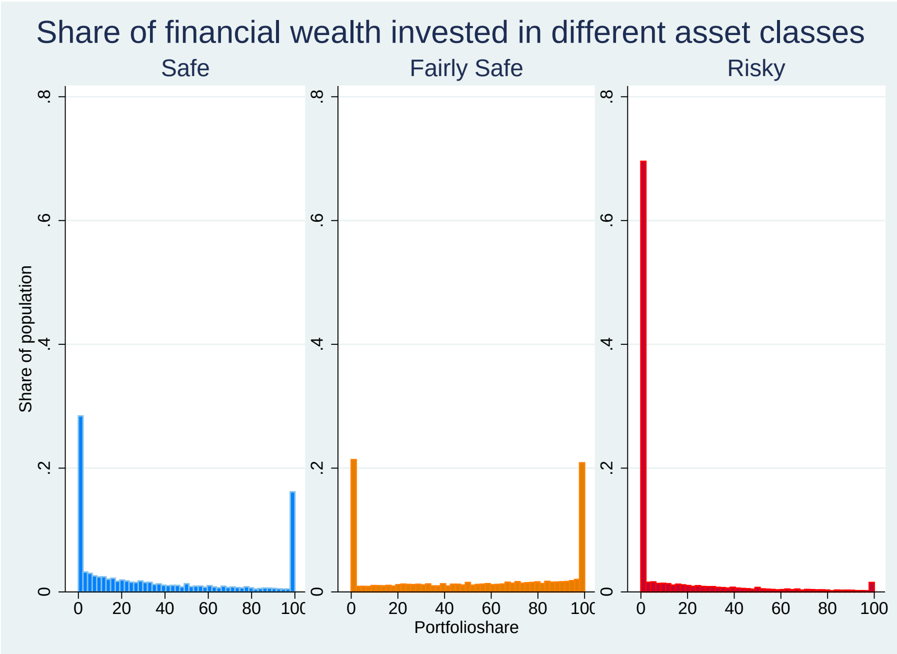


**Figure 1:** Distribution of Shares for Risk Classes

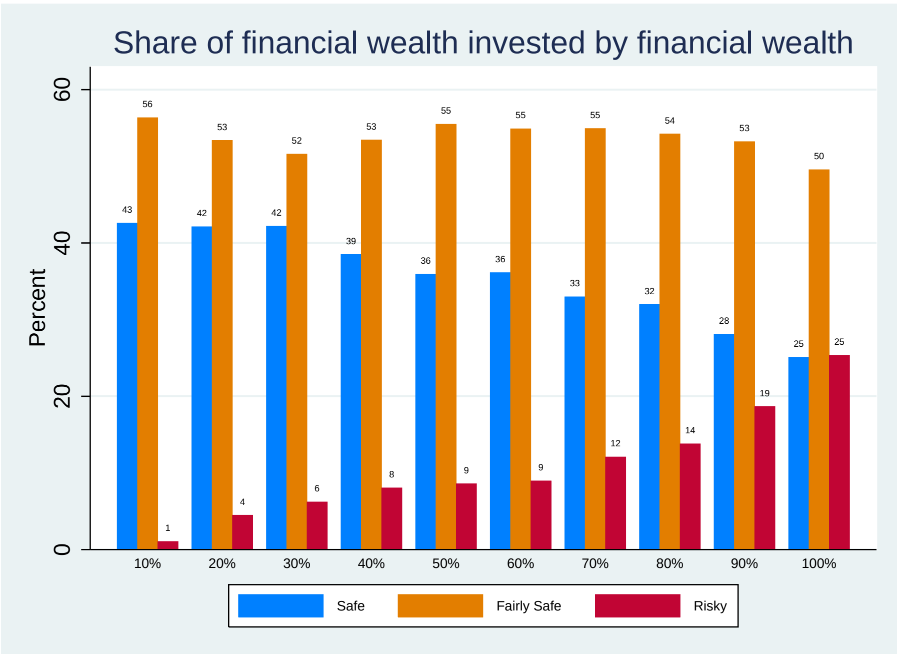


**Figure 2:** Distribution of Shares by Deciles of Financial Wealth

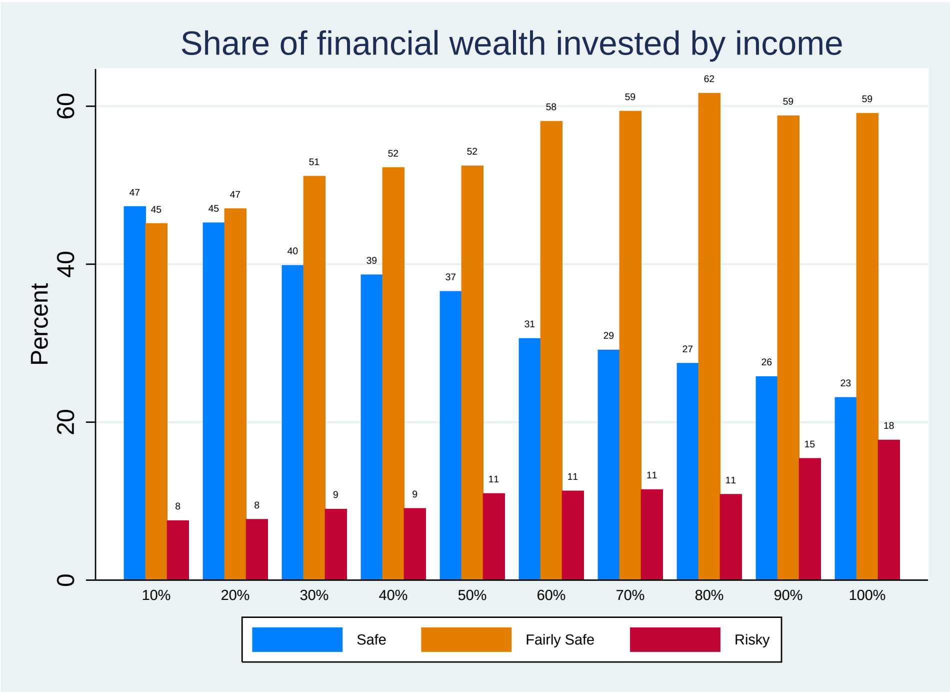


**Figure 3:** Distribution of Shares by Deciles of Income

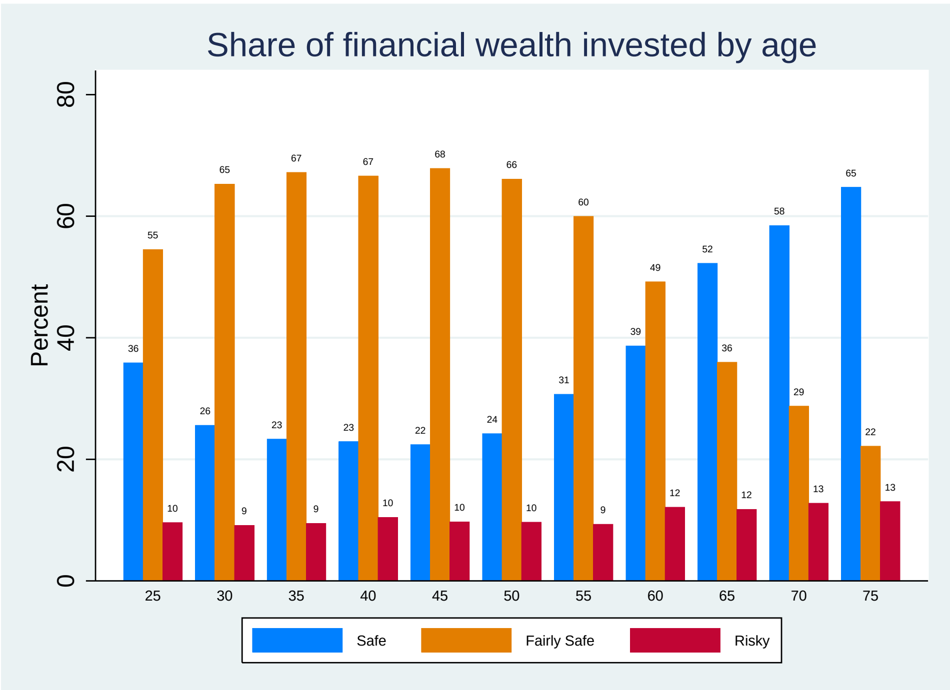


**Figure 4:** Distribution of Shares by Age Groups

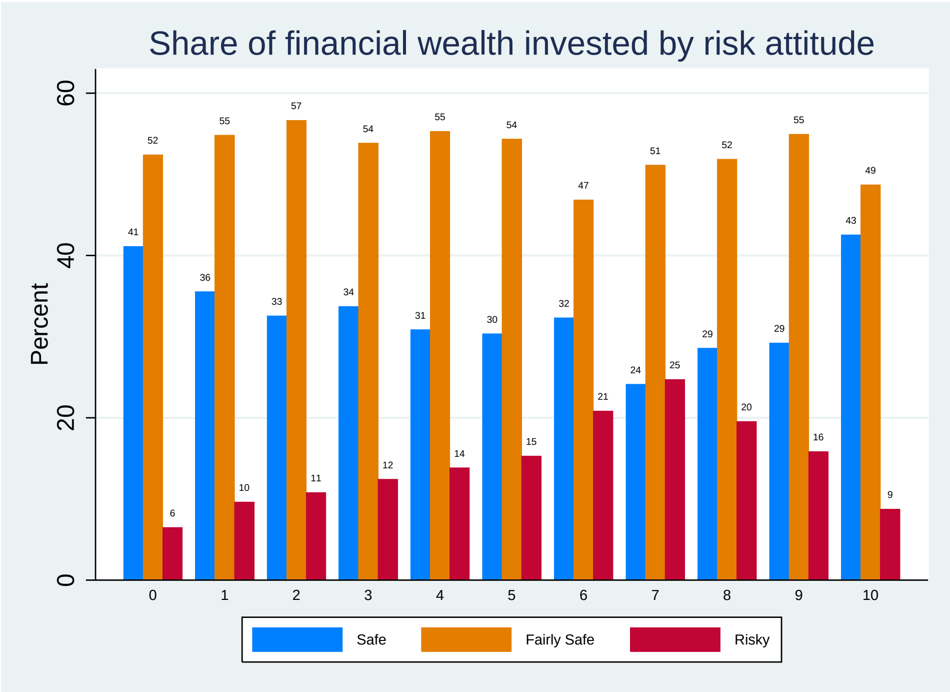


**Figure 5:** Distribution of Shares by Risk Attitude

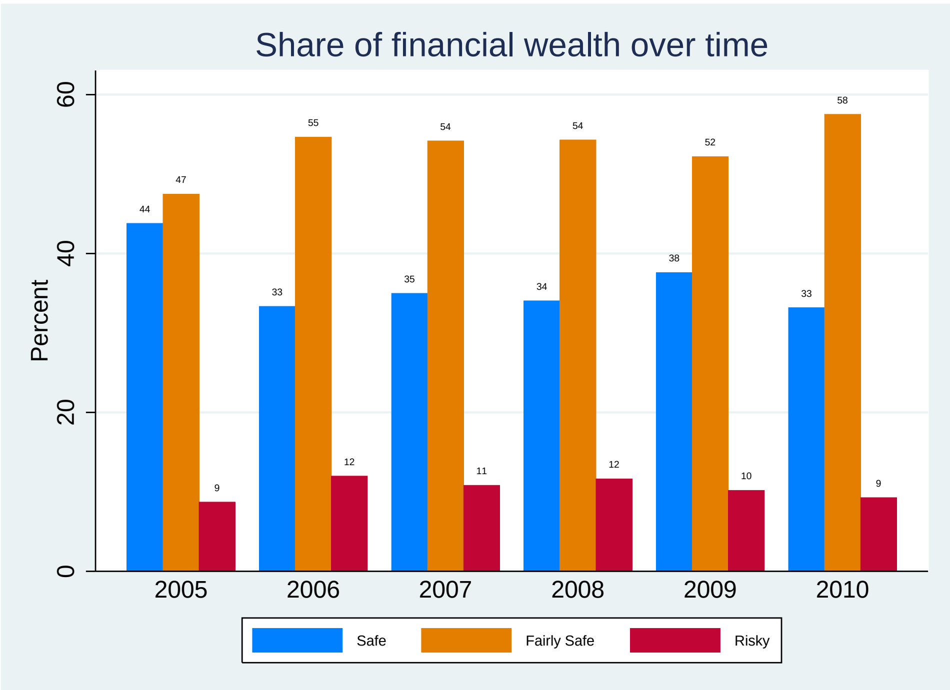


**Figure 6:** Distribution of Shares by Years