Econometric Studies on Personal Wealth and Household Finance in Germany

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Gideon Becker
aus Paderborn

Tübingen
2015

Tag der mündlichen Prüfung:	13. Oktober 2015
Dekan:	Professor Dr. rer. soc. Josef Schmid
Erstberichterstatter:	Professor Dr. rer. pol. Martin Biewen
Zweitberichterstatter:	Professor Dr. rer. pol. Joachim Grammig

Publications

Chapter 2 of this doctoral thesis has been submitted to and accepted for publication by the *Studies in Economics and Finance*:

Becker G. and Dimpfl T. (presumably 2016): "Labor income risk and households' risky asset holdings." *Forthcoming in Studies in Economics and Finance* (Volume 33, Issue 3).

Funding

The research for this doctoral thesis was in parts supported by the German Research Foundation (DFG) under the name "Ökonometrische Analyse der Vermögensstruktur und Vermögensverteilung in Deutschland" (Grant No. BI 767/2-1).

Aaron Wuensch

*30.10.1983 - †09.06.2009

Danksagung

Diese Dissertation ist über mehrere Jahre am Lehrstuhl für Statistik, Ökonometrie und Quantitative Methoden der Eberhard Karls Universität Tübingen enstanden. Auf dem Weg dorthin habe ich die Hilfe einer Vielzahl von Menschen erfahren, ohne die diese Arbeit nicht denkbar gewesen wäre. Vor allem die Mitglieder der beiden Ökonometrie Lehrstühle haben dafür gesorgt, dass ich gerne an diese Zeit zurückdenke. Neben gemeinsamen Seminaren und Diskussionrunden bleiben mir viele gemeinschaftliche Mittagessen und die Weihnachtsfeiern im 3. Stock der Mohlstraße 36 in schöner Erinnerung. An dieser Stelle möchte ich mich dafür herzlich bedanken. Zuallererst bei Martin, der mir die Möglichkeit zu dieser Dissertation gegeben und mich mit seiner ruhigen, freundlichen Art und seinem geballten Fachwissen über alle Maße gefördert hat. Mit deiner tollen Betreuung hast Du diese Arbeit maßgeblich vorangetrieben. Bei Joachim, der überhaupt erst meine Begeisterung für Ökonometrie geweckt und mir das notwendige Rüstzeug für meine spätere Forschung mitgegeben hat. Bei Andos und Markus, die mich zu Beginn meiner Promotionszeit sofort herzlich willkommen geheißen und unter ihre Fittiche genommen haben. In all den Jahren wart ihr mir Vorbild und Ansporn zugleich. Bei Stefanie, mit der ich fast meine gesamte Promotionszeit verbringen, noch einmal die "Uni-Bank" drücken und viel schöne Gespräche führen durfte. Bei Madalina und Matthias, die den Lehrstuhl in die nächste Generation führen und mir meine ehemaligen Kurse und Räumlichkeiten abgenommen haben. Bei Frau Eiting, die als gute Seele des Lehrstuhls zusammen mit Frau Bürger alle administrativen Unannehmlichkeiten von uns ferngehalten hat und dabei immer noch Zeit für ein Schwätzchen fand. Bei allen unseren unglaublich motivierten Hiwis und Tutoren, insbesondere Lukas und Timo. Ihr habt den Lehrstuhlbetrieb mit am Laufen gehalten und bereichert. Bei meinem Koautor und Flurnachbar Thomas, von dessen positiver Einstellung und methodischer Arbeitsweise ich mir einiges abschauen konnte. Bei Franzi, Tobi, Eva und Jantje für viele gemeinsam verbrachte Bachelorseminare und Kuchenrunden. Bei Luis für R Lektionen und Risikoanalysen.

Grenzenlose Dankbarkeit empfinde ich für meine Familie, die mich immer dazu ermutigt hat meinen Weg zu gehen und mich in meinem akademischen Werdegang bedingungslos unterstützt hat. Zu guter Letzt möchte ich mich bei der Person bedanken, ohne deren liebevollen Beistand ich diese Arbeit niemals beendet hätte - Stefanie Du warst in all der Zeit mein Fels in der Brandung und zugleich mein schärfster Lektor.

Gideon Becker, Mainz im April 2015

Contents

List of Figures

List of Tables

Chapter 1

Dissertation Introduction

Traditionally, financial economics has focused on asset pricing and corporate finance and was less concerned with the issue of how individuals use financial instruments to increase their well-being. Over the past twenty years, however, an increasing number of financial economists have turned their attention toward studying the financial situation of private households. In his presidential address to the American Financial Association in 2006, John Campbell emphasized the growing importance of this area of research by assigning this thriving field its own name: household finance (Campbell, 2006). Household finance is foremost concerned with the specific features of households which distinguish their financial decision making from that of corporations or other financial actors. Researchers often focus on individual behavior that is not easily reconciled with the predictions of standard financial theory. The best known example of this is the so-called "stock market participation puzzle", i.e. the finding by Haliassos and Bertaut (1995) and others that only a minority of individual investors own stocks even though standard financial theory suggests that everyone should hold at least some risky assets. Another such peculiarity is the recognition that financial portfolios of most households are poorly diversified (Barasinska, Schäfer and Stephan, 2012). All this raises the question whether investors behave irrationally or if there are important household characteristics that explain this behavior. The field has seen intense research in this area with researchers scrutinizing the importance of factors such as human capital (Campbell, 2006), risk aversion (Guiso and Paiella, 2008), a lack of financial literacy (van Rooij, Lusardi and Alessie, 2011), health risks (Rosen and Wu, 2004), substantial levels of illiquid housing wealth (Cocco, 2005) or generally unavoidable background risks (Heaton and Lucas, 2000b; Guiso and Paiella, 2008). The field has also been influenced by the findings of behavioral finance which emphasize the importance of trust in markets (Guiso, Sapienza and Zingales, 2008) or the influence of social interactions (Hong, Kubik and Stein, 2004) in this matter.

As Guiso and Sodini (2013) note, one reason for the rise of household finance is the increasing complexity of financial matters for the typical household. A shift away from pay-as-you-go pension schemes toward self-funded pensions in many western countries have made investment decisions more meaningful for many households. At the same time, investments that were previously only available through intermediaries have become directly accessible to retail investors through financial innovations such as exchange traded funds (ETF) or crowdfunding platforms. Therefore, investment decisions play an ever more important role in the lives of many individuals as these have to take financial affairs increasingly into their own hands.

Another cause for the increasing interest in the investment behavior of private households is the financial crisis of 2007-08 which gave rise to the Great Recession and the European debt crisis. The financial crisis originated primarily in the USA where a massive real estate bubble led many individuals to believe they could safely borrow against their supposedly ever increasing home property values. Combined with loose credit standards, this behavior resulted in a large number of subprime-mortgages and sharply increasing household debt (Wolff, 2012). These over-leveraged households were in turn left vulnerable to a drop in housing prices. When the housing bubble finally burst, the financial situation of many private households suffered tremendously. Ackerman, Fries and Windle (2012), for example, find that the median net worth of US households fell by almost 40 % from 2007 to 2010 - mainly due to falling housing and stock prices. Thus, on the one hand households contributed to the rise of the financial crisis and were, on the other hand, massively affected by it.

Another trend in recent years has been a revived interest in the distribution of private wealth. This discussion has received considerable momentum from the influential book by Piketty (2014) which has become an instant classic. In his book, Piketty (2014) demonstrates empirically that wealth inequality in many rich countries has soared considerably since the 1980's and is now back to levels previously seen in the roaring twenties. The book has sparked a debate about the legitimacy of this claim as well as its potential causes. The financial crisis has fostered renewed interest in this topic as well. It is natural to ask what effect the sharp drop in housing and stock prices has on the distribution of wealth. Wolff (2012), for example, finds that wealth inequality in the United States has increased drastically from 2007 to 2010. In this respect, an increasing number of studies examine the distribution of wealth in a counterfactual decomposition framework. For instance, Bilias, Georgarakos and Haliassos (2013) analyze the influence of equity culture on wealth inequality in the United States by means of counterfactual distributions.

Empirical household finance blends the theoretical models used in traditional finance research with the microeconometric toolkit which is more common in fields such as applied labor market economics. One main theme of this doctoral thesis is the importance to control for unobserved household characteristics, as failing to do so is likely to yield spurious relationships. To give an example, it is a well-known fact that wealthier households are more likely to hold risky assets such as stocks or business assets (Wachter and Yogo, 2010). This empirical finding could represent a causal relationship, i.e. wealthy households can tolerate the potential losses associated with such assets more easily due to the high shock absorbing capacity of their wealth. Thus, they might be able to take advantage of the higher returns of risky assets more readily compared to their less well-off compatriots. However, it is far from clear that this relationship is indeed a causal one. It might simply arise due to the inability of the researcher to adequately measure certain characteristics of households which are related to both the level of wealth and the propensity to hold risky assets. It is entirely plausible to assume, as Carroll (2002) does, that more risk-tolerant individuals are both more likely to hold risky assets and to obtain a relative high level of wealth because they are generally more willing to seize opportunities that offer themselves. Consequently, such individuals would end up with both higher levels of wealth and money invested in risky assets. Similar examples, which illustrate the need to control for unobserved factors, abound when it comes to analyzing the investment decisions of private investors.

An important prerequisite for quality research in this context is the availability of suitable micro-data sets, as Campbell (2006) and Guiso and Sodini (2013) emphasize. Household surveys have long avoided this sensitive topic since asking questions regarding the asset situation of participants makes it more likely that these will refuse to answer or drop out all together. However, as the growing need for such data has become more apparent, more household surveys have begun to ask for wealth related items or were specifically designed for this purpose. Moreover, many surveys on household wealth now offer a panel design, i.e. they track the same individuals and households over time. Such a set-up offers great potential for the applied econometrician. For one, it enables the researcher to study the dynamics of a household's financial situation, such as the evolution of its property assets over time. More importantly, this data design makes it possible to disentangle spurious relationships such as the one described above. For this reason panel data methods are likely to become more prevalent in studies on household finance.

In this doctoral thesis I empirically scrutinize various aspects of the financial situation of private households and individuals in Germany by applying microeconometric methods such as non-linear panel data methods and counterfactual decomposition techniques. Germany is a particularly interesting case in the context of household finance for several reasons: As the largest European economy with one of the highest savings rates in the world, Germany exhibits high levels of private wealth. At the same time, this wealth is distributed very unequally compared to most other developed countries. Moreover, the participation of households in many asset classes is comparatively low. Germany has, for instance, the lowest rate of home ownership among countries in the euro area (Eurosystem Household Finance and Consumption Network, 2013). Generally, Germans tend to be quite risk averse in financial matters. They shun financial assets such as stock and rather opt for safe investments such as deposit accounts or life insurance policies. In the following, I sketch the difficulties faced by German retail investors in the current financial environment to motivate the significance of the subsequent studies in this dissertation.

One important aspect in this regard is pension provision. Traditionally, the main building block of old-age provision for most Germans is a public pay-as-you-go pension scheme. In such an inter-generation contract current contributors fund the pension of current retirees. The system was not designed as a mere basic pension but intended to maintain the standard of living. However, an increasingly unfavorable demographic situation in Germany has undermined the foundation of this statutory pension insurance scheme. Germany has a rapidly aging population with one of the lowest birth rates in the world. Thus, fewer and fewer prime age earners have to finance ever more retirees with their contributions. Starting around the turn of the millennium, policy makers therefore began to reduce benefits and raise the statutory retirement age in order to prevent contribution rates getting out of hand. To close the resulting pension gap, legislators decided to make the system rely more on supplemental private pension schemes, in effect passing on longevity risks to the individual. In this context, the introduction of the so-called Riester pension, a type of heavily subsidized funded pension plan, was supposed to incentivize more people to assume responsibility for their old-age provision themselves.[1] These reforms have made it increasingly crucial for individuals to choose appropriate financial instruments in order to ensure their long term financial well-being.

More recently, the European Central Bank (ECB) has lowered interest rates towards the zero lower bound to encourage growth in the stagnating euro zone and prevent a slip into Japan-style deflation. This expansive monetary policy adopted by ECB in the wake of the European debt crisis is bound to have considerable consequences for private investors in Germany. As of 2015 the central bank has even launched a campaign of quantitative easing, i.e. the buying-up of financial assets on the open market. This led to a situation where safe financial investments, such as savings accounts, yield no or even negative nominal interest rates. The changes in the financial environment force individual investors to question their investment strategies. This applies in particular to two financial instruments which are quite popular with German investors: life insurance policies and building loan contracts. Even though these instruments are not intended for pure capital investment, they are widely used for this purpose in Germany due to their alleged security. Yet, the current environment has exposed the vulnerability of these products to prolonged periods of low interest rates. In the face of the adverse interest rate environment,

[1] For a general overview of the pension system in Germany and its recent reforms see Borsch-Supan, Bucher-Koenen, Coppola and Lamla (2014).

issuers are lowering guaranteed interest rates and terminate long-standing contracts that still exhibit high fixed interest rates.

At the same time, rock bottom interest rates on property loans have awoken the interest of Germans in real estate investments. This is noteworthy as a high share of tenants which enjoy a strong legal position has meant that Germany has been spared from the real estate bubble which has afflicted other western countries. Yet, now the purchase of home equity appears more affordable to many households and might eventually lead to higher levels of household leverage. The flight to real assets, or "Betongold" as it is called in Germany, is reinforced by the lack of promising investment alternatives in the current interest environment. This has led the Deutsche Bundesbank (2014a) to express its concern regarding increasing overvaluation of residential property in Germany's main metropolitan areas.

The phenomenon of low stock market participation in Germany is best illustrated by reference to the DAX, Germany's key stock index containing 30 major blue-chip companies. During the late 1990s there was a short period of increased interest in the stock market by the German public. On the one side there were several IPO's (initial public offerings) of recently privatized former state-owned enterprises that attracted significant publicity. This phenomenon was epitomized by the Deutsche Telekom which emerged from the former state-owned Deutsche Bundespost and whose stock became known as "Volksaktie" or people's share. On the other side the internet based new economy promised unprecedented growth potential and a subsequent participation in profits by retail investors. During that time the DAX rose from less than 4,000 base points in October 1998 to an all-time high of more than 8,000 base points in March 2000. At the time Börsch-Supan and Essig (2002) expected that stock market participation in Germany would become more common - not least because the pension reforms made additional private provision increasingly important. Yet, after the burst of the dot-com bubble due to over-optimistic growth expectations and sometimes outright fraud, the DAX fell below 2,500 base points in March 2003. The index reached its previous height only four years later in June 2007. By that time many Germans felt vindicated in their believe that the stock market is not a place to prudently invest one's hard earned money but rather a playground for speculators. German equity culture never recovered from this shock: since its peak in 2001 the number of Germans holding stocks either directly or via equity funds has declined steadily. In 2014 only a small minority of less than 8.5 million retail investors participated in the stock market compared to about 12.9 million in 2001 (Deutsches Aktieninstitut, 2015). Thus, about 4.4 million Germans have turned away from the stock market during that time. After the index again lost half its value in the wake of the financial crisis and fell below 4,000 basis points in early 2009, it experienced an almost uninterrupted bull market since then. Propelled by monetary policy of the ECB, the index more than tripled to more than 12,000 basis points in March 2015. However, the German Central Bank found that the share of DAX held by private German investors even decreased during this major stock market boom and was less than 12 % as of 2014 (Deutsche Bundesbank, 2014b). Thus, even though most Germans were not affected by the drop in share prices, they also did not benefit from the following surge in equity prices due to their non-participation. Given the current interest environment and an increasing importance of supplementary old-age provision, this trend bodes ill for the future wealth of many Germans as stocks and other risky securities reliably offer the highest return over the long term to ensure a certain standard of living at retirement.

The following three chapters constitute the core of this doctoral thesis. Chapter 2 examines the potentially adverse effect of labor income risk on the propensity to hold risky assets among the German population by means of binary panel data models. In Chapter 3 I analyze the determinants of the composition of financial portfolios of German households using a fractional multinomial logit model and controlling for unobserved heterogeneity. Chapter 4 presents a counterfactual decomposition analysis for differences in the distribution of per capita net wealth between East and West Germany. In the next paragraphs, I give a short outline of these studies.

Stock Market Participation and Labor Income Risk

The first part of this doctoral thesis, joint work with Dr. Thomas Dimpfl[2], is concerned with the stock market participation of German households. In this study we investigate the determinants of a household's decision on whether to participate in risky asset markets. Standard financial theory states that any investor should hold at least some amount of her wealth in risky financial assets. It is an empirical fact, however, that only a small minority of households directly invests in this type of assets (see the seminal work by Haliassos and Bertaut, 1995). As mentioned above, Germany exhibits one of the lowest shares of private stockholders among developed countries which makes it all the more important to ascertain potential reasons for this non-participation. Recent theoretical research suggests that with increasing labor income risk, the reluctance of households to hold risky financial assets increases. In this regard, the uncertain return to human capital can be understood as a substitute for risky returns of financial capital. We therefore investigate the determinants of a household's decision on whether to invest in risky financial assets using the German Socio-Economic Panel (GSOEP) for the years 2001 to 2010. We propose to measure income risk as the observed variation of household income over a five year period. We find that indeed higher variation, i.e. higher income risk, reduces the propensity to invest in risky securities even after controlling for other important characteristics such as household income, wealth or socio-demographic make-up. We also find that it is important to include subjective measures of a household's financial situation such as income satisfaction or expectations towards one's future financial situation.

However, in such an cross-sectional framework unobservable household characteristics could bias the results. For instance, households that feature unobserved unfavorable characteristics such as an inability to defer gratification that could be associated with a less stable employment history as well as a reluctance to hold assets that require a longer investment horizon as is usually the case for stocks. We take advantage of the panel dimension of the GSOEP in order to control for such unobserved heterogeneity by means of binary panel data methods. In doing so, the impact

[2]Contact: *University of Tübingen, Faculty of Economics and Social Sciences, Tübingen, Germany; thomas.dimpfl@uni-tuebingen.de*

of realized labor income variation decreases considerably and is no longer significant. We thus conclude that the negative effect of higher labor income volatility is in part driven by unobserved differences in household characteristics.

Multifractional Logit Estimation of Household Portfolio Shares

What determines the risk structure of financial portfolios of German households? The third chapter of this thesis is aimed at answering this question. The existing literature usually focuses on the analysis of a single financial asset class, usually stocks or bonds. By contrast, this study analyzes the composition of a household's financial wealth in its entirety in a joint framework. Specifically, I focus on the determinants of the share of financial wealth invested in three broad risk classes, thus obtaining a more comprehensive picture of a portfolio's risk structure and its determinants.

The data for my empirical analysis stems from the SAVE survey, a representative panel study on some 2,500 private German households established in 2001. The SAVE survey provides detailed information on the financial situation of households, for example, the amount of money invested in a wide range of financial assets such as stocks, bonds or savings accounts.

On the methodological side I contribute to the existing literature by modeling shares via fractional response models as proposed by Papke and Wooldridge (1996). This approach is arguably more appropriate than other widely used models in this context. It seems natural to expand this approach to multiple shares in a joint modeling framework by using a fractional multinomial logit model as suggested by Mullahy (2011) and Murteira and Ramalho (2014). In this fashion, one can incorporate interdependences between risk classes compared to a situation where one investigates each share separately. Finally, I extend the multivariate fractional response model in the spirit of Papke and Wooldridge (2008) via a correlated random effects (CRE) approach. In this manner, I exploit the panel-structure of the SAVE survey in order to control for unobserved time-invariant household characteristics. CRE methods have been widely used in many areas of applied economics. However, to the best of my knowledge, they have not been applied to a household finance context so far.

Overall, most discovered effects are as expected from financial theory. 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 me to get a more complete picture of household portfolios. For instance, 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.

Once I control for household specific effects via correlated random models I find that most variables are no longer significant. This is not surprising as this method relies on the variation over time for identification which is typically lower than the variation across households. Nevertheless, I 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 I conclude that the observed effects of these variables are not mainly attributable to unobserved differences between households but rather represent some genuine relationship.

Counterfactual Analysis of Regional Wealth Differentials in Germany

In chapter four I turn to analyzing the distribution of wealth within Germany - more precisely the wealth gap between East and West Germany. Nearly 25 years after German reunification, vastly different living conditions still remain between these two regions. This is particularly true for the distribution of net wealth which is of special importance for the well-being of individuals. Using the wealth component of the German Socio-Economic Panel for the years 2002, 2007 and 2012, I find that, on average, members of western households exhibit a per capita net worth more than twice as high as their eastern counterparts. This wealth gap remains roughly stable over the observed period of time and is even more pronounced for upper parts of the wealth distribution. Such wealth differentials matter for several reasons. For one, wealth provides utility for individuals, for instance, by acting as a buffer against negative income shocks or by serving as as retirement provision. Moreover, the German constitution explicitly calls for the creation of equivalent living conditions across Germany. It is therefore of general interest to learn where these differences arise from.

In this study, I analyze how much of the gap in per capita net wealth at different parts of the distribution can be attributed to observable factors such as permanent income or socio-demographic characteristics. To this end, I employ the counterfactual decomposition approach by DiNardo, Fortin and Lemieux (1996) which allows me to identify the effect of certain variables on the wealth gap by computing counterfactual distributions. I carry out the decomposition analysis via a multistage reweighting approach. Thus, the population under observation is reweighted in such a way that the distribution of the variables of interest are the same in both regions. In this fashion one obtains, for example, the distribution of wealth in West Germany that would prevail if the distribution of permanent income was like in the East.

I find that for the lower part of the distribution, most of the gap can be attributed to observable wealth determinants while this share is much lower for the upper part. The most important contributing factors in this regard are the lower levels of income still prevailing in East Germany as well as differences in labor market outcomes. Moreover, I observe that the wealth gap for younger Germans is both much smaller and easier explicable by observable factors. For the young, success on the labor market is by far the most important explanatory factor. I also find that home ownership rates differ markedly between the two regions. This is associated with a substantial part of the wealth gap even though differences in housing prices also seem to matter here.

Chapter 2

Stock Market Participation and Labor Income Risk

2.1 Introduction

The way in which households invest their earnings during their working life largely determines their welfare over the life cycle. Therefore, it is of great interest to learn which factors determine how people allocate their funds to different asset classes. Historically, gains from investing in the stock market or in markets for similar financial instruments outperformed more conservative investment options like savings accounts or insurance plans by far. Yet, private investors largely forgo this equity premium as only a minority of households holds any stocks at all. This well-known reluctance is referred to as "stock market participation puzzle" (Haliassos and Bertaut, 1995).

Underinvestment can also be observed for other types of risky securities both in the United States and in many European countries (Guiso, Haliassos and Jappelli, 2002). For Germany, Börsch-Supan and Eymann (2002) find that the direct holding of risky financial assets has increased between the mid 1970s and the late 1990s. Still, like most continental European countries and in contrast to Anglo-Saxon countries, Germany only has a low share of households participating in risky asset markets.

One factor that potentially explains why households are less likely to hold risky assets than traditionally suggested by financial literature is labor income risk. Labor income is uncertain with respect to future income streams and it constitutes the primary source of income for most households. Higher levels of uncertainty in this income component might induce households to limit the deliberate exposure to other kinds of financial risk. In his 2006 presidential address to the American Financial Association, John Campbell provides an intuitive explanation: *"[...] Households receive labor income but cannot sell claims to that income [...], much of the risk in labor income is idiosyncratic and therefore unhedgeable. This background risk increases effective risk aversion and leads households to invest more cautiously"* (Campbell, 2006, p.1559). One should note, though, that the effect would be reversed if households hedged against labor income risk by investing into assets for which the return is negatively correlated with their own labor income (for an overview, see Campbell, 2006).

A number of empirical studies assess whether labor income risk reduces the share of risky assets in a household's portfolio (Heaton and Lucas, 2000b; Hochguertel, 2003; Cardak and Wilkins, 2009), i.e. confining the analysis to households that hold any kind of risky asset. For example, Betermier, Jansson, Parlour and Walden (2012) document a 20% decrease of a household's share in risky assets in its total portfolio composition when labor income risk rises by 20%. In contrast to these studies we follow Haliassos and Bertaut (1995) and focus on the extensive margin, i.e. the overall propensity of households to hold any risky assets. The distinction between risk-taking households and households that do not invest in risky securities at all can be regarded as the single largest determinant of the average risky asset share in the population.

The topic of private risky asset holding has received increasing attention in recent years and - besides labor income risk - several other influencing factors have been identified. In the first place, housing (Fratantoni, 1998; Cocco, 2005) and borrowing constraints (Guiso, Jappelli and Terlizzese, 1996) play an important role as they limit the amount of money that is available for investments in risky assets. Also, Kaustia and Torstila (2011) find that political preferences have an effect on the investment decision. Financial literacy is identified by Christelis, Georgarakos and Haliassos (2011) and van Rooij et al. (2011) as a further determinant. More recently, behavioral factors have gained more attention as the traditional rational portfolio optimization theory is not able to fully explain the low stock market participation rate of households. Guiso et al. (2008) identify lack of trust in the stock market as a key to non-investment, for example. Hong et al. (2004) and Dierkes, Klos and Langer (2011) find social households - households that frequently interact with neighbors or regularly go to church - to invest in the stock market more often than other households. Breuer, Riesener and Salzmann (2014) have psychological traits like individualism pegged as drivers in stock market participation. The book edited by Bogan (2014) provides a good summary of recent advances in the field.

We contribute to the existing literature by using data from the German Socio-Economic Panel which comprises approximately 12,000 households. Unlike most other studies, which draw on cross-sectional data, we explicitly employ panel data methods in order to correct for household heterogeneity. Following Cardak and Wilkins (2009), we measure labor income risk by the coefficient of variation, relying on the normalized realized variability, rather than the deviation from some expected value of income. We find a significant negative effect of labor income variability on the propensity of households to hold risky securities. This effect remains robust even after accounting for a wide array of covariates, among them measures of the perceived financial situation. However, once we control for unobserved heterogeneity between households, labor income risk does not seem to be a major determinant anymore. We attribute this outcome to the fact that labor market success and the tendency to undertake risky investments are likely to be driven by the same household characteristics. Furthermore, it turns out that accounting for the subjective assessment of a household's current and its expectation regarding its future financial situation is even more important than actual income variability.

The remainder of the chapter is organized as follows: Section 2.2 presents the dataset and illustrates the key variables of the analysis. We explain the estimation strategy in Section 2.3. Section 2.4 presents the empirical results. Section 2.5 summarizes our findings and concludes.

2.2 Data

2.2.1 The German Socio-Economic Panel

The data for our analysis originates from the German Socio-Economic Panel (GSOEP), a panel study of German households, similar to the Panel Study of Income Dynamics (PSID) in the USA. It was established in 1984 as a representative sample of approximately 4,500 West German households. The survey is conducted annually by the German Institute for Economic Research (DIW) and is currently in its 27^{th} wave. At present, it covers some 12,000 households with about 20,000 individuals. The GSOEP is one of the most extensive longitudinal micro datasets worldwide and covers a wide range of socio-economic variables such as employment status, income sources, education level, or attitudes towards different aspects of life (see Haisken-DeNew and Frick (2005) or Wagner (2008) for an in-depth overview of the methodological aspects and data coverage of the GSOEP).

More importantly for this study, the GSOEP contains several questions regarding the financial situation of the adult household members, both on the personal and the household level. Specifically, it is reported whether households invest in certain asset classes such as savings accounts, life insurance policies or securities. Since 2001 the question concerning the holding of securities has distinguished between "fixed-interest securities (e.g. saving bonds, mortgage bonds, federal saving bonds)" and "other securities (e.g. stocks, funds, bonds, equity warrant)". We use the latter variable to investigate whether a household invests in risky assets or not. We therefore employ a subsample of the data which covers the period 2001 (the first year for which our dependent variable exists) to 2010.

The GSOEP reports the holding of risky securities at the household level. However, the corresponding investment decision is in large parts determined by personal characteristics such as age, gender, education, or attitude towards risk, which are individual, but not household characteristics. To be able to combine household information with personal characteristics we choose the head of a household as unit of observation as is standard in the literature.

Our main interest lies in the effect of labor income risk on the decision to participate in risky asset markets. Yet, although labor income is measured on the personal level, we are rather interested in the fluctuation of the total labor income of the household. Thus, we aggregate the labor income of the household head and a potential spouse before computing our measure of income risk. Labor income of children or other dependents is disregarded as these income streams are rather unsteady and are most likely not taken into account by the household head when making investment decisions.

2.2.2 The Coefficient of Variation as a Measure for Income Risk

The classical measure of risk in the financial literature is the variance. However, the major drawback of this measure is that for highly skewed variables such as labor income the results would be largely driven by few households with very high labor income. We thus follow the approach of Cardak and Wilkins (2009) and use the coefficient of variation instead. The coefficient of variation is the ratio of the standard deviation to the absolute value of the mean. It is a normalized measure of variation that does not depend on the unit of measurement and is therefore not distorted by very high values of our labor income variable. We calculate the coefficient of variation for each household based on income data of the past five years prior to the time of observation.[1]

The coefficient of variation seems particularly suitable for our analysis because the fluctuation of the income stream is a feature that a household can observe itself, even without statistical knowledge. The variation of income is observable through account statements or pay slips and the absolute deviation is a crude measure for income variability. Looking at the variable income stream is probably the most objective way for a household to evaluate its exposure to income risk. The coefficient of variation standardizes the variation in order to allow for a cross-household comparison. Subjective measures like the subjective variance of real income as proposed by Guiso et al. (1996) have two drawbacks: first, data is not available to calculate the measure in all time

[1] In order to check robustness of our findings, we varied the length of the window when calculating the coefficient of variation. Our results turn out to be qualitatively robust.

periods. And second, it is unlikely that in particular financially less literate households really sit down on a yearly basis and think about expected inflation and expected income growth.
The simplicity in calculating the coefficient of variation comes at the cost that it is a backward looking measure drawing on past income data. The subjective variance of real income of Guiso et al. (1996), in contrast, is forward looking. It is next to impossible to state which measure is preferable. Most likely households to some extent take past income risk and future income risk into account when deciding to invest in risky financial assets.
Income risk is also influenced by the nature of the job, whether it is a permanent or a temporary employment, or whether one works full-time or part-time (cp. Grande and Ventura, 2002). Such state variables, however, are binary by nature and cannot capture the feature whether the income stream is stable or highly volatile. A full-time worker who has a contract of a duration of five years might very well perceive her income to be less risky than an otherwise identical worker on a one-year contract. However, a binary measure would classify them as equally exposed to income risk. The major risk a household faces is, of course, job loss. This event is captured by the coefficient of variation as the variance increases once labor income drops while the mean decreases. The measure will thus rise substantually.
Finally, the coefficient of variation is a model-free measure of risk and thus free of assumptions as opposed to model-based measures like the one proposed by Angerer and Lam (2009). The advantage of the latter is that it provides a decomposition of shocks to labor income risk into permanent and transitory components and thus allows for a finer distinction. Angerer and Lam (2009) find that permanent income risk reduces the investment in risky assets while transitory income risk has virtually no effect on the investment decision. This is an important distinction we cannot explicitly make using the coefficient of variation. However, a permanent positive (negative) income shock should in the long run leave the variance unchanged and raise (lower) the mean, such that the coefficient of variation falls (rises) to a new steady level.

2.2.3 Sample and Variable Selection

First and foremost, we restrict the analysis to household heads. We also exclude pensioners because this group is not exposed to labor income risk in the usual sense. The final sample then comprises approximately 69,000 observations over the ten-year observation period 2001-2010.

Variable	Description
Risky	Dummy variable for household investment in risky assets (Participation = 1)
Male	Dummy variable for gender of household head (Male = 1)
Age	Age of household head
College	Dummy variable for college education
Yeduc	Years of education of household head
Married	Dummy variable for marital status of household head (Married = 1)
Minor	Dummy variable for minor living in household (Minor present = 1)
East	Dummy variable for household head lived in GDR (Lived in GDR = 1)
Foreign	Dummy variable for household head is foreign (Foreigner = 1)
Finrisk	Attitude towards financial risk of head (0 = no risk taking - 10 = risk seeking)
Nethhinc	Net household income
Netwealth	Net wealth
Wealth Terc.	Wealth terciles
Savings	Dummy variable for household saves money (HH saves = 1)
Loans	Dummy variable for household holds loan (HH holds loan = 1)
Nasset	Number of different assets held
Inher	Dummy variable for ever inherited money (Inheritance = 1)
Finlit	Dummy variable for financial literacy of household head (Financial literate = 1)
Incsat	Income satisfaction of household head (0 = not satisfied at all - 10 = perfectly satisfied)
Worrfin	Worries about the financial situation of household head (1 = very concerned; 2 = concerned; 3 = not concerned)
Coefvar	Coefficient of variation of total household labor income

Table 2.1: Variable Description. The table presents the variables used in the estimation along with a description.

In our sample, about one in three households invests in some kind of risky asset. One should note, though, that we see a constant decline in the participation rate over time. While in 2001 about 34 percent of all households participated in the asset market, this was true for only about 27 percent in 2010. This is consistent with the broad decline in stock market participation described in the introduction of this thesis. Besides this general pattern, one can observe pronounced disparities between household types. Especially income levels and wealth exhibit a strong positive correlation with the participation decision of a household. For instance, while only about 10 percent of the households in the lower tercile of the wealth distribution invest in risky assets, the corresponding fraction for households in the upper tercile is about 50 percent. Finally, most households adjust their portfolio composition only infrequently: in the course of any two years about 80 percent of all households did not alter their participation decision.

Previous research has identified several socio-economic variables that strongly affect the participation decision of households. For an overview, see Haliassos and Bertaut (1995), Campbell (2006), or Guiso and Sodini (2013). In our model we try to account for as many of these influences as possible. Table 2.1 provides an overview and a description of the variables, which are motivated subsequently.

Among the strongest predictors of stock market participation are wealth and income levels. Hence, we include the log of the net household income along with a one-year lag as well as dummy variables, which indicate to which tercile of the wealth distribution a household belongs. Moreover, more educated households are more likely to invest in risky assets, possibly because they are better able to understand sophisticated financial products. To capture this effect we use a dummy variable for college-educated household heads in the model. Furthermore, it should be self-evident that the more comfortable someone is with taking financial risk, the more likely he or she will invest in stocks or other risky assets. We use the information provided by the GSOEP to account for self-assessed attitude towards financial risk of the household head. Age is also an important factor as individuals face different goals, financial constraints, and saving horizons in different phases of their life. To represent the age structure, we include dummy variables for age groups of ten years. We account for these influences in addition to our measure of labor income risk and time dummy variables (for every year but 2001) in our baseline model.

Variable	Mean	Std. Dev.	Min.	Max.	N
Risky	0.318	0.466	0	1	68924
Male	0.610	0.488	0	1	68924
Age	44.521	12.07	18	88	68924
Yeduc	12.372	2.644	7	18	68924
College	0.219	0.414	0	1	68924
Married	0.505	0.5	0	1	68924
Minor	0.307	0.461	0	1	68924
East	0.101	0.301	0	1	68924
Foreign	0.077	0.267	0	1	68924
Finrisk	2.485	2.257	0	10	68924
Nethhinc	2360	1531	28	99999	68924
Loginc	7.608	0.574	3.332	11.513	68924
Loginc lag	7.599	0.569	3.332	11.35	68924
Netwealth	126608	415625	-1452000	31665604	13763
Wealth Terc.	1.772	0.785	1	3	68924
Savings	0.607	0.488	0	1	68924
Loans	0.262	0.44	0	1	68924
Nasset	2.315	1.446	0	6	68924
Inher	0.242	0.428	0	1	68924
Finlit	0.149	0.356	0	1	68924
Incsat	6.009	1.908	0	10	68924
Worrfin	1.987	0.546	1	3	68924
Coefvar	0.490	0.564	0	2.449	68924

Table 2.2: Summary Statistics. This table presents summary statistics (mean, standard deviation, extreme observations and number of observations) for the variables used in the estimation. Net household income (nethhinc) is also reported together with the logarithmic data.

The variables household wealth and self-assessed risk aversion are only available in two waves (2002 and 2007). We therefore assume some stability of these factors over a short time horizon. For wealth this means that the household would remain in the same tercile while of course actual wealth as measured in Euro may change until we revise the classification. We find that a large majority of households pertained to the same wealth tercile. Furthermore, only 1% of households moves from the first to the third tercile or vice versa. We therefore consider it unproblematic to treat the sorting into the wealth tercile as constant between these two waves. Changes in self-assessed risk aversion are slightly more notable, but the second wave for which it is available coincides with the initial period of the financial crisis. We therefore assume that risk aversion is stable in the short run, but experienced a level shift due to the financial crisis.

In addition, we control for demographic characteristics that are likely to have an effect on the decision process. Men and women often behave differently when it comes to personal finance decisions, and foreigners often face circumstances and constraints that differ from those of natives. It will also matter whether the household head is married and whether a minor lives in the household. Finally, we expect the investing behavior of people who where socialized in the German Democratic Republic (GDR) to deviate from that of individuals who where born in West Germany prior to 1990 due to the different socio-economic environment they were raised in.

Other aspects regarding a household's financial situation are likewise of importance. We include dummy variables that indicate whether a household is saving on a regular basis or holds loans. We also control for inheritance and whether the household head can be considered financially literate, i.e. whether he or she has ever received some kind of formal training in a finance related field. Lastly, we include the number of different asset classes that a household invests in besides risky assets.

Along with these objective, measurable indicators of a household's financial situation, subjective perception of the latter is most likely to be of equal importance as investment decisions depend on the expectations and concerns of the decision maker. The GSOEP offers two variables that measure this type of subjective perception of a household's finances. First, individuals are asked to rate their satisfaction with their income. Second, they state how worried they are about their financial situation. We expect these variables to capture the subjective factors that influence the investment decision, which by nature cannot be covered by the (objective) income or wealth variables. We include the average of these two variables over the five years prior to the observation, as we believe that subjective feelings are likely to have a long lasting impact on investment decisions.

Summary statistics of all variables are provided in Table 2.2. It shows that the average household head is a 45-years-old male with 12 years of education and moderate willingness to bear risk.

Variable	holding		non-holding	
	Mean	Std. Dev.	Mean	Std. Dev.
Male	0.648	0.478	0.593	0.491
Age	44.561	11.652	44.502	12.261
College	0.344	0.475	0.161	0.368
Bilzeit	13.416	2.794	11.885	2.422
Married	0.558	0.497	0.480	0.500
Minor	0.308	0.462	0.306	0.461
East	0.094	0.292	0.104	0.305
Foreign	0.033	0.179	0.097	0.297
Finrisk	3.393	2.312	2.061	2.101
Nethhinc	2954	1904	2082	1226
Loginc	7.857	0.508	7.492	0.565
Loginc Lag	7.831	0.515	7.491	0.561
Netwealth	216491	626679	83868	249995
Wealth Terc.	2.114	0.759	1.613	0.745
Savings	0.821	0.384	0.508	0.500
Loans	0.212	0.409	0.285	0.451
Nasset	3.017	1.260	1.987	1.410
Inher	0.358	0.479	0.188	0.391
Finlit	0.223	0.416	0.114	0.318
Incsat	6.802	1.618	5.639	1.921
Worrfin	2.197	0.515	1.889	0.532
Coefvar	0.356	0.432	0.552	0.605

Table 2.3: Summary Statistics by Participation. The table presents summary statistics (mean and standard deviation) for the variables used in the estimation grouped by participation in risky asset holding (variable "risky"). Columns labeled "holding" refer to households which hold risky assets (N=23781), columns labeled "non-holding" refer to households who do not hold any risky assets throughout the sample period (N=44139).

The average household in our sample has 2,360 Euros net income, is based in West Germany, and holds two different assets. We also present summary statistics by participation outcomes in Table 2.3. One can see that households that do invest in securities with variable rates are much more prosperous compared to their non-investing counterparts. Additionally, their heads are on average better educated and more willing to take risks in financial matters. Also, the volatility of the labor income seems to be considerably lower for these households.

2.3 Estimation Strategy

In order to estimate the effect of an increase in the variation of household labor income on the probability that a household will hold risky financial assets, we employ a twofold estimation strategy to assure robustness of our findings: on the one hand, we incrementally add different control variables (as described in Section 2.2) to the baseline model. On the other hand, we use several econometric models in order to rule out that our results are driven by unobserved heterogeneity.[2]

We are faced with a classical discrete choice problem in a panel data framework where we model the latent utility y_{it}^* that household i derives in year t from investing in risky financial products as

$$y_{it}^* = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it} \ , \tag{2.1}$$

where $\boldsymbol{x}_{it}$ is the vector of covariates for household i in period t, as described in Section 2.2, and $\boldsymbol{\beta}$ is the corresponding coefficient vector. c_i denotes heterogeneous, household specific characteristics and ε_{it} is a standard idiosyncratic error term. Define y_{it} as an observable binary variable that equals 1 if household i holds any form of risky asset in year t and 0 else. A household is said to invest in this asset type if the unobserved utility associated with holding the asset is positive, i.e. $y_{it} = 1[y_{it}^* > 0]$. Then the conditional probability for risky asset holding is given by

$$P(y_{it} = 1 | \boldsymbol{x}_{it}, c_i) = F(\boldsymbol{x}_{it}\boldsymbol{\beta} + c_i) \qquad t = 1, \ldots, T. \tag{2.2}$$

$F(\cdot)$ is a cumulative density function (CDF), which is usually either a standard normal CDF $\Phi(\cdot)$ or a logistic CDF $\Lambda(\cdot)$. The former case implies a probit model, the latter a logit model. To obtain the standard random effects (RE) probit model, one assumes that the household fixed

[2] We also check for robustness by estimating the models on several sub-samples. Results are not reported as our results are qualitatively robust in this respect.

effects c_i are conditionally normal distributed: $c_i|\boldsymbol{x}_i \sim \mathcal{N}(0, \sigma_c^2)$. One can then integrate out c_i to get consistent estimates and use σ_c^2 to compute marginal effects.

The RE assumption implies that household heterogeneity only affects the outcome of the dependent variable but is not related to the covariates. However, this assumption is unlikely to hold. Consider, for instance, a household composed of members who find it easy to defer gratification. Such a household can (*ceteris paribus*) be expected to be more prone to invest in risky financial assets, which entail a long investment horizon. At the same time, though, this type of household is also likely to have a more stable employment history and, in consequence,F a smoother labor income stream compared to households that consist of members with a high time preference. In such a situation, as long as we are not able to control for this characteristic, our estimate for the effect of labor income risk on investment decisions will be biased in a simple model. Given panel data, we can address the problem of unobserved heterogeneity, which is potentially correlated with explanatory variables, by using a fixed effects logit model. The basic intuition behind this model is that by conditioning on the total number of years S in which a household actually held risky assets, we can eliminate the household fixed effects c_i in Equation (2.1). Thus, one can estimate the conditional probabilities for risky asset holding $P\left(y_{it} = 1 \mid \boldsymbol{x}_{it}, \sum_{t=1}^{T} y_{it} = S\right)$ while allowing for a relationship between c_i and $\boldsymbol{x}_{it}$.

Yet, the fixed effects logit estimator has several disadvantages. First, we cannot estimate the coefficient for time-invariant variables such as gender. Additionally, we can only conduct the estimation for households that exhibit some variation of the dependent variable within the observed time period. This leads to a significant loss of observations as the decision whether or not to invest in risky financial products is quite long-lasting in our sample. Finally, the estimation of marginal effects is unfeasible as it would require the observation of c_i. Still, statistical significance, the sign of the estimates, and the relative effect associated with two variables can be interpreted.

Another model which avoids most of the disadvantages of the fixed effect logit model, but still allows for some correlation between the unobservables c_i and the covariates $\boldsymbol{x}_{it}$, is Chamberlain's correlated random effects model. We employ the Mundlak (1978) version of this model, which assumes that part of the individual heterogeneity is a linear function of the time average of the covariates:

$$c_i = \psi + \bar{\boldsymbol{x}}_i \boldsymbol{\xi} + a_i \tag{2.3}$$

where ψ and $\boldsymbol{\xi}$ are coefficients and a_i is an i.i.d. error term. The model presumes that most of the individual heterogeneity is captured by a linear combination of the time averages $\bar{\boldsymbol{x}}_i$. For the residual term a_i the usual random effects assumptions apply. The result is a random effects probit model which accounts for the potentially endogenous nature of the fixed effects and has the following form:

$$P(y_{it} = 1 | \boldsymbol{x}_{it}, c_i) = F(\psi + \boldsymbol{x}_{it}\boldsymbol{\beta} + \bar{\boldsymbol{x}}_i \boldsymbol{\xi} + a_i) \qquad t = 1, \ldots, T. \tag{2.4}$$

In contrast to the fixed effects logit model, the form of the relationship between the individual heterogeneity and the covariates has to be specified explicitly and rather strictly. But, as mentioned above, these stronger assumptions entail the advantage that households that did not alter their investment decision over time remain in the sample and can be used for estimation. More importantly, we can compute marginal effects as in the standard random effects setting. Of course both the fixed effects logit as well as the correlated random effects model have their disadvantages if the underlying assumptions are not met. However, as stated before we are convinced that it is important to try and address the problem of unobserved heterogeneity in the present context. By applying two different models we seek to provide robust interpretations which are not simply model driven.

2.4 Empirical Results

In the following, we first present the results of the random effects probit estimation (RE) for the different model specifications. Afterwards, we compare these results to those obtained by controlling for unobserved heterogeneity via the fixed effects logit (FE) and the correlated random effects probit (CRE) model. In all Tables 2.4, 2.5, and 2.6 coefficient estimates for the dummy variables for years and age groups are not reported to conserve space, but are available upon request.

2.4.1 Random Effects Probit Model

Estimation results based on the random effects probit model are summarized in Table 2.4. Inference for the average marginal effects is based on 400 clustered bootstrap samples.[3] We estimate four different specifications where we gradually add further control variables as outlined in Section 2.2.

The first column in Table 2.4 shows the average marginal effects of the main explanatory variables identified by the literature. The main focus of this study, the coefficient of variation, is highly statistically significant and exhibits the expected negative sign. Thus, higher income variability is associated with a reduction in the propensity to invest in risky assets. More precisely, a one-unit increase in the variation of a household's labor income, as measured by the coefficient of variation, reduces the probability that the household in question holds risky securities by 3.5 percentage points. The other variables are also statistically significant and display the expected sign. As reasoned in Section 2.2, we find that indeed wealth, income, education level, as well as risk appetite increase the likelihood of a risky investment. For instance, a household in the upper tercile of the wealth distribution has a 14.2 percentage points higher probability of holding stocks or the like as compared to a household in the lower tercile.

In the next step we add the demographic variables. Controlling for these household characteristics does not affect the coefficient estimate of our variable of interest: income variability is still significant and has the same sign and magnitude as before. By and large, the same holds true for the other control variables. With respect to the demographic variables we find that men are somewhat less likely to invest in risky assets. Whether the household head is married does not seem to be of importance, while the presence of a minor in the household slightly reduces the propensity to invest in risky assets by 1.4 percentage points. People born in the GDR are slightly more willing to invest in risky assets than West Germans in this setting. Foreigners

[3] We draw 500 samples with replacement and calculate marginal effects for each bootstrap sample. The standard error of the marginal effect is then estimated by the standard deviation across all bootstrap repetitions.

are decisively less likely to invest in risky assets: the probability is significantly reduced by 16.4 percentage points.

	Base	Demographic	Financial	Perception
College	0.117***	0.110***	0.100***	0.090***
	(0.008)	(0.008)	(0.007)	(0.007)
Finrisk	0.028***	0.028***	0.027***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)
Wealth 2nd	0.087***	0.085***	0.057***	0.051***
	(0.007)	(0.007)	(0.007)	(0.007)
Wealth 3rd	0.142***	0.139***	0.094***	0.082***
	(0.009)	(0.009)	(0.009)	(0.009)
Loginc	0.089***	0.092***	0.057***	0.047***
	(0.005)	(0.005)	(0.005)	(0.005)
Loginc lag	0.035***	0.037***	0.030***	0.018***
	(0.004)	(0.004)	(0.004)	(0.005)
Coefvar	-0.035***	-0.035***	-0.027***	-0.021***
	(0.006)	(0.006)	(0.005)	(0.005)
Male		-0.017***	-0.014**	-0.015***
		(0.006)	(0.006)	(0.006)
Married		0.001	-0.008	-0.006
		(0.007)	(0.007)	(0.007)
Minor		-0.014**	-0.013**	-0.011*
		(0.006)	(0.006)	(0.006)
East		0.016**	0.014**	0.032***
		(0.007)	(0.007)	(0.007)
Foreign		-0.164***	-0.129***	-0.120***
		(0.016)	(0.014)	(0.014)
Savings			0.065***	0.057***
			(0.004)	(0.004)
Loans			-0.012***	-0.008**
			(0.004)	(0.004)
Nasset			0.036***	0.034***
			(0.002)	(0.002)
Inher			0.059***	0.055***
			(0.007)	(0.007)
Finlit			0.094***	0.089***
			(0.007)	(0.007)
Incsat				0.014***
				(0.002)
Worrfin				0.044***
				(0.007)
Constant	-1.214***	-1.229***	-1.045***	-1.060***
	(0.048)	(0.049)	(0.048)	(0.048)
age	yes	yes	yes	yes
year	yes	yes	yes	yes
Pseudo R^2	6.33	6.64	8.76	9.18

Table 2.4: Random Effects Probit Model Estimation Results. The table presents the calculated marginal effects based on the estimation results of the random effects probit model. Clustered bootstrap standard errors are in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively. The Pseudo R^2 is reported in percent.

Continuing, we control for the financial behavior of households. The loan dummy variable has a very small, but as expected negative effect. All other financial variables (number of asset classes held, inheritance and financial literacy) exert a strong positive influence on the probability to hold risky assets and are highly statistically significant. Accounting for the household's overall financial background leads to a decrease of the effect of a volatile labor income stream from -3.5 percentage points to -2.7 percentage points. The estimate still remains highly statistically significant, though. This reduced magnitude is probably due to the effect of different levels of labor income uncertainty on saving and investment patterns. For several of the other variables, which are carried over from the previous model, we see a decrease in the size of their effect as well. Again, this is likely due to correlation of these control variables to the household's finances. This is particularly true for wealth and income effects as these are the most important determinants of the financial behavior of a household. For instance, the effect of log-income declines from 8.9% to 5.7%.

The fourth column of Table 2.4 displays the full model after adding variables which pick up the household head's perception of the household's financial situation. The more satisfied a household's head is with its income, the more likely (1.4 percentage points for a one unit increase) she or he is to invest in risky assets. Similarly, for a household which is less worried about the financial environment, the probability that this household holds risky assets is increased by 4.4 percentage points for each one unit increase. Both variables are also statistically significant.

As before, the variables most affected by adding the subjective measures are wealth and income. A possible explanation is that with increasing levels of income and wealth the household will generally be more satisfied with and less worried about its financial situation. The subjective measures are highly important even though we control for the objectively measurable financial indicators. This reflects findings in the behavioral finance literature that people's expectations and concerns are very important determinants of the investment process.

Turning again to the coefficient of variation, we observe that while it is still statistically significant, the marginal effect is now down to -2.1 percentage points. This leads us to the conclusion that objective factors are not the only driving force in the decision making process when it comes to investing in stocks. Perception of one's own financial situation and of the financial environment are likewise of importance. This conclusion is supported by the decreasing magnitude of the coefficient estimate of income variability across the four model specifications: if we account for the objective factors only (as in the baseline case), they will pick up a certain amount of the subjective influences. As soon as we control for these subjective factors, we get a more accurate estimate of how income variability affects the investment decision.

To sum up, the effect of realized labor income variation on the propensity to make risky financial investments is negative and highly significant. Controlling for a variety of variables does not change this finding. The magnitude of the effect, however, declines from -3.5 percentage points in the baseline model to -2.1 percentage points in the model with all control variables. One should note that this effect is not very large, as the reported magnitude is associated with a one-unit increase in the coefficient of variation. Yet, the measure takes values between 0 and 2.45 in the data and exhibits a standard deviation of about 0.564. Thus, a one unit increase can be seen as a fairly large increase in the variability of the household's labor income stream. Nevertheless, the magnitude of our results is in line with previous findings in the literature. This also holds true for the overall explanatory power of our models as measured by the Pseudo R^2. It is also reassuring that it increases from 6.33% in the baseline case to 9.18% in the fully-specified model.

2.4.2 Accounting for Unobserved Heterogeneity

As outlined in Section 2.3, the results presented so far are potentially biased due to correlation of the covariates with unobserved household characteristics. In the following, we examine the results obtained after accounting for unobserved heterogeneity by means of fixed effects logit and correlated random effects probit models.[4] The approach of an incremental addition of control variables is the same as before.

It should be emphasized that the fixed effects logit model cannot identify the effects of time-invariant variables, such as gender. For the correlated random effects probit model the effect of

[4]We test whether the inclusion of time averages has no effect in our model. The resulting p-value is lower than 0.1% and we therefore conclude that household fixed effects are indeed present.

variables with no or only little variation over time can be observed, but is in fact indistinguishable from the effect of the corresponding time averages. The models do control for these variables, though. Variables that are not completely time-invariant but do change only slightly within households over the observed time span usually are statistically insignificant due to the lack of variation. In our model this is true for the dummy variables for college and foreign nationality, which are highly significant in the random effects probit model but are not significant at all in the other two models.

Coefficient estimates of the FE model are summarized in Table 2.5. Their interpretation is limited to statistical significance and their sign. The relative effect of two variables can also be compared to each other. This fact can be used to further compare the fixed effect logit model to the two models using the random effects assumption.

One should also note that we observe a different sample for the FE logit model compared to the RE models. As mentioned before, we only observe results for households that switch at least once between holding and non-holding of risky assets in our sample period. Consequently, the number of observations decreases from about 69,000 (about 11,000 households) for the RE models to about 28,500 observations (about 3,800 households) for the FE logit model. More importantly, the composition of our sample is likely to change. For instance, as many low-income, low-wealth households never hold risky assets and many wealthy households always hold risky assets over the observed time period, the wealth and income distribution for the FE logit sample is more compressed compared to the RE sample. This will potentially affect the estimated coefficients for these variables and it also leads to a lower Pseudo R^2 as compared to the RE model. However, as we will see, the sign, significance and relative effects of the variables in this model are very similar to those for the correlated random effects model. We thus conclude that sample selection issues do not seem to matter much for the FE logit model in this context.

We find that the main determinants, which exhibit some variation over time, such as wealth and income, are still highly statistically significant and exhibit the same sign as in the RE probit model. When we add the additional variables, the magnitude of these coefficients decreases, but they remain statistically significant, similar to what we observe for the random effects probit

	Base	Demographic	Financial	Perception
College	0.117	0.108	0.068	0.074
	(0.238)	(0.239)	(0.238)	(0.237)
Finrisk	0.136***	0.136***	0.136***	0.137***
	(0.020)	(0.020)	(0.020)	(0.020)
Wealth 2nd	0.267***	0.247***	0.173*	0.158*
	(0.091)	(0.091)	(0.091)	(0.091)
Wealth 3rd	0.446***	0.425***	0.317***	0.293**
	(0.116)	(0.116)	(0.116)	(0.117)
Loginc	0.631***	0.596***	0.382***	0.349***
	(0.064)	(0.065)	(0.065)	(0.066)
Loginc Lag	0.233***	0.210***	0.165***	0.121**
	(0.056)	(0.056)	(0.056)	(0.056)
Coefvar	-0.163**	-0.166**	-0.139*	-0.099
	(0.080)	(0.080)	(0.080)	(0.081)
Married		0.251***	0.189*	0.185*
		(0.097)	(0.097)	(0.097)
Minor		0.0740	0.0551	0.052
		(0.072)	(0.073)	(0.073)
Foreign		0.304	0.333	0.329
		(0.383)	(0.395)	(0.397)
Savings			0.432***	0.415***
			(0.046)	(0.046)
Loans			-0.007	-0.000
			(0.047)	(0.047)
Nasset			0.232***	0.229***
			(0.023)	(0.023)
Inher			0.235**	0.222**
			(0.105)	(0.105)
Incsat				0.080**
				(0.032)
Worrfin				0.236**
				(0.103)
Pseudo R^2	2.92	2.99	4.25	4.37

Table 2.5: Fixed Effects Logit Model Estimation Results. The table presents the coefficient estimates based on the fixed effects logit model. Clustered standard errors are in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively. The Pseudo R^2 is reported in percent.

model. The same can be said for most of the additional variables where we observe significant positive effects of saving, number of asset classes held, and satisfaction with income, for example. Thus, even after we control for unobserved differences between households, many of the effects remain stable, at least in terms of direction and significance.

Now that we have eliminated inherent household differences, we are most interested in the impact of the coefficient of variation on the asset market participation probability. We find that the negative sign of our income risk measure remains unchanged with a coefficient equal to -0.163. However, the significance of the coefficient is lower than before with a p-value of about 4 percent. When adding the demographic and financial control variables, the magnitude of the coefficient

	Base	Demographic	Financial	Perception
College	0.006	0.000	-0.001	0.000
	(0.018)	(0.018)	(0.018)	(0.018)
Finrisk	0.012***	0.012***	0.012***	0.012***
	(0.001)	(0.002)	(0.002)	(0.002)
Wealth 2nd	0.023***	0.022***	0.015**	0.014*
	(0.007)	(0.007)	(0.007)	(0.007)
Wealth 3rd	0.037***	0.036***	0.026***	0.024**
	(0.009)	(0.009)	(0.009)	(0.010)
Loginc	0.059***	0.058***	0.038***	0.035***
	(0.005)	(0.005)	(0.005)	(0.005)
Loginc Lag	0.019***	0.018***	0.014***	0.010**
	(0.004)	(0.004)	(0.004)	(0.004)
Coefvar	-0.015**	-0.015***	-0.012**	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)
Married		0.021***	0.016**	0.016**
		(0.008)	(0.008)	(0.008)
Minor		0.006	0.005	0.004
		(0.006)	(0.006)	(0.006)
Foreign		0.024	0.027	0.026
		(0.031)	(0.032)	(0.032)
Savings			0.038***	0.036***
			(0.004)	(0.004)
Loans			0.000	0.000
			(0.004)	(0.004)
Nasset			0.022***	0.022***
			(0.002)	(0.002)
Inher			0.017**	0.016**
			(0.008)	(0.008)
Incsat				0.007***
				(0.002)
Worrfin				0.021***
				(0.008)
Constant	-1.352***	-1.539***	-1.267***	-1.252***
	(0.052)	(0.054)	(0.053)	(0.053)
age	yes	yes	yes	yes
year	yes	yes	yes	yes
Pseudo R^2	7.93	8.52	10.85	10.96

Table 2.6: Correlated Random Effects Probit Model Estimation Results. The table presents the calculated marginal effects based on the estimation results of the correlated random effects probit model. Clustered bootstrap standard errors are in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively. The Pseudo R^2 is reported in percent.

estimate decreases only slightly which is in line with what we have seen before. Also, the coefficient of variation remains significant at the 5% or 10% significance level, respectively. Once we account for the measures of perceived financial situation, though, it is no longer significant at any conventional significance level with a p-value of about 21 percent. The magnitude of the coefficient further decreases to -0.1. Thus, it seems that once we control for household heterogeneity along with all other control variables - especially the ones capturing subjectivity - the effect of labor income variability seems negligible. However, before drawing conclusions we check the robustness of our results by turning to the CRE model.

The CRE model allows for computation of average marginal effects, which enable us to interpret the size of the effects. From Table 2.6 we see that the estimates of the CRE model point towards the same conclusions as those of the FE logit model. For all four model settings, we find that significance of the estimates is quite similar to that of the corresponding estimators in the FE model. The signs are also identical (with the exception of the insignificant variable "college"). Similar to Table 2.5, the coefficient of variation is significant for the first three model specifications. Once we add the variables that account for income satisfaction and worries about financial situation, though, the income variation is not statistically significant anymore. Still, the relative effect of the two variables is roughly comparable between the FE and the CRE model. These findings are reassuring as they indicate that the previous results are not driven by the characteristics of the FE logit model.

After controlling for the unobserved heterogeneity, we find the effect of a one-unit increase in the coefficient of variation to lead to a decrease in the participation probability by 1.5 percentage points in the baseline model. This is approximately two thirds of the magnitude of the effect in the fully-specified RE model. Controlling for all variables in the CRE framework further reduces the marginal effect to less than one percentage point for a one-unit increase of the coefficient of variation. Even though most variables exhibit a decrease in magnitude, this reduction is more important for the coefficient of variation as its effect was already rather small. However, the negligible size of its marginal effect together with the lack of significance in the fully-specified model for both the FE logit and the CRE probit estimation strongly indicate that labor income risk at best plays a very small role in determining the propensity of a household to invest in risky asset classes.

The above findings are summarized in Table 2.7 where we compare the results of all three models with all covariates. Note again that the columns RE and CRE contain marginal effects while

	RE	CRE	FE
College	0.090***	0.000	0.074
	(0.007)	(0.018)	(0.237)
Finrisk	0.027***	0.012***	0.137***
	(0.001)	(0.002)	(0.020)
Wealth 2nd	0.051***	0.014 *	0.158*
	(0.007)	(0.007)	(0.091)
Wealth 3rd	0.082***	0.024**	0.293**
	(0.009)	(0.010)	(0.117)
Loginc	0.047***	0.035***	0.349***
	(0.005)	(0.005)	(0.066)
Loginc Lag	0.018***	0.01**	0.121**
	(0.005)	(0.004)	(0.056)
Coefvar	-0.021***	-0.009	-0.099
	(0.005)	(0.006)	(0.081)
Married	-0.006	0.016**	0.185*
	(0.007)	(0.008)	(0.097)
Minor	-0.011*	0.004	0.052
	(0.006)	(0.006)	(0.073)
Foreign	-0.120***	0.026	0.329
	(0.014)	(0.032)	(0.397)
Savings	0.057***	0.036***	0.415***
	(0.004)	(0.004)	(0.046)
Loans	-0.008**	0.000	-0.000
	(0.004)	(0.004)	(0.047)
Nasset	0.034***	0.022***	0.229***
	(0.002)	(0.002)	(0.023)
Inher	0.055***	0.016**	0.222**
	(0.007)	(0.008)	(0.105)
Incsat	0.014***	0.007***	0.080**
	(0.002)	(0.002)	(0.032)
Worrfin	0.044***	0.021***	0.236**
	(0.007)	(0.008)	(0.103)
Observations	68,924	68,924	28,477
Number of HH	10,933	10,933	3,781

Table 2.7: Model Comparison. The table presents the estimation results of the fully specified model for the random effects, the correlated random effects and the fixed effects specification. Columns "RE" and "CRE" hold marginal effects while column "FE" holds coefficient estimates. Clustered (bootstrap, if applicable) standard errors are in parentheses. ***, ** and * denote significance on the 1%, 5% and 10% significance level, respectively.

the column FE contains parameter estimates only. The difference for the coefficient of variation is quite striking, while other important variables such as the log of household income or saving behavior are not that different across models. Most interestingly, the subjective measures of the financial situation ("incsat" and "worrfin") are highly significant across all models and their impact is not negligible in economic terms. In particular in the CRE and the FE models, adding these subjective variables, which account for income satisfaction and the perception of the financial environment, seems to drive out the objective measure of income risk.

2.5 Conclusion

We investigate the determinants of a household's decision to invest in risky assets. Financial theory suggests that households that face higher labor income risk are less likely to hold risky financial assets. We model the probability of holding risky assets using the coefficient of variation of household labor income as a measure for labor income risk. Along with a large number of explanatory variables, which have already been considered in the literature, we include measures of a household head's income satisfaction and whether she or he is worried about her/his financial situation.

Our results do not contradict the predictions of financial theory, but also do not stress them. At first, we find a significant negative impact of income variability on the propensity to hold risky assets, which is in accordance with previous research and theoretical predictions. This finding still holds even after controlling for all our covariates, in particular the perception variables. The size of the effect is rather small, but again the magnitude is in line with previously reported results. However, once we control for unobserved heterogeneity, we observe a sharp drop in the magnitude of the estimated effect. Income variability remains significant until we arrive at the fully-specified model, which accounts for the household head's income satisfaction and worries about the financial situation.

We therefore conclude that much of the previously observed effect of labor income risk may have been driven by underlying unobserved household characteristics, as well as by the perception of the investment environment. It is thus not clear that less labor market uncertainty would lead to a higher level of investment in stocks by private households. Differences between household characteristics, as well as confidence and trust in the financial system, appear to play an important role. The latter, however, are hard to influence and in particular hard to build, but they are easily destroyed as the current situation of financial markets clearly demonstrates. A recent study of Bucher-Koenen and Ziegelmeyer (2014) investigates this issue and reveals that in particular financially illiterate households are prone to turn away from stock market investments once they have made a negative experience, i.e. experienced losses. This material loss may well be accompanied by a loss of trust in the financial system. In order to increase households' participation in the stock market, it seems therefore appropriate to undertake confidence-building measures (which might include reducing financial illiteracy) to stimulate households' risky asset holding.

Chapter 3

Multifractional Logit Estimation of Household Portfolio Shares

3.1 Introduction

In this chapter I 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, I analyze the composition of a household's financial wealth in a joint framework. Specifically, I 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 I follow the approach of Papke and Wooldridge (1996) who show that one can model the 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 (2014) who also account for the fact that shares have to sum up to one in such a framework. I use their extension in order to model the shares of the three aforementioned risk classes jointly. Furthermore, I adapt their model to a panel data in the spirit of Train (2009) to control for unobserved heterogeneity across households.

My analysis contributes to the field in several ways: I 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 I employ the SAVE survey - a rich micro survey on saving and investment decisions of German households which is still underexploited in my opinion. Furthermore, I 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 I obtain a more comprehensive picture of the portfolio structure and its determinants. On the methodological side I contribute by modeling shares via fractional response models. This approach is arguably

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

more appropriate than other often used models in this respect. What is more, by modeling shares in a joint framework I incorporate the interdependences between asset classes compared to a situation where one investigates each share separately. Finally, I extend multivariate fractional response models to the panel data case. Specifically I 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 chapter is organized as follows: Section 3.2 gives an overview of the current state of empirical research on the portfolio composition of private households. In Section 3.3 I motivate the econometric models which are employed in the subsequent estimations. Description of the data-set and summary statistics for my sample are presented in Section 3.4 before I report the empirical results in Section 3.5. Section 3.6 sums up the analysis and proposes potential future work.

3.2 Literature Review

This study 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 I review some of the main findings in the household finance literature. To explain the portfolio composition, empirical analyses typically control for demographic factors (such as 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 (for example risk-aversion and preferences) as well as other risk factors (i.e. health risk).

In the following I present results from the previous literature on some of the key variables for my 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, McCarthy and Turner, 1997; Halko, Kaustia and Alanko, 2012). Felton, Gibson and Sanbonmatsu (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.

3.3 Econometric Models

In this section I 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, I 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.

I 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, I show how this approach can be extended to the multivariate case, as suggested by Mullahy (2011) and Murteira and Ramalho (2014). Finally, I combine these two extensions of the original model to allow for the estimation of mean shares for several assets jointly in a panel data framework via simulation methods as described in Haan and Uhlendorff (2006).

3.3.1 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 a financial context are, for instance, the share of portfolio invest in stocks or the debt to asset ratio - see Cook, Kieschnick and McCullough (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:

$$E[y_i|\boldsymbol{x}_i] = \boldsymbol{x}_i\boldsymbol{\beta} \tag{3.1}$$

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

[2] Both Ramalho, Ramalho and Murteira (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 I 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).

overview of issues encountered in empirical studies in household finance as well as a survey of appropriate microeconometric models in this regard.

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:

$$E[y_i|\boldsymbol{x}_i] = G(\boldsymbol{x}_i\boldsymbol{\beta}) \in (0, 1) \tag{3.2}$$

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 $G(\cdot) = \Phi(\cdot)$ 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) I will focus on the Flogit specification where the conditional density for the ith 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:

$$E[y_i|\boldsymbol{x}_i] = \frac{exp(\boldsymbol{x}_i\boldsymbol{\beta})}{[1 + exp\,(\boldsymbol{x}_i\boldsymbol{\beta})]} \tag{3.3}$$

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

$$\mathcal{L}_i(\boldsymbol{\beta}) = f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = [G(\boldsymbol{x}_i\boldsymbol{\beta})]^{y_i}[1 - G(\boldsymbol{x}_i\boldsymbol{\beta})]^{1-y_i} \tag{3.4}$$

is maximized to obtain the QML estimator:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{L}_i(\boldsymbol{\beta}) \tag{3.5}$$

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: $\widehat{APE}_k = \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i\hat{\boldsymbol{\beta}})\hat{\beta}_k$.

3.3.2 Fractional Response Models with Unobserved Heterogeneity

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

[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, Hochguertel and Van Soest (2004) who model the joint decision to hold stocks and bonds over time.

An important issue in this respect 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 I denote as α_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 ith 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 one does not restrict the relationship of the variables over time. In effect, one treats the unobserved time-invariant characteristics α_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 α_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 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 I 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 α_i enter the link function additively in the fashion of a single index model. Assuming that the covariates are strictly exogenous, given α_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] I 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:

$$E[y_{it}|\boldsymbol{x}_{it}, \alpha_i] = \Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i) = \frac{exp(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)}{[1 + exp(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]} \in (0, 1) \tag{3.6}$$

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 α_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 α_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 α_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

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

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 α_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 α_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 α_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 ith household, i.e. the joint distribution conditional on α_i, becomes:

$$\mathcal{L}_i(\boldsymbol{\beta}) = f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}) = \prod_{t=1}^{T} [\Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]^{y_{it}}[1 - \Lambda(\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i)]^{1-y_{it}} \tag{3.7}$$

The unconditional joint distribution of shares for the ith household which no longer depends on α_i is obtained by integrating out the unobserved heterogeneity:

$$f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \boldsymbol{\beta}) = \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} f(y_{it}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta})\right] h(\alpha_i) d\alpha_i \tag{3.8}$$

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, I follow an approach similar to Papke and Wooldridge (2008) and assume that α_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 $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log f(y_{i1}, \ldots, y_{iT}|\boldsymbol{X}_i, \boldsymbol{\beta}) \tag{3.9}$$

[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.

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 α_i in Equation 3.8. 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 α_i are assumed to be independent from the covariates. As mentioned before, one does apply panel data models in large part to account for the potential confounding influence of α_i. On the other hand, fixed effects models, if at all feasible, are quite restrictive in a non-linear framework. 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 Chamberlain (1984). In the CRE approach one models the time-constant unobserved heterogeneity α_i as a linear combination of the time averages of the time varying covariates:

$$\alpha_i = \bar{\boldsymbol{x}}_i \boldsymbol{\xi} + a_i \tag{3.10}$$

Where $\bar{\boldsymbol{x}}_i$ is the vector of time averages of the covariates for the ith 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, α_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.

[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).

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 α_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 I 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, one is in need of simulation methods for the multivariate case anyway. In the following I 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 I 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 (2009).

Consider a situation, similar to the one in Equation 3.8, where one wants to solve an intractable integral of the form

$$f(y_i|\boldsymbol{x}_i,\boldsymbol{\beta}) = \int_{-\infty}^{\infty} f(y_i|\boldsymbol{x}_i,\alpha_i,\boldsymbol{\beta})h(\alpha_i)d\alpha_i \tag{3.11}$$

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 α_i^r into the function to be integrated. I obtain the Monte Carlo integral, which is the simulated likelihood contribution of agent i, by averaging over all resulting expressions:

$$\mathcal{SL}_i(\boldsymbol{\beta}) = \widehat{f}(y_i|\boldsymbol{x}_i, \alpha_{iR}, \boldsymbol{\beta}) = \frac{1}{R}\sum_{r=1}^{R} f(y_i|\boldsymbol{x}_i, \alpha_i^r, \boldsymbol{\beta}) \tag{3.12}$$

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 $R \to \infty$:

$$\widehat{\boldsymbol{\beta}}_{MSL} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{SL}_i(\boldsymbol{\beta}) \tag{3.13}$$

From this it is easy to see how to to obtain the partial effect $PE_{ik} = E_\alpha\left[g(\boldsymbol{x}_i\boldsymbol{\beta} + \alpha_i)\right]\beta_k$. One plugs 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:

$$\widehat{APE}_k = \frac{1}{NR}\sum_{r=1}^{R}\sum_{i=1}^{N} g(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}} + \alpha_i^r)\widehat{\beta}_k \tag{3.14}$$

Compared to the conditional mean approach used for cross-sectional analysis, as described in Subsection 3.3.1, 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.

3.3.3 Fractional Multinomial Response Models

The aim of this 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 I 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, I return to the cross-sectional case. I denote the share of the jth asset held by the ith individual as y_{ij}. A suitable model for this situation must reflect the bounded nature of each individual share (i.e. $0 \leq y_{ij} \leq 1$ 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 (2014). They both concentrate on multivariate fractional dependent data where the main focus is on modeling the conditional mean of shares jointly.

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

[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.

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 respect 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 3.3.1 in this fashion is straight forward. Using a multinomial logit specification as link function results in the so called fractional multinomial logit model (to which I will refer to as FMlogit). This is the main model specification in both Mullahy (2011) and Murteira and Ramalho (2014) and is given by:[11]

$$E[y_{ij}|\boldsymbol{x}_i] = \Lambda(\boldsymbol{x}_i\boldsymbol{\beta}_j) = \frac{exp(\boldsymbol{x}_i\boldsymbol{\beta}_j)}{\left[\sum_{h=1}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}, \qquad j = 1 \dots J \tag{3.15}$$

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:

$$E[y_{ij}|\boldsymbol{x}_i] = \frac{1}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}, \qquad j = 1 \tag{3.16}$$

$$E[y_{ij}|\boldsymbol{x}_i] = \frac{exp(\boldsymbol{x}_i\boldsymbol{\beta}_j)}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_i\boldsymbol{\beta}_h\right)\right]}, \qquad j = 2, \dots, J \tag{3.17}$$

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

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 kth regressor on the jth share:

$$PE_{ijk} = \frac{\partial E[y_{ij}|\boldsymbol{x}_i]}{\partial x_{ik}} = E[y_{ij}|\boldsymbol{x}_i] \cdot \left[\beta_{jk} - \frac{\sum_{h=2}^{J} \beta_{hk} exp(\boldsymbol{x}_i \boldsymbol{\beta}_j)}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_i \boldsymbol{\beta}_h\right)\right]} \right] \tag{3.18}$$

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

$$\mathcal{L}_i(\boldsymbol{\beta}) = \prod_{j=1}^{J} E[y_{ij}|\boldsymbol{x}_i]^{y_{ij}} \tag{3.19}$$

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

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{L}_i(\boldsymbol{\beta}) \tag{3.20}$$

Murteira and Ramalho (2014) 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 (2014) suggest alternative models, such as the nested logit or the mixed logit which are not afflicted with

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

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 I 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 (2014) 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 in the same way 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 (2014) 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.

3.3.4 Fractional Multinomial Response Models with Unobserved Heterogeneity

Murteira and Ramalho (2014) suggest that an application of the fractional multinomial logit model in a panel data framework looks very promising. A combination of the two extensions of the basic Flogit model presented in Subsection 3.3.2 and Subsection 3.3.3 does indeed seem to suggest itself. However, to the best of my knowledge, mine is the first study to extend the fractional multinomial logit model in this fashion. I 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®. Here, 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. I 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, I 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:

$$E[y_{ijt}|\boldsymbol{x}_{it}, \boldsymbol{\alpha}_i] = \frac{exp(\boldsymbol{x}_{it}\boldsymbol{\beta}_j + \alpha_{ij})}{\left[1 + \sum_{h=2}^{J} exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}_h + \alpha_{ih}\right)\right]} \tag{3.21}$$

I allow the unobserved heterogeneity $\boldsymbol{\alpha}_i = (\alpha_{i1} \ \alpha_{i2} \ \dots \ \alpha_{iJ})'$ to affect each share differently. In addition, I 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 I only allow for random intercepts α_{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 (2009) 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, 2009, 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 (2014). Furthermore, accounting for unobserved heterogeneity in a panel data framework 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 3.3.2 and Subsection 3.3.3. I write the likelihood contribution of a single household i for all shares and time periods conditional on the unobserved effects as:

$$\mathcal{L}_i(\boldsymbol{\beta}) = f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\alpha}_i, \boldsymbol{\beta}) = \prod_{j=1}^{J}\prod_{t=1}^{T} E[y_{ijt}|\boldsymbol{x}_{it}, \boldsymbol{\alpha}_i]^{y_{ijt}} \tag{3.22}$$

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

$$E_{\boldsymbol{\alpha}}[\mathcal{L}_i(\boldsymbol{\beta})] = \int_{-\infty}^{\infty} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\alpha}_i, \boldsymbol{\beta}) f(\boldsymbol{\alpha}_i) d\boldsymbol{\alpha}_i \tag{3.23}$$

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:

$$\mathcal{SL}_i(\boldsymbol{\beta}) = \frac{1}{R}\sum_{r=1}^{R} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\alpha}_i^r, \boldsymbol{\beta}) \tag{3.24}$$

The resulting estimator maximizes the sum of log simulated likelihoods:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} log\mathcal{SL}_i(\boldsymbol{\beta}) \tag{3.25}$$

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 I implement the model via maximum simulated likelihood estimation.[13] When I described the MSL approach in Subsection 3.3.2 I did not say anything about how to actually 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 (2009) 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 my model I 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. I come to the same conclusion in my 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 α_1 have been

[13] I implement the model in Mata® due to speed gains.

normalized to zero and the unobserved heterogeneity $\boldsymbol{\alpha}_i$ is assumed to follow a multivariate normal distribution. Allowing for correlation across equations, I 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 $\boldsymbol{\beta}$.

$$\boldsymbol{\alpha}_i = \begin{pmatrix} \alpha_{i2} \\ \alpha_{i3} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{i2} \\ \epsilon_{i3} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma_\alpha}) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha_2} & \sigma_{\alpha_2\alpha_3} \\ \sigma_{\alpha_2\alpha_3} & \sigma^2_{\alpha_3} \end{pmatrix} \right) \tag{3.26}$$

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 I 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:

$$\boldsymbol{\alpha}_i | \boldsymbol{X}_i \sim \mathcal{N}(\boldsymbol{f}(\bar{\boldsymbol{x}}_i), \boldsymbol{\Sigma_a}) = \mathcal{N}\left(\begin{pmatrix} \bar{\boldsymbol{x}}_i\boldsymbol{\xi}_2 \\ \bar{\boldsymbol{x}}_i\boldsymbol{\xi}_3 \end{pmatrix}, \begin{pmatrix} \sigma^2_{a_2} & \sigma_{a_2a_3} \\ \sigma_{a_2a_3} & \sigma^2_{a_3} \end{pmatrix} \right) \tag{3.27}$$

3.4 Data

3.4.1 SAVE Survey

The data for my 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

[14] As of 2010 for its ninth wave.

[15] See Börsch-Supan, Coppola, Essig, Eymann and Schunk (2008) for a more detailed account on the survey.

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 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 my 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 my analysis than the aforementioned surveys.

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

[17] For more information on these surveys see Wagner, Göbel, Krause, Pischner and Sieber (2008) for the GSOEP and von Kalckreuth, Eisele, Blanc, Schmidt and Zhu (2012) for the PHF.

3.4.2 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.

I employ a sample consisting of six consecutive waves of the SAVE survey from 2005 to 2010. My 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 study as it includes the financial crisis of 2007 to 2008 which potentially had a huge impact on the behavior of retail investors.

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 my analysis these households constitute the target population as only they allow me to construct financial asset shares.

In addition, I 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] I 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 me 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

[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, I stick to the lower threshold.

have to be interpreted as being conditional on positive amounts of financial wealth as well as a lack of business activities.[19]

3.4.3 Financial Asset Classes

The most interesting aspect of the SAVE survey for my research question 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 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").

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 3.1: Participation Rates and Shares of Financial Wealth invested in different financial asset classes. Source: Own computation using SAVE weights.

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 my data where more than 50 % of households in the

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

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, one can see in Table 3.1 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.

Looking more closely at the composition of household portfolios, Table 3.1 list the make up of the different assets in my 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 my 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.

3.4.4 Construction and Composition of Risk Classes

My main interest is to get a more complete picture of the risk profile of households' financial portfolios. Hence, I am not interested in individual assets that are similar in nature to each other.[21] In order to divide financial products into different categories I 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 allocation of assets by means of a mean-variance approach is not feasible here. Instead, I opt for a classification strategy which is widely used in the household finance literature[22] - I 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

[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, Coppola and Reil-Held (2012).For this chapter, however, such a distinction is of no interest.

[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, Brunetti and Maestas (2012) or Barasinska et al. (2012).

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 I denote as ”fairly safe”.

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

Table 3.2: Classification of Financial Assets by inherent Risk.

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 regard is that I cannot distinguish between the different types of bonds and assign them to the appropriate category.

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 3.3: Shares of Financial Wealth by Risk Class. Source: Own computation using SAVE weights.

The categorization of asset classes according to their inherent risk is summarized in Table 3.2. Table 3.3 and Figure 3.1 give an impression of the distributions of the share of financial wealth invested in each of the three risk categories. Table 3.3 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. The information in Table 3.3 complement those given in Table 3.1 which provides summary statistics for the risk classes along with those for the financial assets themselves. One 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 3.1 where I plot histograms for each category next to each other.

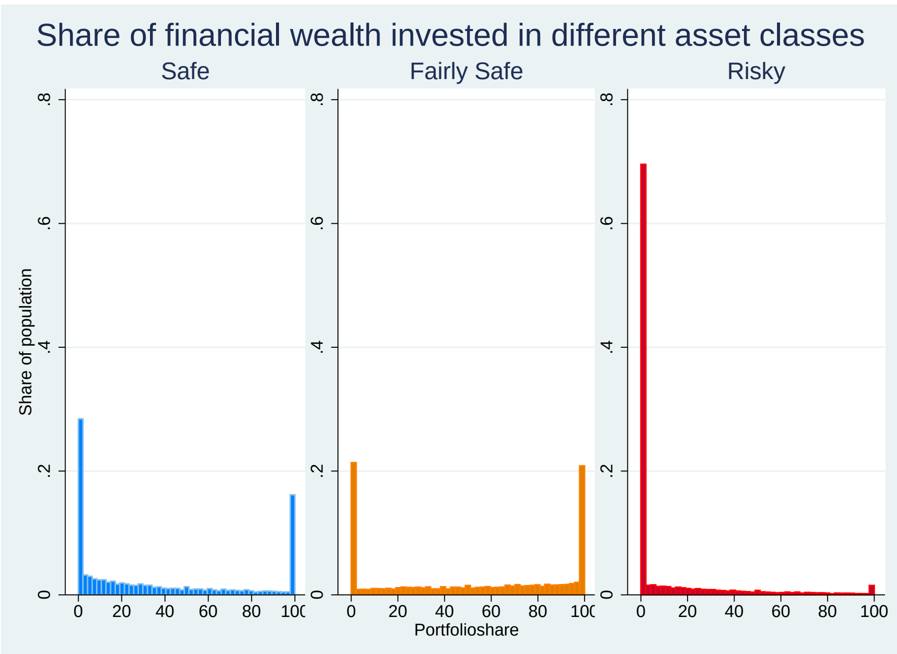


Figure 3.1: Distribution of Shares for Risk Classes. Source: Own computation using SAVE weights.

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 I have emphasized in Section 3.3, the prevalence of such corner values does not affect the validity of my modeling strategy as I am mainly interested in making statements about the conditional mean of shares.

3.4.5 Explanatory Variables and Descriptive Analysis

At this point I give a descriptive overview over the explanatory variables for the regressions. I use a variety of well-established covariates like the ones used in the studies in Guiso et al. (2002).

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 3.4: Regressor Description. This table outlines the definition of the variables used in the regressions.

I control for the wealth and income levels of a household as well as self-assessed risk aversion. In addition, I 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, I account for education level and employment status. Finally, I 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 3.2. In Table 3.4 I provide the definition of each covariate while Table 3.5 gives the descriptive statistics for the covariates.

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 3.5: Summary Statistics. This table gives descriptive statistics for the main covariates in the estimation sample. Source: Own computation using SAVE weights.

From this one 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.
Before I 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 me a good first impression of the relationships in the data.

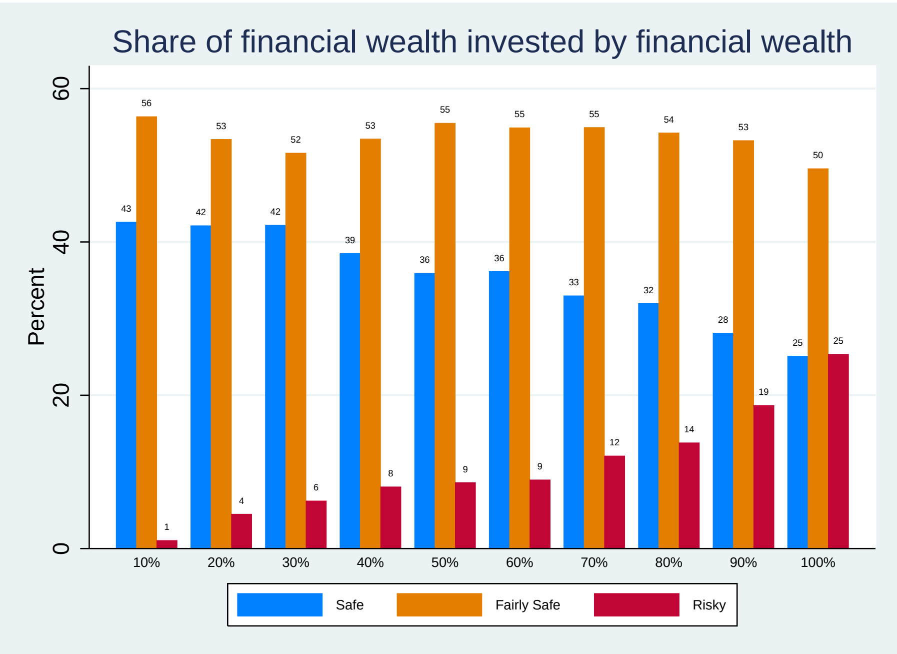


Figure 3.2: Distribution of Shares by Deciles of Financial Wealth. Source: Own computation using SAVE weights.

Figure 3.2 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

[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 my opinion, the most plausible option.

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 %). From the impression in Figure 3.2 I conclude that using a log transformation as is often done in empirical research on risky asset shares is probably not appropriate here. Instead, I 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, Hochgürtel and van Soest (2002) and Banks and Tanner (2002).[24]

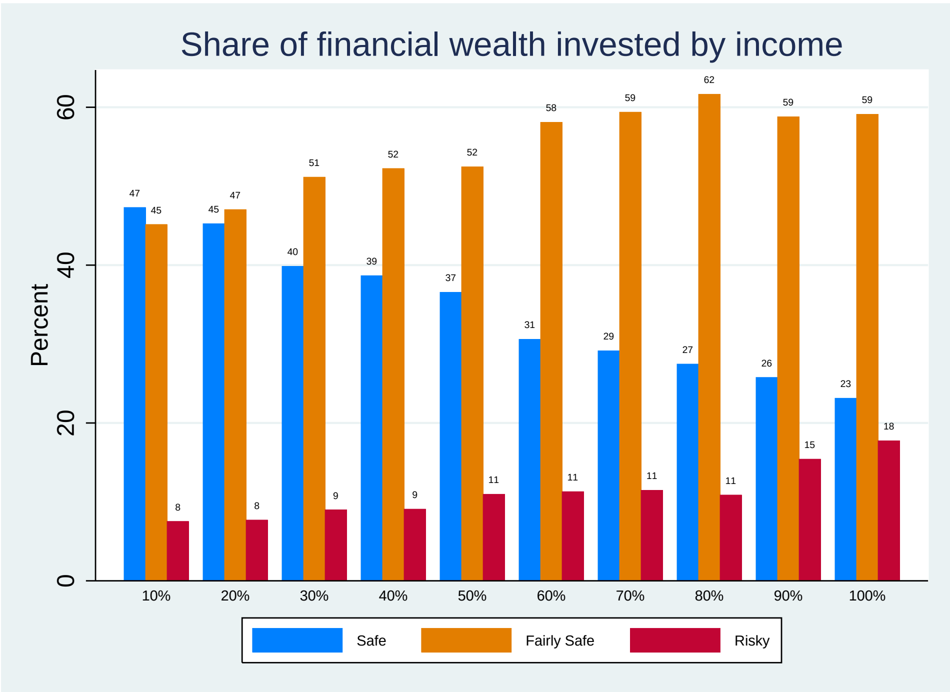


Figure 3.3: Distribution of Shares by Deciles of Income. Source: Own computation using SAVE weights.

Looking at the evolution of the portfolios over income deciles in Figure 3.3 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

[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 my opinion, the most plausible option. I also find that using finer quantiles such as deciles does not add to the explanatory power of the model.

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 I use the same modeling strategy for household income as for wealth.

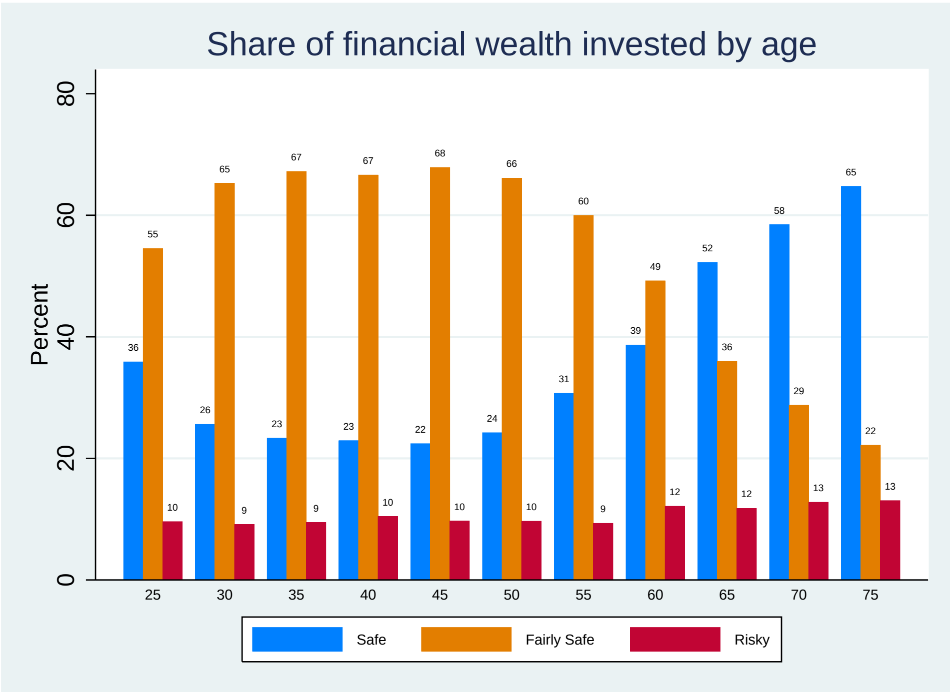


Figure 3.4: Distribution of Shares by Age Groups. Source: Own computation using SAVE weights.

The change of portfolio shares across age is given in Figure 3.4. One 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, one observes 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 safe assets" I introduce age with quadratic terms in my model even though this is probably less suitable for the share of "clearly risky" assets.

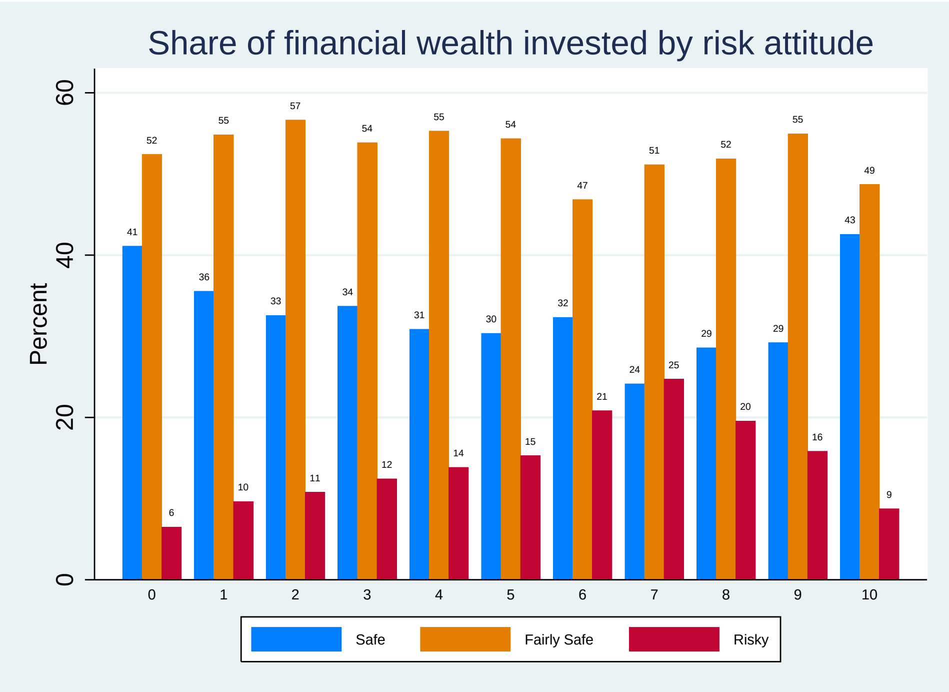


Figure 3.5: Distribution of Shares by Risk Attitude. Source: Own computation using SAVE weights.

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 3.5 I 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

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

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, I use a dummy for a score higher than 0 in the regressions as I think it gives more reliable information regarding the true risk aversion of the sample.

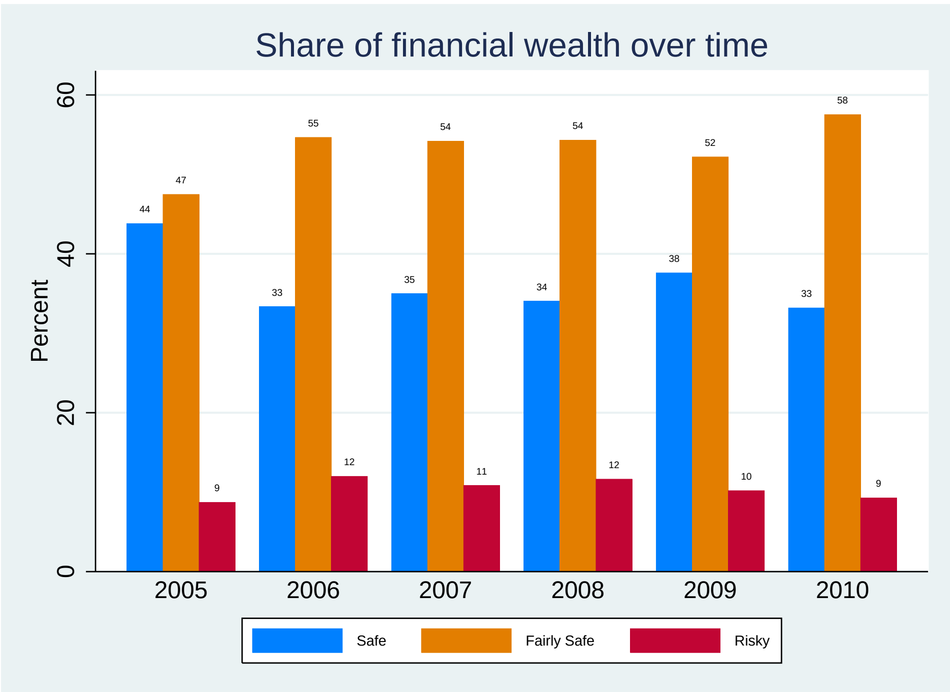


Figure 3.6: Distribution of Shares by Years. Source: Own computation using SAVE weights.

Finally, before I turn to the regression results I 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 3.6, I do not find evidence for this 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.

3.5 Empirical Results

In the following, I present the regression results for the multivariate regression models introduced in Section 3.3. As noted there, the estimated coefficients in these models do not give me much information about the effect of a variable on the means of shares. Thus, I mainly report the computed average partial effects as described in Section 3.3.[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 3.2 and described in Subsection 3.4.5. Year dummies are included in each regression but the corresponding effects are not reported to conserve space.

3.5.1 Pooled Model

Table 3.6 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.

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" assets the turning point is the same but the relationship is reversed - for older investors the

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

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	

Table 3.6: Marginal Effects for Pooled Model. 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. Source: Own computation.

negative effect dominates. These findings correspond to the shape of the age curves for these two asset classes that can be seen in Figure 3.4. As hypothesized in Subsection 3.4.5 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 one does see much smaller effects. I 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. I find that households with a male head allocate on average 1.6 % more of their funds to equities even as I control for risk-attitudes and expectations of investors.[27] For the remainder of the portfolio I 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 I cannot be sure that the person answering the questionnaire is also the one making the financial decisions in all households.

For education I 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. I 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 3.2 in Section 3.4 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 3.4.5. 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 my 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 I have noted in Section 3.2. However, my 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 I 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 3.2.

In accordance with the reasoning in Section 3.2 I 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 I 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. My 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 my results I 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 I 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, I 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 - one can 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.

3.5.2 Random Effects Model

In the following I present the results for the fractional multinomial logit model where I allow for unobserved heterogeneity via maximum simulated likelihood as previously described in Subsection 3.3.4. The statistical inference of the model is based on bootstrap resampling. Specifically, I compute fully robust cluster-corrected standard errors on the basis of 500 bootstrap samples. I 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. One thus can reject the null hypothesis of no random effects on any conventional significance level.

The marginal effects for the random effects fractional multinomial logit model are presented in Table 3.7. Notably, the marginal effects are very similar in magnitude to those for the pooled

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	

Table 3.7: Marginal Effects for Random Effects Model. 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. Source: Own computation.

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 I observe this only for some of my regressors. This is likely due to the fact that I have also made the random effects model fully robust to any departures from the standard assumptions regarding the error term.

Due to these similarities I 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 I have noted in Section 3.3 the main reason for introducing panel data methods is to account for correlations of the time-constant household-specific effects and the covariates. Hence, I rather focus on the CRE model in the next subsection.

3.5.3 Correlated Random Effects Model

Finally, I turn to the results of the correlated random effects framework. I 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 3.3. 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 I report in Table 3.8 only the marginal effects for those variables for which I also included their time-averages in the regression. I focus on households which have been observed for at least three years to ensure enough variation over time. However, my results are quite robust to different sub-samples.

Looking at Table 3.8, I find that most effects are no longer significant once I control for potential correlations with unobserved heterogeneity. This is not an uncommon finding for CRE appli-

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

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	

Table 3.8: Marginal Effects for Correlated Random Effects Model. 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. Source: Own computation.

cations 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 I 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. I find that the marginal effect (3.8 %) is almost identical to the ones found in Subsections 3.5.1 and 3.5.2. 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, I 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 I account for household-specific characteristics. It exhibits the anticipated positive sign while its magnitude is somewhat diminished compared to before. I 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 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.

I find that wealth quartiles exhibit by far the strongest effects within the CRE framework which is in accordance with my 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 3.5.1 and 3.5.2. This difference in the magnitude

[29]Table A.1 in the appendix illustrates this by listing the overall, between and within variation of the dependent and independent variables.

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 3.5.1 and 3.5.2 where I 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 3.2.

When looking at the effect of the means of wealth quartiles over time (given in Table A.2 in the appendix) I 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 I discount the effect of the household-specific characteristics I 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, I find that the effects of most variables become quite small and insignificant when I 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 3.5.1 and 3.5.2. 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.

3.6 Conclusion

This chapter explores the determinants of the risk structure of financial portfolios of German households. To this end I 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. I 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 (2014). Furthermore, I utilize the panel dimension of the SAVE survey to account for unobserved heterogeneity across households. In this way, I control for potential confounding effects on the explanatory variables which might otherwise influence the results. I 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.

My 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 me to get a more complete picture of household portfolios. For instance, my 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 I control for household specific effects via a correlated random effects approach I 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, I 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 I 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 I have outlined in Section 3.3, 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 my application can be regarded as unproblematic. Nevertheless, in the spirit of Mullahy (2011) and Murteira and Ramalho (2014) it might still be interesting to employ models that allow for more general features of the conditional share distribution in future research.

Appendix A: Additional Tables

Variable	$\sigma_{overall}$	$\sigma_{between}$	σ_{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

Table A.1: Overall, Between and Within Variation invested in different financial asset classes. 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. Source: Own computation using SAVE weights.

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)

Table A.2: Marginal Effects for Time Averages. 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. Source: Own computation.

Chapter 4

Counterfactual Analysis of Regional Wealth Differentials in Germany

4.1 Introduction

The 9th November 1989 went down as a watershed event in the history of Germany. After 40 years of separate statehood following the Second World War, the fall of the Berlin Wall, the symbol of separation between East and West Germany, ushered in a new era. The chain of events following this date culminated about one year later, on 3rd October 1990, in the German reunification, when the former communist German Democratic Republic (GDR) joined the Federal Republic of Germany (FRG) in German unity. The expressed goal of politicians at the time was to rapidly bring living standards for citizens of the former GDR on par with their western compatriots (in this context, chancellor Kohl expressed his now famous vision of impending "blooming landscapes" in East Germany). However, even today, after nearly 25 years of German unity, vast differences in terms of economic well-being remains a reality between the two parts of the country. While differentials in wages, household incomes or unemployment rates between both parts of Germany have been studied fairly well (see, for instance Biewen, 2001; Fuchs-Schündeln, Krueger and Sommer, 2010), the analysis of differences in wealth levels is still at its beginning. The main reason for this situation is the usually more problematic data collection process for personal wealth compared to other measures of economic well-being. Respondents are often unwilling to give a detailed account of their wealth situation or are simply not able to assess it accurately. As Davies and Shorrocks (2000) point out, household surveys are less reliable for questions on wealth than for those concerning income. Nevertheless, several recent studies have been looking into the distribution of wealth for Germany, at least on a descriptive level (Hauser, 2010; Frick and Grabka, 2010; Grabka, 2014).

Why does the asset situation of Germans matter in the first place? If one is interested in the overall well-being of a group of people, assessing their financial situation is of key importance. Wealth provides utility for members of a household in a number of different ways:[1] First of all, wealth can yield income streams through interest or dividends on capital investments. It can also provide direct utility - for instance, through owner-occupied housing. Moreover, precautionary wealth can ensure financial stability by serving as a buffer against idiosyncratic risks such as negative income shocks. In addition, a person can achieve a high social status as well as economic and political power by accumulating wealth. Finally, a high net worth enables one to care for one's children via gifts or bequests.

Against this backdrop it is important to assess the factors that contribute to the still prevailing differences in wealth levels and wealth inequality between East and West Germany.[2] There are several institutional reasons that can explain the lower levels of wealth in East Germany today.[3] Probably the most important factor is the strongly restricted opportunity for wealth accumulation in the former GDR. Namely, private ownership of business assets and real estate wealth were largely prohibited. Thus, only few eastern households owned these kinds of assets at the time of reunification. Furthermore, while wages and pension payments where converted to Deutsche Mark (DM) on a one-to-one basis, this was not true for notable financial wealth.[4] Once households where legally able to acquire ownership on named assets, it was often not feasible for them to do so due to the quickly deteriorating financial situation of many eastern

[1] See, for instance, Grabka and Frick (2007) and Frick and Grabka (2010).

[2] Grabka (2014) discusses this wealth gap as part of a special issue of the DIW Wochenbericht series on the socio-economic reality in Germany 25 years after the fall of the Berlin wall.

[3] See Hauser (2010), Frick, Grabka and Hauser (2010) and Frick and Grabka (2010) for more background information.

[4] Financial wealth of more than 6,000 DM where converted on a two-to-one basis only.

households in the face of mass unemployment and starkly declining wage levels. A more subtle point in this respect is the social learning process. Children acquire attitudes towards saving behavior and the ownership of certain assets in large parts from their parents. This might have helped to perpetuate negative attitudes towards wealth accumulation in the former GDR. It is therefore of interest how these very different surrounding conditions continue to shape the levels of wealth accumulation in these two regions today.

With this in mind, an important question is how much of the observed wealth difference between the two parts of Germany can be explained by observable differences in factors that are associated with higher wealth levels such as income, educational attainment and age. Previous studies tried to decompose differences in wealth levels between certain groups. For instance, Cobb-Clark and Hildebrand (2006) and Bauer, Cobb-Clark, Hildebrand and Sinning (2011) look into the determinants of the wealth gap between the native population and immigrants in the United States and Germany, respectively. Both studies find that the adverse financial situation of immigrant households, compared to native households, can in large parts be explained by their unfavorable socio-demographic make-up. Specifically, migrant households are on average younger and less educated compared to native households in these countries. Sierminska, Frick and Grabka (2010) assess the difference in personal wealth between men and women in Germany. They find that the on average lower wealth levels of females are largely determined by their lower levels of income and labor market experience. In the following I try to decompose the wealth gap between East and West German households in a similar fashion.

The rest of the chapter is structured as follows. In Section 4.2 I describe key aspects of the wealth distribution in Germany as well as the findings of previous research of the wealth situation of individuals and households in East and West Germany since reunification. Summary statistics for per captia net wealth and potential wealth determinants in my sample are presented in Section 4.3. Next, Section 4.4 covers the foundations of my decomposition method via a reweighting approach. In Section 4.5 I present the empirical results of my decomposition analyses. Finally, Section 4.6 concludes.

4.2 Wealth in Germany

In order to analyze the wealth gap between East and West Germany, one first has to define what one means by "wealth". First of all, it is decisive whether one is interested in the wealth positions of individuals or households. Most surveys ask for information on asset and debt positions only on the household level. Consequently, most studies focus on household wealth. However, even if data on personal wealth is available, it is unclear how to disentangle the wealth position of individual household members as they are unlikely to take financial decisions independently of each other. Frick and Grabka (2010) touch upon this problem by stressing that it comes down to how one sees the implicit redistribution within a household. If one does focus on individual wealth, one makes the implicit assumption that redistribution does not occur on the household level. Looking at the average wealth within a household, on the other hand, implies that personal wealth is completely redistributed on the household level. Thus, the former type of analysis will typically unveil higher levels of wealth inequality compared to the latter one. In reality, the true level of redistribution between household members will likely lie somewhere in-between these two extremes. Here, I will focus on per capita net wealth as this is the more conservative measure and I believe it to be the more meaningful level of observation for this type of analysis.

Usually wealth studies concentrate on disposable wealth which Davies and Shorrocks (2000) define as nonhuman wealth minus debt. Human capital arguably constitutes the largest fraction of wealth for many individuals, especially for younger ones. However, it is notoriously hard to measure with any degree of accuracy. Thus, human capital is generally excluded from the definition of wealth. A similar argument applies to the expected value of vested pension rights and other social security wealth. In a country with a generous welfare state, such as Germany, these entitlements probably make up a large part of the wealth of a typical households. At the same time, the cash value of future entitlements is afflicted by uncertainty with respect to appropriate discount rates, risk adjustment factors and the like. For this reason, most wealth surveys do not ask for these wealth positions.[5] Another point in this regard is that the aforementioned benefits

[5] In Frick and Grabka (2010) an attempt is made to estimate the impact of entitlements to government transfers on the distribution of wealth in Germany. They find that accounting for these claims lowers the inequality by about 20 % as they are very evenly distributed across the population.

provided by wealth do not apply to human capital and social security wealth in the same way as for financial and housing wealth. The reason for looking at wealth net of debts is that it is more informative of the well-being of household members compared to gross wealth. Consider, for example, gross housing wealth which would portray the financial situation of a person too positive if one did not take into account the mortgage debt that usually accompanies such an investment.

In order to analyze the wealth gap in a sensible fashion, one must first have a solid understanding of the key properties of such a distribution. Both Davies and Shorrocks (2000) and Jenkins and Jantti (2005) provide a good summary of the stylized facts of wealth distributions. First of all, the distribution of net wealth is generally much more unequal than the distribution of income.[6] For most people, wealth is the result of an accumulation process of income streams over time. Thus, the prevailing income inequalities are also added up and intensified so that the resulting wealth distribution is much more spread out. Moreover, the distribution of wealth is typically right skewed with a very fat right tail due to a few individuals owning a disproportionally large fraction of the total wealth. Leaving out such households due to sample variation can significantly affect the sample composition with respect to the distribution of wealth. Hence, it is important to oversample rich households in surveys on household wealth. Furthermore, financial wealth is usually much more unequally distributed compared to non-financial assets - at least if owner-occupied housing constitutes a significant fraction of wealth. Finally, many household members never accumulate much wealth so that one usually observes a high number of households with non-positive net wealth. Specifically, most wealth distributions feature a pronounced mass-point at zero net wealth as a large fraction of households report exactly zero net worth.

As Jenkins and Jantti (2005) point out, these stylized facts have important implications for the modeling of the wealth distribution. First of all, it should be obvious that the mean is generally not a very informative moment for such a spread-out and skewed distribution. Thus, it is not sufficient to estimate only the mean wealth difference, for instance via an Oaxaca-Blinder

[6] In Germany wealth inequality, as measured by the Gini coefficient, is about double as high as income inequality (see Hauser and Stein, 2006).

decomposition. To get a more complete picture of the wealth gap, one should investigate different parts of the distribution - e.g. by looking at several quantiles such as the median. Additionally, the high number of households with zero or negative net wealth suggests looking at the proportion of household members with non-positive wealth levels. The often found spike-at-zero means that smoothing techniques, like kernel-density estimation, are usually not feasible. In Section 4.4 I will lay out how one can analyze the wealth gap along the entire distribution taking into account the peculiarities of the wealth distribution.

4.2.1 EVS

Early studies on the distribution of private wealth in Germany mainly relied on the Income and Expenditure Survey (Einkommens- und Verbrauchsstichprobe: EVS) a microcensus of 1 % of the German population surveyed every five years since 1964 by the Federal Statistical Office of Germany. One of the first studies looking into the financial and asset situation of German households in detail is Börsch-Supan and Essig (2002). The authors focus on the composition of household portfolios for the time period from the 1970s to the mid 1990s. When comparing the situation for East and West Germany, Börsch-Supan and Essig (2002) use the 1993 wave of the EVS. They find that households in East Germany posses much lower levels of wealth compared to those in the West and that it is distributed more unevenly. They speculate that part of this higher inequality might be due to a small group of profiteers of the transition process. The authors also highlight that the wealth distribution of the youngest cohort in the sample is much more similar between the two regions which might reflect the different opportunities to accumulate wealth before reunification. Moreover, the average portfolio composition differs widely between the two regions: on average, eastern portfolios are much less diversified than their western counterparts, feature almost twice the share of wealth in safe assets (43 % to 22 %) and are much less likely to contain real estate (28 % to 51 %).

Hauser (2010)[7] uses the EVS to compare the wealth of households in East and West Germany in the time directly following reunification from 1993 to 2003.[8] He finds a trend of convergence between East and West Germany in terms of inequality and level of disposable household wealth for this time period. Nonetheless, households in the eastern part of reunified Germany lag their western counterparts considerably in both income and wealth levels. In 1993 the net worth of an average household in the East was 36,400 € while this figure was 125,400 € for the West. Thus, the average eastern household commanded only about 29 % of the wealth of the average household in West Germany. Until 2003 these figures increased notably for both parts of Germany to 59,600 € in the East and 148,800 € in the West. Thus the relative wealth level of eastern households increased to 40 % over this time period. In the same fashion, one could see a convergence in wealth inequality as measured by the Gini coefficient. In 1993 East Germany featured a much higher wealth inequality with a Gini coefficient of 0.72 - compared to 0.63 in the West. In 2003 these figures had aligned: while the Gini in the East declined to 0.68 it increased for the West to 0.67. Somewhat surprisingly, while the income inequality is lower in the East than in the West, the wealth inequality is actually higher in the former GDR. Fuchs-Schündeln et al. (2010) use the same data but also look at the financial wealth of households. They find that inequality, as measured by the Gini coefficient, increases for both net total wealth and financial wealth over this time period and are roughly the same in 2003 (about 0.7). However, the increase is more pronounced for net financial wealth which started from a much lower level in 1978 (0.56 vs. 0.64).

4.2.2 GSOEP Wealth Module

The EVS exhibits several methodological shortcomings which makes it less than ideal for the assessment of the wealth situation of Germans according to Frick et al. (2010) and Grabka (2014). The main weakness of the EVS, in regard to the analysis of wealth, is that it is top-coded, i.e. it does not contain high-income households (those with a monthly household income of 18,000

[7] As well as Hauser and Stein (2006) and Hauser (2009).

[8] He also looked into the data for West Germany alone prior to reunification back till 1978 but notes that the data from those waves are not necessarily comparable.

€ or more). High income households are usually also the households with the highest levels of net worth. As mentioned previously, such households make up a large fraction of the national wealth. Thus, leaving out these households will unquestionably distort the observed wealth distribution in the survey. Moreover, the EVS does not ask for business assets which typically make up a large part of the wealth for high net-worth households. For both these reasons Hauser and Stein (2006) suspect that they capture only the lower bound of the inequality and the level of wealth in Germany.[9] Finally, the EVS is not a random sample of the German population but a quota sample which makes it harder to look more closely at certain parts of the population, such as households in East Germany.

As of 2002 many of these issues have been addressed by the German Socio-Economic Panel (GSOEP), the longest running panel study of German households, which is comparable in its design to the Panel Study of Income Dynamics (PSID) for the USA. The survey has been established in 1984 as a representative sample of approximately 4,500 West German households and was extended to East Germany shortly before reunification. It is conducted annually by the German Institute for Economic Research (DIW) and is currently in its 30th wave. At present it covers some 12,000 households with about 20,000 individuals. The GSOEP is the most extensive longitudinal micro data set of its kind in Germany and covers a wide range of socio-economic variables such as employment status, income sources, education level or attitudes towards different aspects of life. Since 2002 the GSOEP survey also collects data on a household's balance sheet in addition to its core questionnaire. Participants are asked about the amount of money invested in five different asset categories as well as their level of debt.[10] The wealth module is asked for every five years and currently encompasses three years (2002, 2007, 2012). Also since 2002 high income households are being oversampled in the GSOEP which leads to a large number of high net worth individuals in the survey. For more information on this process, see Schupp, Frick, Goebel, Grabka, Groh-Samberg and Wagner (2009).

[9] Also, they set negative wealth levels to zero when computing the Gini coefficient which further decreases the inequality measure.

[10] More specifically the survey asks for two types of housing assets - owner-occupied and other. Moreover, the levels of financial wealth, private insurance assets, business wealth and tangible assets are inquired. Finally, the levels of mortgage debt and other consumer debt are verified.

Since the introduction of the wealth module in the GSOEP questionnaire, the responsible authors have published several descriptive papers outlining the wealth distribution in Germany as well as its development over time. Mostly these studies focus on net wealth on the individual level as the possibility to study wealth on this level is a unique element that distinguishes the GSOEP from many other surveys on the topic. Grabka and Frick (2007) describe the main results of the first wealth study in the 2002 wave of the GSOEP. Frick et al. (2010)[11] extend the analysis to the 2007 wealth questionnaire and focus on the change in the wealth distribution over this time period. Furthermore, their book gives a detailed account of the wealth situation in Germany prior to the turn of the century as well as an extensive overview of the wealth module itself.

The work by Frick and Grabka (2010) is interesting in the context of my analysis as it also covers the differences in the wealth distribution between East- and West Germany for 2002 and 2007. They find that in the 2002 GSOEP sample, the estimated average household wealth is about 10,000 € higher than the equivalent amount estimated via the 2003 EVS sample. Likewise the Gini coefficient is almost 10 percentage points higher (0.76) - the same applies to the share of wealth accrued by the top 10 %. More generally they find that the wealth distribution is much more spread out than expected. Thus, they hypothesize that previous studies based on the EVS have underestimated both the overall level of wealth as well as its concentration in the past. This is in line with the shortcomings of the EVS mentioned above. To compare the wealth distribution in East and West Germany, the authors resort to individual net wealth which is much lower than household net wealth and even more unequally distributed. Looking at per capita net wealth, the authors find that the average per capita net wealth in the West in the 2002 GSOEP sample is about 10 % lower than in the 2003 EVS sample. At the same time the EVS sample features an 18 % higher average wealth level for East Germany compared to the GSOEP. Thus, the wealth gap between the two countries is larger than previously thought.

Finally, Grabka and Westermeier (2014) look into the change of the distribution over all three sample years. They find that the overall structure of the wealth distribution in Germany remains largely unchanged between 2002 and 2012. The same holds true for the wealth gap between East

[11] See also: Frick and Grabka (2009a) and Frick and Grabka (2009b).

and West Germany. For instance the share of average individual net wealth in the East compared to the West is 41 % (90,000 € to 36,700 €) in 2002 and increases only marginally to 44 % (94,000 € to 41,000 €) until 2012. Also the Gini coefficient is about four percentage points higher in the former GDR over this period of time. Hence, there does not seem to be a narrowing of the wealth gap between East and West Germany since the turn of the century and the overall level of the gap has remained pronounced for more than 20 years after reunification.

4.3 Data

4.3.1 Summary Statistics for Net Wealth

In the following I start my analysis by looking at the descriptive distribution of wealth in the three available wealth modules in 2002, 2007 and 2012. My sample consists of roughly 20,000 adults per year (about 22,000, 20,000 and 18,000, respectively) of which about a fifth lives in East Germany. Table 3.1 illustrates the distribution of per capita net wealth for Germany as a whole over these years.

I find, in line with the results by Grabka and Westermeier (2014) on individual wealth, that the structure of the wealth distribution remains largely unchanged over the observed time period. Most statistical measures change only modestly from 2002 to 2012. It is striking, that for the lower part of the distribution up to the 30 % quantile one sees a trend of decreasing or at least stagnating wealth levels. Starting from the 40 % quantile, one can see a modest increase in the average net worth per household over time. Even though these contrasting developments in the lower and upper parts of the distribution suggest an increase in inequality over time, this is not reflected in the inequality measures. The Gini coefficient as well as different quantile ratios increase only modestly, if at all. All of this is also in line with the findings by Grabka and Westermeier (2014). It is also noteworthy that the wealth levels decrease from 2002 to 2007 for each quantile only to surge again in 2012. As just mentioned, for quantiles above the 30 % quantile net worth ends up even higher than in 2002. It is not clear what might have caused this dip in 2007 as the financial crisis had not even begun at that time. Moreover, Germany did not experience a substantial rise in home equity in the years preceding the crisis and thus there was no potential for a strong deterioration in housing prices. The ensuing drop in the German stock market could only affect relatively few Germans as most do not participate in this asset class.

Stats	**2002**	**2007**	**2012**
Mean	76,483	78,689	82,770
p1	-18,040	-20,228	-21,569
p5	-2,519	-3,585	-3,180
p10	0	0	0
p20	263	0	0
p25	2,075	1,730	1,967
p30	4,896	4,069	4,845
p40	12,482	11,871	13,732
p50	28,728	25,605	31,344
p60	52,236	48,356	54,850
p70	80,000	77,230	86,750
p75	96,739	94,733	104,280
p80	118,033	118,140	125,550
p90	192,625	198,051	200,327
p95	289,260	300,000	301,754
p99	601,366	680,727	754,295
% Nonpos.	19.30	20.48	20.23
% Neg.	6.68	8.45	8.55
% Zero	12.62	12.03	11.68
Gini	71.28	73.49	71.92
p75p50	3.37	3.70	3.33
p90p50	6.71	7.74	6.40
p75p25	46.73	55.21	53.03
N	22,813	20,728	18,151

Table 3.1: Descriptive Statistics for Net Wealth in Germany. Source: Own computations using GSOEP weights.

The typical characteristics of wealth distributions, as described before, are reflected in Table 3.1. Due to the stability of the distribution over the observed time periods I will focus on the latest year in the following. In 2012 the average German holds net assets of almost 83,000 €. A person at the median, in contrast, is only worth around 31,000 €. Thus, the mean of the distribution is more than 2.5 times higher than the median. This emphasizes the right-skewness of the distribution. At the same time, I find that one in five people in my sample has no or even

negative net worth. About 12 % of sample population report exactly zero net assets which is in line with the often reported spike at zero. How spread out the distribution is can be ascertained by looking at different quantiles and their ratios. For instance, a person at the lower quartile owns around 2,000 € in net assets while the corresponding value for her counterpart at the upper quartile is about 100,000 €. The inter-quartile-ratio, therefore, is roughly 50 - meaning the wealth of someone at the upper quartile is more than 50 times higher than for people at the lower quartile in 2012. The wealth levels at the 90 % and 95 % quantile are about 200,000 € and 300,000 €, respectively. Thus, people in this part of the distribution own 100 times or rather 150 times the net wealth of their compatriots at the 25 % quantile. This high level of wealth inequality is also reflected by the Gini coefficient. It amounts to nearly 72 % for per capita net wealth while the corresponding value for per capita net income is only around 27 %. However, as expected, I find that per capita net wealth is still more evenly distributed compared to individual net wealth as reported by Grabka and Westermeier (2014). Nonetheless, they note that Germany, along with Austria, is the most unequal economy in this regard in the euro area.
I am chiefly interested in the difference between the wealth distributions in East and West Germany. For this purpose, Table 3.2 exhibits distributional measures of the per capita net wealth in both regions. The evolution of the wealth levels for the two regions largely follows that of Germany as a whole, albeit on very different levels. Over the observed time period, both East and West Germany exhibit mostly declining or stagnating wealth levels below the 30 % quantile of the distribution. Thus, many low net worth households in both parts of the country are actually worse of in 2012 compared to 2002. At the same time net wealth increase modestly for most of the upper part of the wealth distribution. Nevertheless, this increase in spread is rather modest and the general shape of the distributions remains largely unchanged as recognized by Grabka and Westermeier (2014).
Looking at the mean net wealth, one sees that the average household member in West Germany in 2002 was on average worth 87,427 €. Until 2012 this amount had increased to 94,088 €. For the East the corresponding numbers are 36,703 € and 41,105 €, respectively. From this I conclude two things: first of all, I observe a very pronounced mean wealth gap between East

Netwealth	**West**			**East**		
	2002	**2007**	**2012**	**2002**	**2007**	**2012**
Mean	87,427	91,671	94,088	36,703	32,233	41,105
p1	-17,033	-20,228	-23,200	-20,673	-20,991	-14,150
p5	-2,331	-3,481	-3,230	-3,113	-3,883	-3,133
p10	0	0	0	0	0	-65
p20	500	119	270	0	0	0
p25	2,617	2,269	2,695	1,010	473	353
p30	5,540	5,065	6,075	2,500	2,014	2,110
p40	16,453	15,842	18,121	5,958	5,860	6,020
p50	38,885	34,208	39,390	12,117	11,287	13,645
p60	64,975	61,382	68,284	22,547	20,483	26,060
p70	93,560	93,384	100,475	38,318	33,416	43,915
p75	111,570	113,612	119,260	47,499	42,281	52,651
p80	135,445	138,399	140,555	58,004	52,081	68,213
p90	213,574	227,970	223,721	99,838	85,456	107,457
p95	322,465	333,063	345,900	144,734	121,341	158,314
p99	658,930	749,185	832,383	285,833	260,600	301,800
% Neg.	6.49	8.23	8.18	7.39	9.23	9.92
% Zero	12.31	11.49	11.03	13.74	13.99	14.07
% Nonpos.	18.80	19.72	19.20	21.13	23.21	23.99
Gini	69.47	71.63	70.51	74.10	74.34	72.99
p75p50	2.87	3.32	3.03	3.92	3.75	3.86
p90p50	5.49	6.67	5.68	8.25	7.58	7.88
p75p25	42.71	50.41	44.45	47.04	89.96	207.89
N	16,956	15,238	13,355	5,857	5,490	4,796

Table 3.2: Detailed Summary Statistics for Net Wealth. Source: Own computations using GSOEP weights.

and West Germany. In 2002 an average West German citizens had almost 51,000 € more at her disposal than a comparable East German. Put differently, an average household member in the East owned only about 42 % of the net asset position of an average person in the West. Second of all, one can see that the relative mean wealth positions do not change much over the years. As can be seen in Table 3.3, the average wealth gap widens slightly to about 53,000 € in 2012. For the median wealth level I find a similar pattern: with about 12,000 € in 2002, the median wealth in East Germany is much lower compared to West Germany where it stands at 38,885 €. East Germans exhibit almost 27,000 € less net worth than westerners, which puts their relative wealth levels at only 31 % of the western levels. Until 2012, the median net worth increased in both regions to 13,645 € and 39,390 €, respectively, while the median wealth gap decreases to roughly 26,000 €. These findings apply more or less to the distribution as a whole. If I examine the distribution of net wealth along different quantiles, I find that eastern households hold less wealth at almost any point of the distribution and that this wealth gap remains relatively constant over time. Due to this distributional stability I will again focus on 2012 if not otherwise indicated.

Stats	**2002**	**2007**	**2012**
Mean	50,724	59,438	52,983
p1	3,640	763	-9,050
p5	782	401	-97
p10	0	0	65
p20	500	119	270
p25	1,607	1,796	2,342
p30	3,040	3,051	3,965
p40	10,495	9,981	12,101
p50	26,768	22,921	25,745
p60	42,428	40,898	42,224
p70	55,243	59,968	56,560
p75	64,071	71,331	66,609
p80	77,441	86,318	72,342
p90	113,736	142,514	116,264
p95	177,731	211,722	187,586
p99	373,097	488,585	530,584
% Nonpos.	-2.33	-3.49	-4.78
% Neg.	-0.90	-0.99	-1.74
% Zero	-1.42	-2.50	-3.04
Gini	-4.63	-2.71	-2.48
p75p50	-1.05	-0.43	-0.83
p90p50	-2.75	-0.91	-2.20
p75p25	-4.34	-39.55	-163.44

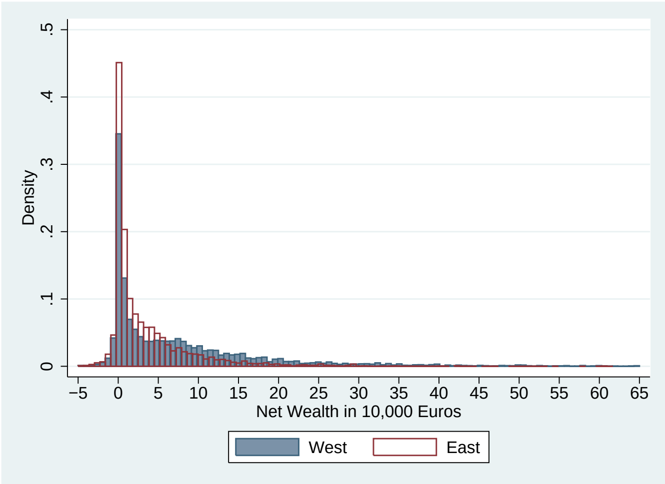
Table 3.3: Absolute Wealth Gap. Source: Own computations using GSOEP weights.

Up till the 25 % quantile one does not observe any noteworthy wealth gap, since household members in this vicinity own similarly low levels of wealth - in fact many own no wealth at all. Above this point, I see a rapidly increasing wealth gap. While western citizens at the 30 % quantile dispose of almost 4,000 € more than their eastern compatriots, this number rises quickly to around 42,000 € at the 60 % quantile and to more than 116,000 € at the 90% quantile. The ratio of eastern to western wealth also increases along the distribution - from 35 % at the 30 % quantile to 48 % at the 90 % quantile.

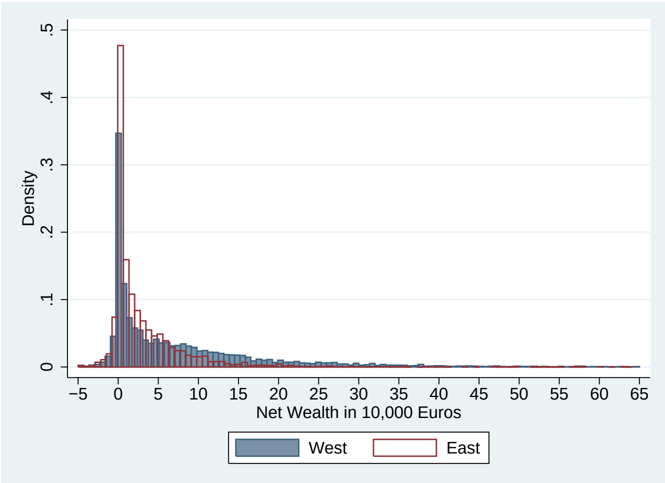
Another way to illustrate the wealth disparities between the two regions is to consider the disproportional low share of total wealth of eastern Germans. While eastern household constitute 21 % of the sample population, they only account for about 10 % of the overall net wealth in the sample. One can again contrast this with net income where the share of East Germans is 19 % and thus much closer to what one would expect under parity. The substantial difference in wealth is also evident when looking at the percentage of household members holding negative or zero net wealth. For West Germany about one in five (19.20 %) individuals hold non-positive levels of net worth while the corresponding number for the East is closer to one in four (23.99 %).

I find, similarly to Hauser (2010), that the distribution of wealth is more unequal in the East than in the West while the opposite is true for the distribution of income. However, the magnitude of this difference is generally not very large. As noted by Grabka and Westermeier (2014), the general inequality does not change much over time. The mostly larger quantile ratios for East Germany indicate that rich citizens in the East tend to accumulate higher multiples of wealth compared to their less fortunate fellow easterners than is the case in West Germany. For instance, the mean-to-median ratio is about 2.4 for West Germany but 3 for East Germany. This suggests a longer right tail for the East. Moreover, the Gini coefficient for East Germany is about 73 % while it is closer to 70.5 % for the West.

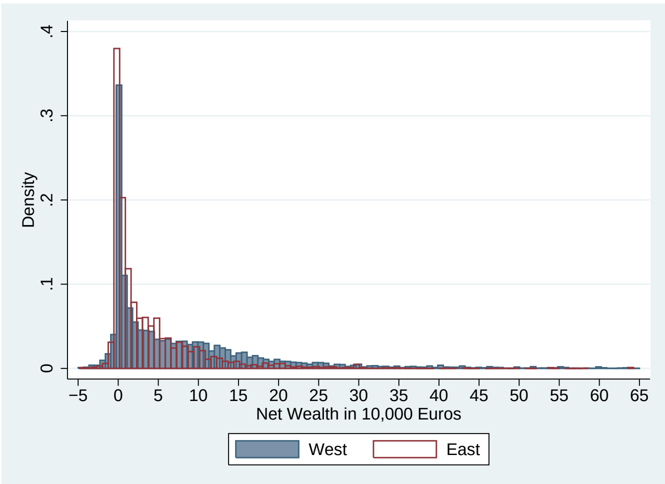
The differences in the distributions of wealth between East and West can also be seen visually. Figure 4.1 shows the histogram for East and West Germany, separately for each sample year. It is obvious that the distribution for East Germany exhibits higher densities at lower levels of net wealth. Especially the peak at zero net wealth is much more pronounced compared to West Germany. Starting from a value of about 50,000 €, the density for West Germany overtakes the eastern distribution. There is not much apparent differences between the distributions in the different years. Looking at the empirical CDFs in Figure 4.2, one can see the wealth gap opens



(a) 2002

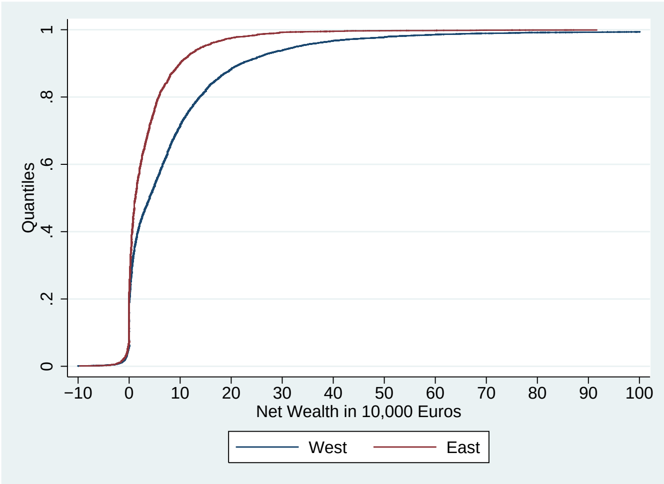


(b) 2007

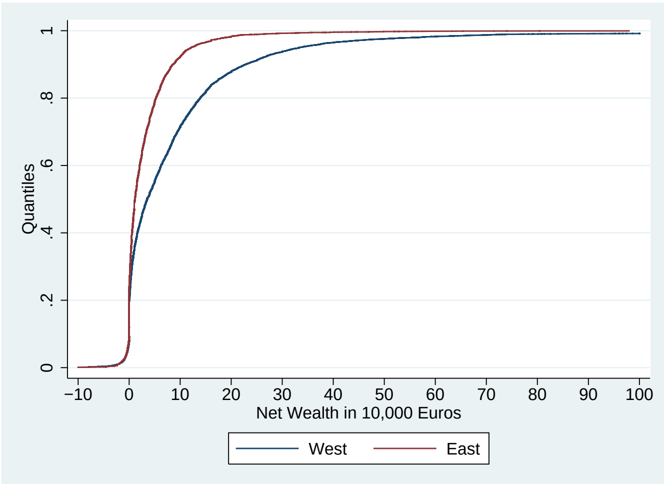


(c) 2012

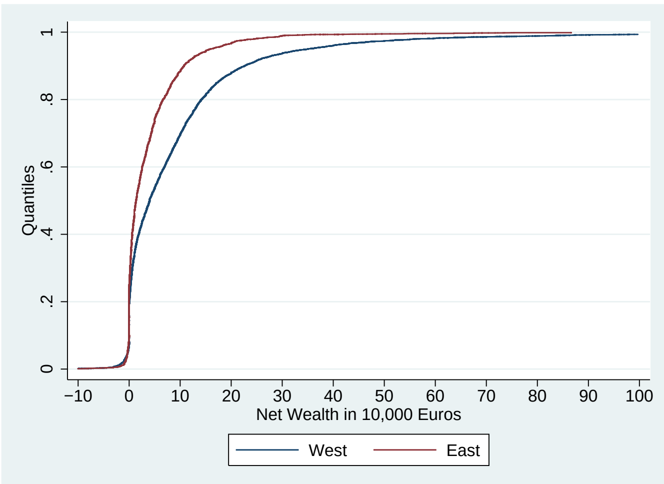
Figure 4.1: Histogram of Net Wealth. East vs. West Germany for each sample year. Source: Own computation using GSOEP weights.



(a) 2002



(b) 2007



(c) 2012

Figure 4.2: Cumulative Density Function of Net Wealth. East vs. West Germany for each sample year. Source: Own computation using GSOEP weights.

after the 20 % quantile. Before that point, the two curves are largely indistinguishable due to the high number of household members with zero wealth in both parts of the country. It is also easy to see, that the gap increases at each quantile.

4.3.2 Composition of Net Wealth

Besides looking at the distribution of per capita net wealth itself, it is also insightful to assess how total net wealth is composed. As already mentioned, the GSOEP questionnaire asks for information on six kinds of assets as well as three different types of debt. The exact definitions of these wealth components are given in Table 3.4.

Wealth Component	**Definition**
House Own	Gross wealth held in the form of owner-occupied real estate
House Other	Gross wealth invested in other types of real estate
Financial Wealth	Gross wealth held in the form of savings accounts, bond, stocks and the like
PandI Wealth	Gross wealth held in the form of life insurances, building loan contracts, private pension schemes and the like
Business Wealth	Gross wealth held as owner of a commercial enterprise
Tangible Assets	Wealth held in tangible form such as gold, jewelery and the like
Gross Wealth	Total gross wealth: sum of all above categories
Mortgage Own	Mortgage associated with owner-occupied real estate
Mortgage Other	Mortgage coupled with other types of real estate wealth
Other Debt	Any type of debt that is not a mortgage such as credit card debt
Total Debt	Sum of all these debt types
Net House Own	Owner-occupied real estate wealth - associated mortgages
Net House Other	Other types of real estate wealth - associated mortgages
Total Net Wealth	Total gross wealth - total debt

Table 3.4: Definition of Wealth Components. Source: Own computations using GSOEP weights.

Furthermore, Table 3.5 entails information on mean participation rates for these components separately for each region and year. From this one can see, for instance, that around one in two West Germans owns owner-occupied real estate and about one in four westerners holds mortgages associated with such housing wealth. The majority of people in both regions owns some kind of financial assets or insurance and pension wealth while only very few people hold business wealth.

	West			East		
	2002	**2007**	**2012**	**2002**	**2007**	**2012**
House Own	50.62	48.61	52.15	36.36	36.52	39.67
House Other	14.82	15.45	15.93	10.05	10.53	10.39
Financial Wealth	54.73	58.80	59.10	55.40	55.69	53.63
PandI Insurance	61.16	66.19	64.04	62.03	62.47	62.10
Business Wealth	7.74	7.62	8.18	7.31	6.52	7.44
Tangibe Assets	14.56	10.25	11.47	7.95	6.08	4.80
Mortgage Own	25.41	24.22	24.64	18.74	18.24	17.10
Mortgage Other	6.78	7.04	7.03	2.78	3.74	4.07
Other Debt	17.06	23.70	23.65	20.44	27.14	28.43
N	16,956	15,238	13,355	5,857	5,490	4,796

Table 3.5: Mean Participation Rates. Source: Own computations using GSOEP weights.

With the purpose of the decomposition analysis in mind, it seems worthwhile to assert the differences in participation rates for these different wealth components across the country. Looking at Table 3.6, it is striking that the by far largest gap in ownership rates between the two regions is found for owner-occupied housing. In 2012, West Germans had a 12.5 percentage points higher probability to own a self-occupied real estate object than their eastern compatriots.

Stats	**2002**	**2007**	**2012**
House Own	14.26	12.09	12.48
Houese Other	4.77	4.91	5.55
Financial Wealth	-0.67	3.11	5.48
PandI Wealth	-0.87	3.72	1.94
Business Wealth	0.42	1.10	0.74
Tangible Wealth	6.61	4.16	6.66
Mortgage Own	6.68	5.98	7.54
Mortgage Other	4.00	3.30	2.96
Other Debt	-3.38	-3.44	-4.77

Table 3.6: Mean Gap in Participation Rates by Year. Source: Own computations using GSOEP weights.

Ownership of almost all other wealth components is also more common in West Germany as is evident from the positive gap in participation rates. The differences in participation rates are, however, usually not as pronounced as is the case for owner-occupied housing. Non-mortgage private debt constitutes a general exception in this regard. In 2012 East Germans were actually almost 5 % more likely to hold this type of debt compared to westerners.

	West			**East**		
	2002	**2007**	**2012**	**2002**	**2007**	**2012**
House Own	57,934	56,308	60,680	24,660	21,976	27,188
House Other	17,200	18,515	16,485	4,402	3,357	4,599
Financial Wealth	10,891	13,167	14,542	6,326	6,242	7,708
PandI Wealth	9,718	11,889	10,066	5,160	5,636	6,289
Business Wealth	6,341	7,912	7,904	4,207	3,143	4,232
Tangible Assets	1,534	1,110	1087	824	305	338
Gross Wealth	103,618	108,903	110,763	45,580	40,659	50,354
Mortgage Own	9,567	10,372	10,755	5,357	5,211	5,228
Mortgage Other	4,448	4,407	3,799	1,243	1,188	1,765
Other Debt	2,176	2,453	2,121	2,278	2,028	2,257
Total Debt	16,190	17,232	16,675	8,877	8,426	9,250
Net House Own	48,367	45,937	49,925	19,303	16,765	21,960
Net House Other	12,752	14,108	12,685	3,159	2,169	2,835
Total Net Wealth	87,427	91,671	94,088	36,703	32,233	41,105
N	16,956	15,238	13,355	5,857	5,490	4,796

Table 3.7: Mean Asset Values. Source: Own computations using GSOEP weights.

Another way to look at this issue is to consider the amount of money that accrues to each of the wealth components. Yet, as one has seen in Table 3.5, most wealth components are not widely owned in the population. Therefore, even the median value is zero for most assets and all debt types. Consequently, I only consider mean asset and debt levels in the following. For this purpose, Table 3.7 and Table 3.8 portray the mean values and the mean differences for each sub-category, respectively. It is evident that the mean amount for any category, except for non-mortgage debt, is higher in West Germany. This difference is especially pronounced for home equity which is responsible for large parts of the average gap in net wealth. Out of the mean wealth gap of roughly 53,000 € in 2012, about 53 % (or 28,000 €) are associated with differences in net owner-occupied housing wealth. This is not entirely surprising as home equity constitutes the main asset in the portfolio of the average German. Around 53 % of the net worth of the average investor in both regions is tied to home equity. Still, these findings emphasize the importance of real estate in the portfolio of the average investor and suggest that differences in real estate investments might help to explain large parts of the wealth gap.

Stats	**2002**	**2007**	**2012**
House Own	33,274	34,332	33,492
House Other	12,798	15,158	11,885
Finanncial Wealth	4,564	6,926	6,834
PandI Wealth	4,557	6,253	3,776
Business Wealth	2,134	4,769	3,672
Tangible Wealth	710	806	749
Gross Wealth	58,038	68,243	60,409
Mortgage Own	4,210	5,161	5,527
Mortgage Other	3,205	3,220	2,035
Other Debt	-102	425	-136
Total Debt	7,313	8,805	7,425
Net House Own	29,064	29,171	27,965
Net House Other	9,593	11,938	9,851
Net Wealth	50,724	59,438	52,983

Table 3.8: Mean Gap for Assets by Year. Source: Own computations using GSOEP weights.

4.3.3 Summary Statistics for Wealth Determinants

In the last section, I have established that a substantial wealth gap between the inhabitants of the former GDR and the western federal states remains even almost 25 years since reunification and shows little sign of diminishing. The main objective of this chapter is to decompose this observed wealth gap into a part that can be attributed to observable differences in potential wealth determinants and an unexplained part that is associated with unobservable factors. These unobserved factors could, for instance, be omitted variables or institutional differences due to the different political systems before reunification. For this subsequent decomposition analysis it is paramount to first examine summary statistics of potential explanatory variables of wealth in my sample. I focus on variables that are likely to play an important role for the wealth formation of individuals. With regard to the explanatory power for the observed wealth gap, it is most interesting to assess how much the distributions of these potential wealth determinants differ between East and West Germany. I group the most promising variables into four main categories: permanent net income, variables associated with the average labor market situation of household members, the average education level within a household as well as socio-demographic and family background variables. Table 3.9 presents descriptions of the variables that I include in each of these categories.

Variable	Description
Permanent Income	Average per capita net monthly income over the past 5 years
Exp. FT	Average household full-time working experience in years
Exp. PT	Average household part-time working experience in years
Exp. UE	Average household unemployment experience in years
High Job	Share of household members with the highest possible job autonomy
Selfemp	Share of household members selfemployed
Retired	Share of household members retired
Middle Vocation	Share of household members with vocational training
High Vocation	Share of household members with high level of vocational training (Abitur + Ausbildung / Meister)
College	Share of household members with college degree
Age	Average age of household members
Male	Share of male household members
Married	Share of married household members
Foreign	Share of foreign household members
Number of Kids	Number of Kids in household
HH Size	Household size
Health Problems	Share of household members with serious health problems
Father College	Share of household members with college educated father
Mother College	Share of household members with college educated mother
Ever Inheritance	Any household member ever received an inheritance, gift or the like
High Inheritance	Any household member received an inheritance or the like of at least 25,000 €

Table 3.9: Definition of Wealth Determinants. Source: Own computations using GSOEP weights.

It is well known that income is among the most important determinants of wealth as a high discretionary income enables one to save and invest in different assets. According to the permanent income hypothesis, long-term income is more important in this regard than transitory income. To proxy for permanent income, I take the average over per capita net monthly income over the past five years. I include the average lifetime experience for full-time, part-time and unemployment in the model in order to account for the labor market experience of household members. Furthermore, I control for the share of members of a household that hold very prestigious jobs, are self-employed or retired. To assess the average education level in a household, I include the share of household members who are highly educated (i.e. college graduates) or have some kind of vocational education (i.e. skilled workers or master craftsmen). Important socio-demographic variables in this regard are the average age of adult household members, the share of male, married and foreign household members as well as persons with chronic health problems. For instance, younger household members are expected to possess smaller fortunes as they had only little time to accumulate wealth. Moreover, I control for the household size and the number of kids up to the age of 16 in the household as the need to finance children or other family members restricts the opportunity to accumulate assets. The social background of a person is likely to have a substantial effect on her financial situation, either indirectly via values learned or directly through financial contribution by close relatives. I include the share of individuals in the household with college educated parents to account for these effects. On the other hand I add dummy variables that indicate if any household member ever received any type of bequest and whether this inheritance was substantial. A detailed account of these potential wealth determinants is given in Table 3.9.

Table B.1 in the appendix illustrates summary statistics for these variables for Germany as a whole pooled over all years and is included primarily for completeness as I am chiefly interested in differences between East and West Germany. For this reason, Table 3.10 features the mean levels of the explanatory factors for each region and year separately. Finally, the average differences in the wealth determinants between West and East Germany for each year are given in Table 3.11. I will again focus on the year 2012 since the regional differences remain steady over time as is evident from Table 3.11. From the tables one can see that a conspicuous difference exists for permanent income between the two regions. On average West Germans dispose of 1,375 € net monthly income while the equivalent figure for East Germans is only 1,153 €. This results in an extra 225 € that the average West German has at her disposal. This substantial income gap is likely associated with the observed wealth gap, especially if one keeps in mind that such income differentials have persisted ever since the reunification. Looking at labor market experience, I find that an average household member in both regions spends by far the longest time of her career in full-time employment. Furthermore, East Germans have three and a half years more full-time experience on average - probably due to the higher labor market participation rate of women in the GDR.

Netwealth	**West**			**East**		
	2002	**2007**	**2012**	**2002**	**2007**	**2012**
Permanent Income	1,342.26	1,361.16	1,374.71	1,123.16	1,081.38	1,153.14
Exp. FT	16.8	17.18	17.59	20.37	19.92	21.03
Exp. PT	2.51	2.98	3.66	1.84	2.09	2.43
Exp. UE	0.58	0.78	0.82	0.98	1.53	1.83
High Job	0.02	0.02	0.02	0.01	0.01	0.01
Selfemp	0.05	0.05	0.05	0.05	0.05	0.06
Retired	0.26	0.27	0.27	0.31	0.31	0.32
Middle Vocation	0.49	0.48	0.48	0.51	0.51	0.52
High Vocation	0.13	0.13	0.13	0.11	0.11	0.12
College	0.13	0.16	0.19	0.21	0.23	0.24
Years Schooling	11.55	11.83	12.08	12.06	12.24	12.46
Age	48.81	49.62	50.87	49.14	49.87	52.17
Male	0.47	0.47	0.47	0.48	0.47	0.47
Married	0.60	0.56	0.56	0.53	0.49	0.51
Foreign	0.10	0.11	0.11	0.02	0.01	0.01
Number of Kids	0.50	0.42	0.37	0.39	0.30	0.31
HH Size	2.58	2.50	2.46	2.44	2.34	2.22
Health Problems	0.18	0.19	0.18	0.19	0.21	0.20
Father College	0.10	0.12	0.14	0.15	0.15	0.17
Mother College	0.05	0.06	0.08	0.10	0.12	0.14
Ever Inheritance	0.25	0.27	0.27	0.22	0.24	0.24
High Inheritance	0.09	0.11	0.10	0.04	0.04	0.05
N	16,956	15,238	13,355	5,857	5,490	4,796

Table 3.10: Wealth Determinants by Year. Source: Own computations using GSOEP weights.

At the same time the average duration spent in unemployment is double as high as for western household members. While East Germans are unemployed for about 22 month on average, the average West German spends only 10 month in unemployment. This is likely to have a negative effect on the relative financial situation of East Germans since one can hardly build up assets during periods of unemployment and maybe even has to deplete one's savings.

Stats	**2002**	**2007**	**2012**
Permanent Income	221.72	283.07	224.74
Exp. FT	-3.52	-2.77	-3.48
Exp. PT	0.69	0.91	1.26
Exp. UE	-0.41	-0.76	-1.04
High Job	0.01	0.01	0.01
Selfemp	0.00	0.00	-0.01
Retired	-0.05	-0.05	-0.05
Middle Vocation	-0.02	-0.03	-0.04
High Vocation	0.02	0.02	0.00
College	-0.08	-0.07	-0.06
Years Schooling	-0.50	-0.41	-0.38
Age	-0.33	-0.29	-1.34
Male	-0.01	0.00	0.00
Married	0.07	0.07	0.04
Foreign	0.08	0.09	0.10
Number of Kids	0.12	0.13	0.05
HH Size	0.15	0.16	0.24
Health Problems	-0.01	-0.02	-0.02
Father College	-0.04	-0.03	-0.03
Mother College	-0.05	-0.06	-0.05
Ever Inheritance	0.04	0.04	0.03
High Inheritance	0.06	0.07	0.05
N	22,813	20,728	18,151

Table 3.11: Gap for Wealth Determinants. Source: Own computations using GSOEP weights.

I also observe that East Germans are slightly better educated - probably due to deliberate political measures during the GDR intended to increase the college attendance of working class citizens. With regard to the social demographic characteristics the most striking difference is the share of foreign household members. While the share of eastern household members with a non-German citizenship is only 1 %, the respective figure for the West is 11 %. Given that individuals with migrant background tend to hold lower levels of wealth compared to natives, this is bound to have an effect on the distribution across the country. What is more, members of East German households are slightly older and less numerous. Referring to socio-demographic background, I find that households in East Germany are more likely to feature members whose parents are college educated. This can probably be attributed to the same reasons as for the case of the education level of the household members themselves. On the other hand, western households are likelier to exhibit members that have received an inheritance. They are even twice as likely to have received a substantial inheritance (10 % vs. 5 %) which should make them more prosperous. Overall about one in four households in Germany received some form of inheritance or gift.

4.4 Decomposition Analysis

The goal of decomposition analyses is to segment differences in distributional statistics for some variables between time-periods or sub-populations. It allows one to quantify the contribution of observable factors to the distributional differences between the groups under consideration for the variable of interest. In the following, I draw heavily on Fortin, Lemieux and Firpo (2011), who provide an excellent overview of decomposition methods in econometrics. The starting point of any decomposition analysis is the observed difference for a given distributional statistic between the sub-populations of interest. I denote this overall gap as Δ_o^ν, where ν denotes the distributional statistic of choice. Due to its predominant role in econometric analyses, the statistic under consideration will often be the sample mean. In the context of this study I am interested, among others, in the average wealth gap between East and West Germany.

The so-called aggregate decomposition is aimed at partitioning the overall difference into an "explained" and an "unexplained" part. The "explained" part is also called composition effect and denoted as Δ_x^ν. It represents the part of the overall gap that is associated with differences in the observable characteristics x. The "unexplained" part will be referred to as wealth structure effect in this chapter and is labeled Δ_w^ν. It captures the contribution of all unobserved characteristics as well as differences in the conditional expected wealth function, i.e. how individuals in the two regions transfer the observed characteristics, such as income, into wealth. I write the aggregate decomposition in the following form:

$$\Delta_o^\nu = \Delta_x^\nu + \Delta_w^\nu \tag{4.1}$$

It is insightful in itself to assess how much of the overall gap is associated with differences in the covariates as well as how much that share changes along the distribution. Yet, one is usually also interested in isolating the contribution of certain parts of the vector of wealth determinants. Regarding the application to the wealth gap it might, for instance, be interesting to know how much variations in permanent income help to explain disparities in per capita net wealth. This approach is known as detailed decomposition and usually requires stronger assumptions than the aggregate decomposition. It is important to note that the resulting effects should not be interpreted in a causal sense. Rather, they are of a descriptive nature and meant to give an understanding of the magnitude of the contribution of observable characteristics towards the observed difference in the variable of interest. Moreover, they are general equilibrium effects as they do not take into account the potential behavioral changes associated with them (see Biewen, 2014).

When it comes to implementing the decomposition analysis, several methods exist, each which its own advantages and limitations. The most popular decomposition method is the Oaxaca-Blinder composition (see Fortin et al., 2011). It owes its popularity to its ease of implementation via OLS and the fact that it naturally allows for detailed composition. However, this method assumes a linear relationship between the variable of interest and the explanatory factors and only allows for the decomposition of mean differentials. This restriction is problematic when it comes to analyzing wealth distributions. As mentioned before, wealth distributions are generally very spread-out and right-skewed. Thus, the mean of the distribution is usually much higher than the median and not necessarily informative regarding the wealth of a typical person or household. More generally, one should try to get a more complete picture of the wealth gap, for instance by studying the difference at different quantiles of the distribution. Moreover, one is often interested in other characteristics of the wealth distribution apart from its level. For example, large fractions of the population exhibit non-positive net wealth. It is of interest to also examine the difference in this fraction between the groups under consideration. In addition, one is often interested in inequality-measures such as the Gini coefficient or the inter-quartile range. Thus, a technique that allows to decompose different summary measures of the distribution is favorable in this respect.

A method making it possible to decompose disparities along the entire distribution with little assumptions was introduced by DiNardo et al. (1996). They propose a reweighting approach, similar to propensity score matching, where one aligns the distribution of the covariates between the two populations in order to identify the composition effect and the unexplained proportion. They first applied this method to analyze changes in the distribution of wages in the USA from 1973 to 1992. Due to its flexibility, the reweighting approach (henceforth DFL method) has become widely-used ever since (see Fortin et al., 2011). Usually, the DFL method is combined with adaptive kernel density estimation where one uses the reweighting factors during the smoothing procedure. The main advantages of this practice is that one can easily visualize the distributional differences and that it leads to variance-reduction of the estimators. However, Biewen (2001) stresses that the correct application of this smoothing method is quite challenging. He continues by pointing out that one is often mainly interested in assessing the effect on certain summary measures of the distribution such as quantiles or inequality indices which are easily obtained from the unsmoothed distribution itself. Thus, smoothing is not necessary unless one is interested in the graphical display of the differences. As both Bover (2010) and Cowell, Karagiannaki and McKnight (2013) note, there is yet another important obstacle to the application of smoothing methods in the context of analyses of wealth distributions. As mentioned before, it is an empirical fact that most distributions exhibit a marked spike at zero net wealth. Bover (2010) emphasizes that the sensitivity of smoothing methods is exacerbated by the presence of such spikes. For this reason I will follow the same approach and focus on cumulative distribution functions (CDF). This makes it unnecessary to apply smoothing techniques as in the case of densities.

The main disadvantage of the DFL method, however, is that it cannot readily be applied to detailed decompositions. Generally one can only assess the contribution of dummy variables on the composition effect (see Fortin et al., 2011). Yet, Cobb-Clark and Hildebrand (2006) show

how one can extend the original reweighting approach in a straight forward fashion in order to differentiate Δ_x^ν by groups of covariates. Still, one is constrained in the number of segments in which the composition effect can be partitioned. Therefore, one usually confines oneself to certain groups of covariates such as socio-demographic variables or labor market indicators (see Cobb-Clark and Hildebrand, 2006; Bauer et al., 2011; Sierminska et al., 2010). Moreover, these detailed effects exhibit sequential ordering. Thus, a Shapley decomposition approach, where one averages over all possible sequences, has to be employed.[12]

I follow the course of action of the aforementioned papers and restrict myself to reweighting the majority group (in my case individuals situated in West Germany) with the characteristics of the comparison group (household members in East Germany). This is necessary due to compressed value ranges of certain variables in the East which would otherwise require extrapolation of values and potentially lead to extreme reweighting factors for certain observations. Hence, one only uses those sequences in which the main group is reweighted and averages over all the resulting sequences. In the following, I describe this decomposition method in more detail.

4.4.1 Aggregate Decompositon

I start out by explaining the mechanics of the aggregate decomposition via reweighting. In order to decompose the wealth distribution between East and West Germany, one has to control for observable factors that might help to explain the wealth differentials between the two regions. Therefore, I am mainly interested in the conditional distribution of wealth given these factors. Remembering that the unconditional distribution can be written as the integral over the conditional distributions weighted by the density of the conditioning factors, one can write the unconditional distribution for the average wealth of individuals in West Germany as:

$$F_{11} = F(w|W=1) = \int\limits_x F(w|x, W=1) \cdot dF(x|W=1) \tag{4.2}$$

Here, wealth is denoted by a lower case w while the indicator for residency in West Germany is given by a capital W. Individuals living in West Germany are indicated by 1 while those situated

[12] See Shorrocks (2013).

in East Germany are labeled by 0. $F(w|x, W = 1)$ is the conditional expected wealth function for individuals belonging to West German households while $dF(x|W = 1)$ is the distribution of the wealth determinants in the same region. Consequentially, F_{11} denotes the distribution of wealth that prevails if both the wealth function and the distribution of observable factors are as in West Germany.[13] In the same fashion, the observed distribution for East German household members can be written as:

$$F_{00} = F(w|W = 0) = \int_x F(w|x, W = 0) \cdot dF(x|W = 0) \tag{4.3}$$

Now one can easily compute the overall difference in wealth between the two regions for a specific summary measure ν by taking the difference between the corresponding statistics for the two observed distributions: $\Delta_o^\nu = \nu(F_{11}) - \nu(F_{00})$. To decompose this difference, one needs to estimate a corresponding counterfactual distribution first. The counterfactual distribution of wealth for West Germany is the distribution one would obtain if the conditional wealth function remained as in the West whereas the distribution of observable characteristics was as in the East. This counterfactual distribution is given by:

$$F_{10} = \int_x F(w|x, W = 1) \cdot dF(x|W = 0) \tag{4.4}$$

It is not immediately obvious how one can compute such a counterfactual distribution. DiNardo et al. (1996) argue that a straight forward way to obtain the counterfactual distribution is to reweight the distribution of covariates in the majority group in such a way that it matches the corresponding distribution found in the comparison population. This can be seen by rewriting the counterfactual for the net wealth distribution in West Germany as:

$$F_{10} = \int_x \left[\frac{dF(x|W = 0)}{dF(x|W = 1)} \right] \cdot F(w|x, W = 1) \cdot dF(x|W = 1) \tag{4.5}$$

[13] Here, I make use of the following equality for the distribution of the covariates: $\mathrm{d}F(x|W = 1) = f(x|W = 1)\mathrm{d}x$.

One can see immediately that this expression is just the observed distribution for the West multiplied by the factor $dF(x|W=0)/dF(x|W=1)$. This term is called the reweighting factor and will be denoted henceforth by $\Psi(x)$. It ensures that the distribution of x is the same in both populations. Therefore, one can write the counterfactual distribution slightly more compact as:

$$F_{10} = \int_x \Psi(x) \cdot F(w|x, W=1) \cdot dF(x|W=1) \tag{4.6}$$

The main question is how to estimate the reweighting factor, which is the ratio of two multivariate density functions. DiNardo et al. (1996) show that the reweighting factor can be written in terms of ratios of probabilities via Bayes' theorem:

$$\Psi(x) = \frac{dF(x|W=0)}{dF(x|W=1)} = \frac{P(x|W=0)}{P(x|W=1)} = \frac{P(W=1|x)}{P(W=0|x)} \cdot \frac{P(W=0)}{P(W=1)} \tag{4.7}$$

In this case, $P(W=1|x)$ is the probability of living in West Germany conditional on the wealth determinants in x. $P(W=0|x)$ gives the same measure for East Germany while $P(W=0)$ and $P(W=1)$ represent the unconditional probabilities of living in East or West Germany, respectively. The probabilities needed for the reweighting factor can be readily estimated from a given sample. For instance, $\widehat{P}(W=0)$ and $\widehat{P}(W=1)$ are just the observed fraction of individuals living East and West Germany in my sample. In order to estimate the conditional probabilities, one can do so either non-parametrically or by imposing more structure on the relationship - usually via a probit or logit model. The non-parametric approach offers the advantage that the relationship between factors can be modeled with a maximum of flexibility. This point is stressed by Barsky, Bound, Charles and Lupton (2002) who use such a reweighting scheme to decompose the wealth gap between black and white households in the United States with respect to household income. They argue that such flexibility is especially crucial in their case due to the unknown relationship between household wealth and income. This relation is likely highly non-linear and thus hard to parametrize. However, any non-parametric method suffers from the well-known curse of dimensionality, i.e. the inability to model the influence of several covariates in such a context. For this very reason, Barsky et al. (2002) do not control for other relevant variables besides household income. This is a major limitation of the non-parametric approach. Therefore, most studies using reweighting decomposition employ parametric specifications (see Fortin et al., 2011). In this chapter, I use logit models to estimate the conditional regional probabilities. Thus, the estimator for the reweighting factor is written as:

$$\widehat{\Psi}_x = \frac{\widehat{P}(W=1|x)}{\widehat{P}(W=0|x)} \cdot \frac{\widehat{P}(W=0)}{\widehat{P}(W=1)} \tag{4.8}$$

With this estimator, I can compute the counterfactual wealth distribution for western individuals if these had the same distribution of characteristics as their eastern compatriots, F_{10}. Now, one can finally put all these concepts together to illustrate the implementation of the decomposition method in Equation 4.1. I write the decomposed wealth gap between West and East Germany for a distributional statistic ν as:

$$\underbrace{\nu(F_{11}) - \nu(F_{00})}_{\Delta_o^\nu} = \underbrace{\nu(F_{11}) - \nu(F_{10})}_{\Delta_x^\nu} + \underbrace{\nu(F_{10}) - \nu(F_{00})}_{\Delta_w^\nu} \tag{4.9}$$

Here $\nu(F_{11}) - \nu(F_{10})$ represents the composition effect because only the distribution of the x vector is changed while the conditional wealth functions are the same. $\nu(F_{01}) - \nu(F_{00})$, on the other hand, is the wealth structure effect since the distributions of wealth determinants are identical but the conditional wealth functions differ from each other. In this fashion, I can assess how much of the observed gap along the wealth distribution can be attributed to the combined influence of observable factors.

4.4.2 Detailed Decompositon

As previously mentioned, one is often not interested in the composition effect Δ_x^ν alone but also in the contribution of specific elements of x to the differences in relative wealth positions. In the context of this study it is reasonable to follow the approach by previous papers on this topic[14] and split the vector of covariates into four sub-vectors: permanent income (y), variables associated with the average labor market situation of a household (l), the average education

[14]See Cobb-Clark and Hildebrand (2006), Sierminska et al. (2010) and Bauer et al. (2011).

level within a household (e) and a vector of socio-demographic background variables (d). All these factors play an important role in the formation of wealth. My aim is to attribute parts of the composition effect to these factors such that $\Delta_x^\nu = \Delta_y^\nu + \Delta_l^\nu + \Delta_e^\nu + \Delta_d^\nu$. As for the aggregate decomposition, I start out by writing the observed wealth distribution in West Germany in terms of conditional distributions with respect to the four subsets of the wealth determinants:

$$F_{11111} = \int_y \int_l \int_e \int_d F(w|y,l,e,d,W=1) \cdot dF(y|l,e,d,W=1) \cdot dF(l|e,d,W=1) \cdot dF(e|d,W=1) \cdot dF(d|W=1) \tag{4.10}$$

Here, the conditional wealth distribution for West Germany, $F(w|y,l,e,d,W=1)$, is identical to $F(w|x,W=1)$ in Equation 4.2 for the aggregate case as $x = \{y,l,e,d\}$. However, unlike before this conditional wealth function is not only weighted by the joint distribution of all covariates in the West, $dF(x|W=1)$. Instead, one has a density for each subset of variables. These weighting factors are ordered in a sequence that is economically plausible. For instance, ones income depends on ones labor market experience, education and general socio-demographic background. Therefore, I consider the distribution of the particular subset conditional on the underlying covariates. In this case $dF(y|l,e,d,W=1)$ is the conditional income function for western household members given their labor market situation, educational attainment and social background. The observed distribution for eastern individuals can be written in an analogous fashion:

$$F_{00000} = \int_y \int_l \int_e \int_d F(w|y,l,e,d,W=0) \cdot dF(y|l,e,d,W=0) \cdot dF(l|e,d,W=0) \cdot dF(e|d,W=0) \cdot dF(d|W=0) \tag{4.11}$$

Because I have split the vector of covariates into several subgroups of covariates, I can theoretically compute counterfactual distributions with respect to any single one of these subgroups as well as any combination of these components. Consider, for instance, the counterfactual distribution of per capita net wealth of West Germans that would occur if everything was as in West Germany but the distribution of permanent income, conditional on the other covariates, was as in East Germany:

$$F_{10111} = \int_y \int_l \int_e \int_d F(w|y,l,e,d,W=1) \cdot dF(y|l,e,d,W=0) \cdot dF(l|e,d,W=1) \cdot dF(e|d,W=1) \cdot dF(d|W=1) \tag{4.12}$$

As for the aggregate case, I can approach this problem by reweighting the observed distribution via an appropriate factor, which in this case is the ratio of two conditional densities:

$$F_{10111} = \int_y \int_l \int_e \int_d \left[\frac{dF(y|l,e,d,W=0)}{dF(y|l,e,d,W=1)}\right] \cdot F(w|y,l,e,d,W=1) \cdot dF(y|l,e,d,W=1) \cdot dF(l|e,d,W=1) \cdot dF(e|d,W=1) \cdot dF(d|W=1) \tag{4.13}$$

When one rewrites this factor via Bayes' law one can see that it is very similar to the aggregate reweighting factor in Equation 4.8. However, instead of a ratio of unconditional probabilites as in the aggregate case, the second part of the factor is a ratio of conditional probabilities here. In the first ratio one conditions on all covariates, including permanent income. For the second ratio one does not condition on income. In this manner one isolates the relationship between income and region of residence. I denote this reweighting factors as $\Psi(y|l,e,d)$ to write the counterfactual distribution more compactly.

$$\Psi(y|l,e,d) = \frac{dF(y|l,e,d,W=0)}{dF(y|l,e,d,W=1)} = \frac{P(W=1|y,l,e,d)}{P(W=0|y,l,e,d)} \cdot \frac{P(W=0|l,e,d)}{P(W=1|l,e,d)} \tag{4.14}$$

$$F_{10111} = \int_y \int_l \int_e \int_d \Psi(y|l,e,d) \cdot F(w|y,l,e,d,W=1) \cdot dF(y|l,e,d,W=1)$$
$$\cdot\, dF(l|e,d,W=1) \cdot dF(e|d,W=1) \cdot dF(d|W=1) \tag{4.15}$$

Similarly, one obtains the counterfactual distribution F_{11011}, which results from setting the western labor market history to its eastern equivalent, by employing the reweighting factor $\Psi(l|e,d) = \frac{P(W=1|l,e,d)}{P(W=0|l,e,d)} \cdot \frac{P(W=0|e,d)}{P(W=1|e,d)}$. On the other hand, if one only changes the vector of socio-demographic background to obtain the counterfactual distribution F_{11110}, the reweighting factor $\Psi(d) = \frac{P(W=1|d)}{P(W=0|d)} \cdot \frac{P(W=0)}{P(W=1)}$ is needed. In such a way, any kind of counterfactual distribution can be calculated. For instance, in order to assess the impact of simultaneously changing y, e and d while holding l fixed, i.e. F_{10100}, one needs to compute $\Psi(y|l,e,d) \cdot \Psi(e|d) \cdot \Psi(d)$. To obtain the decomposition effects, one has to subtract distributional statistics from each other which stem from distributions that differ in one factor only. There are several possibilities how this can be accomplished. One possible decomposition sequence is the following:

$$\underbrace{\nu(F_{11111}) - \nu(F_{00000})}_{\Delta_o^\nu} = \underbrace{\nu(F_{11111}) - \nu(F_{10111})}_{\Delta_y^\nu} + \underbrace{\nu(F_{10111}) - \nu(F_{10011})}_{\Delta_l^\nu} + \underbrace{\nu(F_{10011}) - \nu(F_{10001})}_{\Delta_e^\nu}$$
$$+ \underbrace{\nu(F_{10001}) - \nu(F_{10000})}_{\Delta_d^\nu} + \underbrace{\nu(F_{10000}) - \nu(F_{00000})}_{\Delta_w^\nu} \tag{4.16}$$

For example, $\nu(F_{11111}) - \nu(F_{10111})$ represents the effect of changing permanent income because the two distributions involved differ only in this respect. Actually, there are $n!$ different sequences, in which such a decomposition can be computed, where n is the number of subsets in x plus one. In this case there are $5! = 120$ different sequences to decompose the wealth gap. However, only 24 of these possible sequences depend entirely on the conditional wealth function in West Germany (i.e. involve counterfactual distributions of the form F_{1xxxx}). One can show that Δ_o^ν, Δ_x^ν and Δ_w^ν are the same for each of these 24 decompositions, in the spirit of the aggregate decomposition. However, the subset effects (Δ_y^ν, Δ_l^ν, Δ_e^ν and Δ_d^ν) can be different for each decomposition. The reason for this is that the counterfactual distributions depend on which factor is changed first. For this reason this approach is called sequential decomposition as it depends on the sequence of changes. Thus, one reports the effect for each subset averaged over all possible decomposition sequences. This is known as Shapley decomposition as introduced by Shorrocks (2013).

4.5 Empirical Results

In the following I present the results of the decomposition analyses. I focus again on the latest year available, 2012, since the results for the other two sample years are qualitatively similar. The corresponding tables can be found in the appendix. When reporting the decomposition results, I will omit the ν subscript at this point for the sake of clarity.

4.5.1 Results of Aggregate Decomposition

I start by looking at the aggregate decomposition and concentrate on the overall explanatory power of my wealth determinants, i.e. how much of the observed gap can be attributed to observable differences in the characteristics of eastern and western citizens. Firstly, Table 3.12 presents summary statistics for per capita net wealth in East and West Germany as already reported in Table 3.2 of Section 4.3.1. Secondly, it shows the counterfactual distribution, i.e. the distribution of wealth in West Germany that would arise if western citizens had the same distribution of the observable characteristics that prevails in the East. The CDF's for these three distributions are depicted in Figure 4.3. Moreover, Table 3.12 displays the already familiar observed wealth gap (Δ_o) together with the part of the gap that is associated with the covariates (Δ_x) as well as the part that cannot be explained by these factors (Δ_w). In addition to these figures I report bootstrap standard errors to assess whether the estimates are statistically significant. Table 3.13 displays the overall gap, the composition effect and the wealth structure effect for selected summary measures only along with the relative share of the overall gap corresponding to each component.

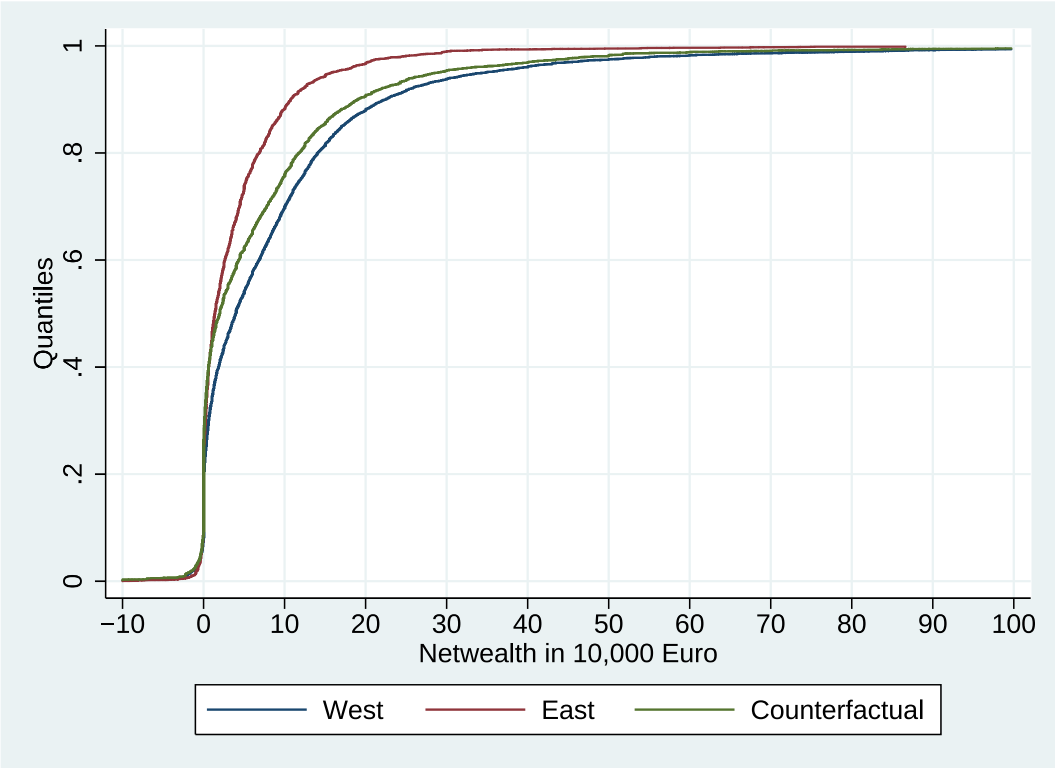


Figure 4.3: CDF of Counterfactual Net Wealth in 2012. Source: Own computation using GSOEP weights.

The first interesting information in these tables is the counterfactual distribution of net wealth. One can see, for instance, that the average West German has total net assets of about 94,000 €. If I reweight the characteristics of the population in the West to match those in the East, I find that an average person would then hold only around 73,500 €. Thus, the more unfavorable composition of the covariates in East Germany is associated with a reduction of the average net wealth by about 20,500 €. Yet, the counterfactual mean wealth is still much closer to the original figure of the West than to that of the East, which is only about 41,000 €. As a rule of thumb, the closer the reweighted distribution for the West is to the distribution in the East, the larger the part of the gap that is associated with the covariates. Looking at Figure 4.3, one can see how the relative position of the counterfactual distribution changes along the levels of net wealth. As one could see in Section 4.3.1, there is little difference between the distributions for East and West Germany up to the lower quartile due to the high numbers of individuals

Stats	West	CF	East	Δ_o	Δ_x	Δ_w
Mean	94,088.36	73,561.62	41,104.91	52,983.45	20,526.74	32,456.70
	[2,770.71]	[3,392.33]	[2,080.86]	[3,640.87]	[3,103.96]	[4,086.77]
p5	-3,230.00	-3,840.00	-3,132.60	-97.40	610.00	-707.40
	[797.82]	[965.40]	[515.74]	[854.52]	[839.84]	[1,046.70]
p10	0.00	0.00	-65.00	65.00	0.00	65.00
	[0.00]	[125.72]	[271.35]	[271.35]	[125.72]	[326.28]
p20	270.00	0.00	0.00	270.00	270.00	0.00
	[228.34]	[0.00]	[0.00]	[228.34]	[228.34]	[0.00]
p25	2,695.00	0.00	353.00	2,342.00	2,695.00	-353.00
	[532.68]	[78.71]	[430.84]	[705.55]	[524.60]	[440.31]
p30	6,075.00	1,008.00	2,110.00	3,965.00	5,067.00	-1,102.00
	[558.39]	[402.17]	[460.36]	[670.86]	[595.63]	[599.33]
p40	18,121.00	5,978.30	6,020.00	12,101.00	12,142.70	-41.70
	[1,825.35]	[1,482.44]	[690.10]	[2,011.26]	[1,606.34]	[1,659.84]
p50	39,390.00	19,235.00	13,644.67	25,745.33	20,155.00	5,590.33
	[1,799.58]	[2,722.73]	[1,389.83]	[2,211.99]	[2,611.46]	[2,894.33]
p60	68,284.15	43,209.46	26,059.90	42,224.25	25,074.68	17,149.57
	[2,982.14]	[4,074.19]	[2,556.01]	[3,551.62]	[3,829.47]	[4,481.56]
p70	100,475.00	77,750.00	43,914.67	56,560.33	22,725.00	33,835.33
	[2,700.19]	[5,017.91]	[2,431.78]	[3,192.46]	[4,989.33]	[5,084.05]
p75	119,260.00	96,583.34	52,650.80	66,609.20	22,676.67	43,932.54
	[3,448.97]	[4,072.60]	[2,625.04]	[4,102.44]	[4,474.17]	[4,416.52]
p80	140,555.00	117,260.70	68,213.34	72,341.66	23,294.30	49,047.37
	[3,217.51]	[5,500.54]	[4,231.21]	[5,022.18]	[5,806.80]	[6,500.96]
p90	223,720.59	190,030.00	107,456.66	116,263.93	33,690.60	82,573.34
	[7,127.75]	[8,709.05]	[4,002.73]	[8,268.28]	[10,847.18]	[10,014.29]
p95	345,900.00	288,187.00	158,313.59	187,586.41	57,713.00	129,873.40
	[15,760.63]	[21,033.95]	[11,687.56]	[21,638.88]	[19,943.52]	[24,567.28]
Gini	70.51	75.85	72.99	-2.48	-5.34	2.87
	[0.85]	[1.40]	[1.32]	[1.73]	[1.29]	[2.06]
% Neg.	8.18	9.26	9.92	-1.74	-1.08	-0.66
	[0.52]	[0.98]	[0.92]	[1.17]	[0.82]	[1.45]
% Zero	11.03	17.13	14.07	-3.04	-6.10	3.06
	[0.40]	[1.20]	[0.78]	[0.87]	[1.09]	[1.41]
% Nonpos.	19.20	26.38	23.99	-4.78	-7.18	2.39
	[0.66]	[1.55]	[1.22]	[1.50]	[1.39]	[2.08]
p90p50	5.68	9.88	7.88	-2.20	-4.20	2.00
	[0.26]	[1.46]	[0.83]	[0.91]	[1.37]	[1.58]
p75p50	3.03	5.02	3.86	-0.83	-2.00	1.16
	[0.11]	[0.66]	[0.36]	[0.35]	[0.64]	[0.75]

Table 3.12: Aggregate Decomposition for 2012. Figures in brackets represent contribution of the respective category to the overall wealth gap in percent. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

with no net worth. Above this point one can observe that at first the CDF of the counterfactual distribution is very close to that for the eastern population. As one moves up in the distribution, however, the counterfactual CDF moves closer and closer to the western CDF. This indicates that the potential wealth determinants help to explain a good amount of the observed wealth differences in the lower part of the distribution but have little explanatory power in the upper part. This is quite reasonable as large fortunes can come about in many ways. Thus, they are not easily reduced to a small number of underlying factors.

Stats	Δ_o	Δ_x	Δ_w
Mean	52,983.45	20,526.74	32,456.70
	(100.00)	(38.74)	(61.26)
	[3,640.87]	[3,103.96]	[4,086.77]
p25	2,342.00	2,695.00	-353.00
	(100.00)	(115.07)	(-15.07)
	[705.55]	[524.60]	[440.31]
p40	12,101.00	12,142.70	-41.70
	(100.00)	(100.34)	(-0.34)
	[2,011.26]	[1,606.34]	[1,659.84]
p50	25,745.33	20,155.00	5,590.33
	(100.00)	(78.29)	(21.71)
	[2,211.99]	[2,611.46]	[2,894.33]
p60	42,224.25	25,074.68	17,149.57
	(100.00)	(59.38)	(40.62)
	[3,551.62]	[3,829.47]	[4,481.56]
p75	66,609.20	22,676.67	43,932.54
	(100.00)	(34.04)	(65.96)
	[4,102.44]	[4,474.17]	[4,416.52]
p90	116,263.93	33,690.60	82,573.34
	(100.00)	(28.98)	(71.02)
	[8,268.28]	[10,847.18]	[10,014.29]
Gini	-2.48	-5.34	2.87
	(100.00)	(215.80)	(-115.80)
	[1.73]	[1.29]	[2.06]
% Nonpos.	-4.78	-7.18	2.39
	(100.00)	(150.03)	(-50.03)
	[1.50]	[1.39]	[2.08]
p90p50	-2.20	-4.20	2.00
	(100.00)	(191.04)	(-91.04)
	[0.91]	[1.37]	[1.58]

Table 3.13: Aggregate Decomposition (Shares) for 2012. Figures in brackets represent contribution of the respective category to the overall wealth gap in percent. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

By looking at the decomposition results, one can more easily quantify the contribution of the covariates to the wealth gap along the distribution. For instance, one quickly notes that out of the nearly 53,000 € mean wealth gap, around 20,500 € are associated with the composition effect while about 32,500 € are due to the wealth structure effect. Thus, the wealth determinants are corresponding to only 38.74 % of the mean wealth gap. Due to the lack of substantive wealth differences below the 25 % quantile, only relatively small values exist for Δ_x and Δ_w over this range. In addition, the bootstrap standard errors are quite large in this area so that the effects are not significant at the 5 % level. Starting from the lower quartile I observe a positive wealth gap which is widening substantially with increasing levels of net wealth. Between the lower quartile and the median the composition effect is as large as or even larger than the observed wealth gap itself. The wealth structure effect in this region is therefore negative. However, Δ_w is usually not statistically significant in this vicinity. As a result, one cannot say whether the share accounted for by Δ_x is actually larger than 100 %. As one moves further up the distribution, it becomes evident that the observable factors contribute less and less to the overall gap. While the variables in the x vector are associated with 78.29 % of the median wealth gap (about 20,000 € out of roughly 25,750 €), this proportion declines quickly and constitutes only about 34 % of the gap at the upper quartile (around 22,700 € of 66,600 €). The share associated with the composition effect stabilizes after the 75 % quantile at about one third of the overall gap. For instance, at the 90 % quantile about 33,700 €, or 29 % of an overall gap of 116,300 €, are due to Δ_x. For the distributional measures, such as the Gini coefficient or the ratio of 90 % quantile to median, I find very large relative composition effects. In fact, the contribution of the covariates is much higher than the gap actually observed. For instance, the relative contribution of the wealth determinants to the difference between the Gini coefficient in the two regions is 215.80 %. Thus, the composition effect is more than twice as high as the observed gap. The results also suggest that the share of individuals with no positive net worth would be even higher in the counterfactual distribution than in the East - more than 26 % compared to the actual share of about 19 % in the West and roughly 24 % in the East. Decomposing such small differences is, however, usually quite difficult. This could also mean that the higher observed inequality in the

East could actually be much worse if people had not accommodated themselves to the prevalent conditions.

4.5.2 Results of Detailed Decomposition

To assess in more detail how these figures come about, I examine the results for the detailed decomposition in Table 3.14 and Table 3.15. As before, I concentrate mainly on Table 3.15 where the share of the overall gap is given for selected summary measures. What is striking in this regard is that for all wealth levels the contribution of educational attainment (Δ_e) as well as socio-demographic background (Δ_d) on the observed gap are consistently negative. This means by assigning the prevailing distribution of these variables in the East to the western population, one would actually increase the wealth gap. The share of the overall gap due to differences in socio-demographic background is mostly in the range of minus 5 % and minus 10 %. For the lower half of the distribution, however, this effect is not statistically significant. The absolute magnitude of the effect of education differences is much larger in the lower half of the distribution and it is nearly always statistically significant. Yet, this effect quickly decreases in size so that its relative share in the upper part of the distribution is usually smaller than that of Δ_d. The combined contribution of these two categories varies somewhere between about minus 25 % at the lower quartile and minus 11 % at the upper quartile. Together they are associated with about minus 7,200 € of the mean wealth gap, which is equivalent to a share of approximately minus 13.5 %.

Even though one cannot assess the contribution of individual variables in this framework, some potential reasons for these findings come to mind. First of all, as one has seen in Section 4.3.3, the education level of household members in the East as well as that of their parents are on average higher than for those in the West. Higher educational attainment is usually associated with higher levels of wealth.[15] Therefore, it is reasonable to assume that if one applied the East German distribution of these variables to West Germany, it would result in higher levels of net worth. A similar argument can be brought forth for the share of foreign nationals. Bauer

[15]See Sierminska et al. (2010).

Stats	Δ_o	Δ_x	Δ_y	Δ_l	Δ_e	Δ_d
Mean	52,983.45	20,526.74	16,737.37	10,969.15	-3,193.39	-3,986.39
	[3,641.46]	[3,124.01]	[1,682.63]	[2,310.12]	[777.17]	[1,605.80]
p5	-97.4	610	230.52	472.74	-213.86	120.60
	[856.98]	[835.00]	[309.49]	[499.80]	[200.08]	[346.60]
p10	65	0	7.98	7.98	-3.02	-12.95
	[277.04]	[109.95]	[62.79]	[79.49]	[22.20]	[56.65]
p20	270	270	243.75	243.75	-162.64	-54.86
	[228.34]	[228.34]	[125.55]	[121.65]	[55.98]	[76.67]
p25	2,342.00	2,695.00	1,552.54	1,700.04	-449.1	-108.47
	[699.58]	[521.57]	[294.20]	[326.42]	[131.59]	[194.85]
p30	3,965.00	5,067.00	2,886.19	3,027.65	-786	-60.84
	[671.87]	[590.34]	[383.47]	[423.85]	[235.05]	[339.05]
p40	12,101.00	12,142.70	7,084.70	7,155.04	-1,479.41	-617.63
	[2,009.89]	[1,625.45]	[792.03]	[895.82]	[398.73]	[877.02]
p50	25,745.33	20,155.00	11,714.44	11,918.37	-2,123.47	-1,354.35
	[2,226.39]	[2,656.81]	[1,313.02]	[1,713.04]	[782.14]	[1,504.92]
p60	42,224.25	25,074.68	15,769.21	14,804.70	-2,674.04	-2,825.19
	[3,542.90]	[3,817.28]	[1,933.82]	[2,430.84]	[974.46]	[1,833.45]
p70	56,560.33	22,725.00	15,970.42	13,394.18	-2,572.16	-4,067.44
	[3,185.80]	[5,026.20]	[2,073.12]	[2,965.19]	[860.20]	[1,962.08]
p75	66,609.20	22,676.67	16,800.19	13,378.59	-2,909.23	-4,592.88
	[4,118.25]	[4,559.43]	[2,188.92]	[2,640.03]	[814.59]	[1,999.61]
p80	72,341.66	23,294.30	19,252.04	13,689.38	-3,771.07	-5,876.05
	[5,053.53]	[5,801.74]	[2,688.43]	[2,973.63]	[1,038.03]	[2,386.63]
p90	116,263.93	33,690.60	32,138.25	19,611.41	-6,704.35	-11,354.71
	[8,233.45]	[10,685.18]	[6,513.20]	[5,790.39]	[2,145.84]	[3,598.56]
p95	187,586.41	57,713.00	66,193.47	21,947.80	-13,683.63	-16,744.65
	[21,720.59]	[19,782.96]	[12,824.81]	[14,539.63]	[4,417.31]	[6,990.42]
Gini	-2.48	-5.34	-2.13	-3.29	0.41	-0.33
	[1.73]	[1.27]	[0.46]	[0.79]	[0.18]	[0.49]
% Nonpos.	-4.78	-7.18	-4.10	-4.22	1.30	-0.16
	[1.49]	[1.39]	[0.49]	[0.71]	[0.26]	[0.54]
% Neg.	-1.74	-1.08	-0.65	-1.00	0.20	0.37
	[1.19]	[0.82]	[0.29]	[0.39]	[0.15]	[0.35]
% Zero	-3.04	-6.10	-3.45	-3.22	1.11	-0.53
	[0.86]	[1.08]	[0.38]	[0.60]	[0.20]	[0.42]
p75p50	-0.83	-2.00	-1.05	-1.19	0.29	-0.06
	[0.35]	[0.64]	[0.31]	[0.35]	[0.21]	[0.20]
p90p50	-2.20	-4.20	-2.04	-2.50	0.54	-0.20
	[0.90]	[1.37]	[0.70]	[0.73]	[0.42]	[0.43]

Table 3.14: Detailed Decomposition for 2012. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

et al. (2011) note that native Germans exhibit much higher levels of wealth compared to those inhabitants with a foreign nationality. West Germany has an 11 times higher share of non-

German nationals than the eastern federal states. Attributing this lower eastern figure to West Germany should therefore lead to higher wealth levels in the counterfactual distribution. Such relationships might explain these negative decomposition results.

Stats	Δ_o	Δ_x	Δ_y	Δ_l	Δ_e	Δ_d
Mean	52,983.45	20,526.74	16,737.37	10,969.15	-3,193.39	-3,986.39
	(100.00)	(38.74)	(31.59)	(20.70)	(-6.03)	(-7.52)
	[3,641.46]	[3,124.01]	[1,682.63]	[2,310.12]	[777.17]	[1,605.80]
p25	2,342.00	2,695.00	1,552.54	1,700.04	-449.10	-108.47
	(100.00)	(115.07)	(66.29)	(72.59)	(-19.18)	(-4.63)
	[699.58]	[521.57]	[294.20]	[326.42]	[131.59]	[194.85]
p40	12,101.00	12,142.70	7,084.70	7,155.04	-1,479.41	-617.63
	(100.00)	(100.34)	(58.55)	(59.13)	(-12.23)	(-5.10)
	[2,009.89]	[1,625.45]	[792.03]	[895.82]	[398.73]	[877.02]
p50	25,745.33	20,155.00	11,714.44	11,918.37	-2,123.47	-1,354.35
	(100.00)	(78.29)	(45.50)	(46.29)	(-8.25)	(-5.26)
	[2,226.39]	[2,656.81]	[1,313.02]	[1,713.04]	[782.14]	[1,504.92]
p60	42,224.25	25,074.68	15,769.21	14,804.70	-2,674.04	-2,825.19
	(100.00)	(59.38)	(37.35)	(35.06)	(-6.33)	(-6.69)
	[3,542.90]	[3,817.28]	[1,933.82]	[2,430.84]	[974.46]	[1,833.45]
p75	66,609.20	22,676.67	16,800.19	13,378.59	-2,909.23	-4,592.88
	(100.00)	(34.04)	(25.22)	(20.09)	(-4.37)	(-6.90)
	[4,118.25]	[4,559.43]	[2,188.92]	[2,640.03]	[814.59]	[1,999.61]
p90	116,263.93	33,690.60	32,138.25	19,611.41	-6,704.35	-11,354.71
	(100.00)	(28.98)	(27.64)	(16.87)	(-5.77)	(-9.77)
	[8,233.45]	[10,685.18]	[6,513.20]	[5,790.39]	[2,145.84]	[3,598.56]
Gini	-2.48	-5.34	-2.13	-3.29	0.41	-0.33
	(100.00)	(215.80)	(85.95)	(132.93)	(-16.60)	(13.51)
	[1.73]	[1.27]	[0.46]	[0.79]	[0.18]	[0.49]
% Nonpos.	-4.78	-7.18	-4.10	-4.22	1.30	-0.16
	(100.00)	(150.03)	(85.80)	(88.24)	(-27.28)	(3.26)
	[1.49]	[1.39]	[0.49]	[0.71]	[0.26]	[0.54]
p90p50	-2.20	-4.20	-2.04	-2.50	0.54	-0.20
	(100.00)	(191.04)	(92.68)	(113.55)	(-24.47)	(9.29)
	[0.90]	[1.37]	[0.70]	[0.73]	[0.42]	[0.43]

Table 3.15: Detailed Decomposition (Shares) for 2012. Figures in brackets represent contribution of the respective category to the overall wealth gap in percent. Figures in square brackets present the standard error from 500 bootstrap samples.

The effects of income differentials (Δ_y) and different labor market outcomes (Δ_l), in contrast, are positive, usually quite substantial in magnitude and highly statistical significant. Unsurprisingly, given the negative effects of Δ_e and Δ_d, their combined contribution is even higher than the total composition effect. Looking at the mean wealth gap, I find that in fact more than half of the observed gap is due to the combined effect of income differentials and labor market differences. Roughly 16,700 € of the difference can be traced back to higher incomes in the West. This is equivalent to 31.59 % of the overall gap or rather 81.54 % of the composition effect. With about 11,000 € the contribution of the differences in labor market histories is somewhat smaller but still presents 20.70 % of Δ_o and 53.44 % of Δ_x. The two effects are usually of the same size although the income effect is larger for high levels of wealth. The combined share of the two categories declines along the distribution and ranges from 138 % at the lower quartile to about 45 % at the 90 % quantile. It is harder to make definitive statements for the distributional measures due to the extreme nature of the effects involved. As for the quantiles, Δ_y and Δ_l are by far the most important contributing factors in this regard. These two effects are roughly comparable in size and each as large as the overall wealth gap itself. The magnitudes of Δ_e and Δ_d, on the other hand, are again rather small. All this indicates that the explanatory power of certain parts of the wealth determinants is actually higher than the aggregated figures suggest. The fact that the direction of the individual contributions go into different directions is lost when one looks at the aggregated effects only.

4.5.3 The Role of Home-Equity

As one could see in Table 3.7, owner-occupied real estate constitutes the main asset in the portfolio of the average German. Having said that, a study by the ECB's Eurosystem Household Finance and Consumption Network (HFCS, 2013) finds that Germany and Austria are the countries in the euro zone with the lowest share of home-owners. The study further suggests that this low propensity to hold self-utilized real estate is related to the relative low levels of median net wealth in these two countries as well as their relative high levels of wealth inequality. This holds true in particular in comparison to countries like Spain. From Section 4.3.2 I know that the chief difference in the investment behavior of East and West Germans lies in their propensity to own a home, which is 12.50 percentage points lower in the East. Thus, it is reasonable to assess whether such a relationship between differences in home-ownership rates and wealth levels can also be found within Germany.

To evaluate the potential effect of differentials in the propensity for home-ownership, I include a housing dummy in the decomposition analysis in addition to the usual wealth determinants. More specifically, I model the decision to own a home as being conditional on ones permanent income, labor market history, education and socio-demographic background. Although this approach makes intuitive sense, it is problematic to use home-ownership in such a way as it is itself a part of net wealth. Nevertheless, I still proceed in this fashion as I primarily seek to explore this relationship in a descriptive manner.

Stats	Δ_o	Δ_x	Δ_h	Δ_y	Δ_l	Δ_{de}
Mean	52,983.45	26,145.41	5,459.04	16,514.34	11,031.07	-6,859.05
	(100.00)	(49.35)	(10.30)	(31.17)	(20.82)	(-12.95)
	[3,624.02]	[3,023.39]	[1,053.50]	[1,647.20]	[2,213.28]	[1,835.55]
p25	2,342.00	2,695.00	464.80	1,322.03	1,434.98	-526.81
	(100.00)	(115.07)	(19.85)	(56.45)	(61.27)	(-22.49)
	[707.73]	[529.53]	[139.88]	[276.48]	[285.34]	[215.93]
p40	12,101.00	13,531.60	2,811.16	6,251.38	6,397.57	-1,928.52
	(100.00)	(111.82)	(23.23)	(51.66)	(52.87)	(-15.94)
	[2,014.20]	[1,565.04]	[752.16]	[748.59]	[850.08]	[861.57]
p50	25,745.33	25,135.00	5,970.82	10,996.67	11,298.91	-3,131.40
	(100.00)	(97.63)	(23.19)	(42.71)	(43.89)	(-12.16)
	[2,205.81]	[2,699.34]	[1,369.63]	[1,228.99]	[1,582.40]	[1,432.06]
p60	42,224.25	33,151.02	8,381.66	15,130.73	14,767.71	-5,129.09
	(100.00)	(78.51)	(19.85)	(35.83)	(34.97)	(-12.15)
	[3,443.56]	[4,407.13]	[2,209.51]	[1,830.25]	[2,174.43]	[1,951.61]
p75	66,609.20	30,096.47	7,358.92	16,680.19	13,550.28	-7,492.92
	(100.00)	(45.18)	(11.05)	(25.04)	(20.34)	(-11.25)
	[4,101.12]	[4,975.14]	[1,752.72]	[2,161.30]	[2,667.79]	[2,371.08]
p90	116,263.93	43,240.60	10,175.33	31,043.90	19,640.82	-17,619.45
	(100.00)	(37.19)	(8.75)	(26.70)	(16.89)	(-15.15)
	[8,381.46]	[10,732.51]	[2,689.14]	[6,018.20]	[5,876.17]	[4,209.53]
Gini	-2.48	-7.25	-1.74	-2.19	-3.36	0.04
	(100.00)	(292.90)	(70.25)	(88.53)	(135.86)	(-1.73)
	[1.73]	[1.35]	[0.36]	[0.48]	[0.80]	[0.55]
% Nonpos.	-4.78	-8.56	-1.27	-4.18	-4.29	1.18
	(100.00)	(178.92)	(26.56)	(87.43)	(89.59)	(-24.66)
	[1.51]	[1.45]	[0.26]	[0.50]	[0.70]	[0.64]
p90p50	-2.20	-7.08	-1.67	-2.57	-3.06	0.22
	(100.00)	(321.63)	(75.72)	(116.63)	(139.08)	(-9.80)
	[0.91]	[2.74]	[0.68]	[0.97]	[1.07]	[0.52]

Table 3.16: Detailed Decomposition including Home-Equity. Figures in brackets represent contribution of the respective category to the overall wealth gap in percent. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

Table 3.16 reports the results for this extension of the original analysis.[16] I find that the housing effect (Δ_h) is positive but its magnitude is usually much lower than that of permanent income (Δ_y) and labor market situation (Δ_l). Including home-ownership increases the composition effect at the mean by about 10 percentage points. This is equivalent to a third (half) of the size of the effect of permanent income (labor market status). At the median the housing dummy increases the explanatory power of Δ_x by about 23 percentage points. This is, however, still only slightly more than half the magnitude of the other two effects. For the measures of inequality and the share of individuals with no positive net worth the pattern is similar: the direction of the effect is the same as for Δ_y and Δ_l but considerably smaller in size.

Wealth Component	Δ_o	Δ_x	Δ_y	Δ_l	Δ_e	Δ_d
House Own Yes	12.48	10.91	4.09	5.50	-0.22	1.55
	(100.00)	(87.42)	(32.75)	(44.09)	(-1.79)	(12.38)
	[1.15]	[0.98]	[0.43]	[0.58]	[0.18]	[0.53]
Mortgage Own Yes	7.54	8.08	2.64	3.66	-0.22	1.99
	(100.00)	(107.15)	(35.02)	(48.58)	(-2.91)	(26.46)
	[0.92]	[0.66]	[0.33]	[0.40]	[0.13]	[0.37]
House Own	33,492.16	13,397.88	8,932.16	7,307.36	-1,101.31	-1,740.33
	(100.00)	(40.00)	(26.67)	(21.82)	(-3.29)	(-5.20)
	[1,398.92]	[1,414.67]	[886.10]	[1,086.64]	[398.54]	[779.90]
Mortgage Own	5,526.66	4,132.29	1,872.13	2,065.16	-188.04	383.03
	(100.00)	(74.77)	(33.87)	(37.37)	(-3.40)	(6.93)
	[527.88]	[376.34]	[171.42]	[261.82]	[79.28]	[185.68]
Net House Own	27,965.51	9,265.59	7,060.03	5,242.20	-913.27	-2,123.36
	(100.00)	(33.13)	(25.25)	(18.75)	(-3.27)	(-7.59)
	[1,273.89]	[1,252.29]	[817.84]	[1,019.74]	[341.61]	[758.05]

Table 3.17: Decomposition of Home-Equity. Figures in brackets represent contribution of the respective category to the overall wealth gap in percent. Figures in square brackets present the standard error from 500 bootstrap samples.

Another way to look at this issue is to decompose the participation rates and wealth levels for home equity directly. In this fashion one can estimate how much of the mean difference in owner-occupied housing is corresponding to differences in the distribution of covariates. The resulting effects are given in Table 3.17. I consider the participation rates for home equity as well as mortgages associated with such real estate. In addition, I look into the mean gross and net wealth

[16] For reasons of clarity I subsume the effects for educational attainment and socio-demographic background under the category Δ_{de} at this point since they have the same direction and a similar magnitude.

levels invested in owner-occupied housing along with the gross value of home loans. Finally, net housing wealth for self-utilized property is scrutinized. What is most notable in Table 3.17, is the discrepancy between the results for the mean wealth levels and the participation rates. The decomposition effects for both gross and net home equity are quite similar to those for net wealth as a whole. In either case the relative share of the composition effect is of the same order of magnitude as for total net wealth (40.00 % and 33.13 %, respectively). Furthermore, one also sees relative large positive effects for income and labor market situation while the contributions of education level and social background are negative and much smaller in magnitude. The gap in participation rates for home equity and corresponding home loans, on the other hand, are much better explained by the differentials in observable characteristics. For instance, nearly 11 percentage points, out of the 12.48 percentage point gap in the participation rates for home ownership, can be attributed to differences in the covariates. This disparity between the relative magnitudes of the composition effect for the participation rates and the actual mean property values is quite striking. It might suggest that the different propensities for owning real estate in the two parts of Germany are only part of the story. Evidently, other factors such as the different price levels for housing property do play a role as well. From all this I cautiously conclude that the different home ownership rates seem related to the wealth gap between East and West Germany. However, the magnitude of the housing effect Δ_h is much smaller than the unconditional summary statistics suggest.

4.5.4 Results by Cohort

The results presented so far apply to the overall population in both regions. Yet, it is likely that various groups experienced the German reunification in different ways. Individuals who were younger or not even born at that time might have been affected differently by the consequences of this event than older Germans who experienced the disparate cultural and political conditions in the opposing systems for most of their life.

To examine whether younger Germans differ from their elder countrymen in regard to the composition of the wealth gap, I conduct a separate analysis with individuals belonging to households

Stats	West	CF	East	Δ_o	Δ_w	Δ_x
Mean	45,396.63	33,461.42	27,374.74	18,021.89	6,086.68	11,935.21
	[3,521.36]	[3,297.91]	[3,882.27]	[5,191.99]	[5,220.59]	[3,701.23]
p5	-8,880.00	-7,950.00	-4,980.50	-3,899.50	-2,969.50	-930
	[1,084.52]	[1,578.14]	[1,464.89]	[1,835.25]	[2,015.07]	[1,361.13]
p10	-1,890.60	-2,136.67	-2,094.40	203.8	-42.27	246.07
	[637.26]	[651.93]	[847.03]	[1,094.16]	[1,035.03]	[679.87]
p20	0.00	0.00	0.00	0.00	0.00	0.00
	[0.00]	[15.03]	[29.85]	[29.85]	[36.17]	[15.03]
p25	0.00	0.00	0.00	0.00	0.00	00.00
	[65.71]	[0.00]	[0.00]	[65.71]	[0.00]	[65.71]
p30	990.00	0.00	0.00	990.00	0.00	990.00
	[385.25]	[101.84]	[116.55]	[409.91]	[154.82]	[367.64]
p40	4,770.00	1,600.00	1,325.67	3,444.33	274.33	3,170.00
	[729.82]	[465.39]	[591.98]	[974.00]	[797.82]	[739.06]
p50	10,303.10	4,790.00	3,538.80	6,764.30	1,251.20	5,513.10
	[1,430.59]	[1,016.68]	[979.92]	[1,595.25]	[1,366.49]	[1,530.26]
p60	19,610.00	9,960.00	8,526.40	11,083.60	1,433.60	9,650.00
	[2,600.27]	[1,546.26]	[1,685.86]	[3,279.15]	[1,982.55]	[2,491.34]
p70	36,262.43	20,752.07	19,520.00	16,742.43	1,232.07	15,510.37
	[3,668.72]	[3,221.71]	[4,470.41]	[6,117.30]	[5,181.28]	[3,443.95]
p75	49,926.67	28,773.33	27,570.70	22,355.97	1,202.63	21,153.33
	[3,646.96]	[4,480.42]	[5,922.55]	[6,811.98]	[6,860.93]	[4,401.91]
p80	66,180.30	42,730.00	39,016.67	27,163.63	3,713.33	23,450.30
	[4,920.66]	[7,085.31]	[5,502.07]	[6,768.61]	[7,858.95]	[6,268.20]
p90	121,250.00	92,894.70	79,230.00	42,020.00	13,664.70	28,355.30
	[8,021.98]	[10,473.16]	[12,235.85]	[14,686.26]	[15,219.66]	[8,389.83]
p95	182,200.00	142,570.16	127,550.00	54,650.00	15,020.15	39,629.85
	[14,248.95]	[15,366.18]	[12,803.88]	[20,237.74]	[20,827.63]	[17,037.50]
Gini	82.03	87.92	85.47	-3.44	2.45	-5.88
	[2.23]	[2.17]	[3.70]	[4.82]	[4.32]	[2.14]
% Nonpos.	26.09	31.14	32.31	-6.21	-1.17	-5.05
	[1.48]	[2.00]	[2.73]	[3.15]	[3.33]	[1.86]
% Neg.	14.27	16.03	17.05	-2.78	-1.03	-1.76
	[1.28]	[1.58]	[2.27]	[2.65]	[2.87]	[1.40]
% Zero	11.82	15.11	15.25	-3.43	-0.14	-3.29
	[0.82]	[1.38]	[1.77]	[1.95]	[2.11]	[1.14]
p75p50	4.86	6.03	7.82	-2.97	-1.79	-1.17
	[0.63]	[1.14]	[2.24]	[2.17]	[2.11]	[1.29]
p90p50	11.80	19.53	22.52	-10.72	-2.99	-7.73
	[1.58]	[4.72]	[6.28]	[6.09]	[8.17]	[4.65]

Table 3.18: Aggregate Decomposition for Young Cohort. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

for which no member is older than 45 years, only.[17] Of course this cut-off is somewhat arbitrary

[17] About 26 % of the sample population fall into this category.

but I examined different cut-off criteria and find that the results are qualitatively unchanged. I conduct this analysis only for the year 2012 as by that time enough people in this cohort had reasonable time to build up meaningful net asset positions. Table 3.18 gives a detailed account of the aggregate decomposition effects across the distribution while Table 3.19 presents results for the detailed decomposition for selected summary statistics along with the share of the overall wealth gap.

The first thing to notice in these tables are the much lower wealth levels found for younger Germans compared to the overall population. Their mean wealth level in the West, for example, is less than half the level of the western population as a whole - less than 45,400 € compared to more than 94,000 € overall. For the East the situation is very similar even if the difference is not as pronounced - the average East German in the younger cohort owns around 27,400 € in net assets while this figure is about 41,100 € for all East Germans. These lower levels of wealth can be found at any point in the distribution. Consequently, the wealth gap for the younger cohort is generally also much lower than for the overall population. For example, the mean wealth gap for this group is only about 18,000 € - compared to an overall mean wealth gap of around 53,000 €. These findings are not unexpected as people in the younger cohort simply had less time to build up assets. For the same reason a much higher share of individuals with non-positive net worth is found in both regions: more than 26 % of young westerners have no or negative net assets compared to only about 19 % for all West Germans. Furthermore, the wealth distribution for young Germans is also very unequally distributed. For example, the Gini coefficient for young East Germans is 85.47 % while it is only 72.99 % for all easterners.

Another striking finding is that a very high share of the observed wealth gap among young Germans can be attributed to the composition effect. The mean composition effect, for instance, is about 12,000 € or 66.23 % of the mean wealth gap for the younger cohort. This stands in contrast to the corresponding share in the general population which is only 38.74 %. Interestingly, this high explanatory power of the wealth determinants is rather stable along the entire wealth distribution and does not decline as strong as for the entire population. The covariates are associated with 81.50 % of the median wealth gap of about 6,700 € and still 67.48 % of the 42,000 € gap at the 90 % quantile.

Stats	Δ_o	Δ_x	Δ_y	Δ_l	Δ_e	Δ_d
Mean	18,021.89	11,935.21	2,413.52	5,668.02	2,190.02	1,663.65
	(100.00)	(66.23)	(13.39)	(31.45)	(12.15)	(9.23)
	[5,191.99]	[3,701.23]	[1,551.84]	[2,793.17]	[1,091.82]	[1,755.42]
p30	990.00	990.00	401.76	527.29	42.16	18.78
	(100.00)	(100.00)	(40.58)	(53.26)	(4.26)	(1.90)
	[409.91]	[367.64]	[149.94]	[234.86]	[93.77]	[180.47]
p40	3,444.33	3,170.00	835.42	1,557.02	261.57	515.99
	(100.00)	(92.04)	(24.25)	(45.21)	(7.59)	(14.98)
	[974.00]	[739.06]	[338.25]	[443.80]	[247.43]	[412.59]
p50	6,764.30	5,513.10	1,427.49	2,473.78	525.09	1,086.74
	(100.00)	(81.50)	(21.10)	(36.57)	(7.76)	(16.07)
	[1,595.25]	[1,530.26]	[505.07]	[791.51]	[427.85]	[694.39]
p60	11,083.60	9,650.00	1,828.27	4,220.20	1,080.75	2,520.78
	(100.00)	(87.07)	(16.50)	(38.08)	(9.75)	(22.74)
	[3,279.15]	[2,491.34]	[701.08]	[1,526.97]	[691.90]	[1,358.05]
p75	22,355.97	21,153.33	3,767.93	8,591.27	3,042.95	5,751.19
	(100.00)	(94.62)	(16.85)	(38.43)	(13.61)	(25.73)
	[6,811.98]	[4,401.91]	[1,361.05]	[2,361.74]	[1,653.88]	[2,934.08]
p90	42,020.00	28,355.30	2,699.77	15,056.86	4,102.68	6,495.99
	(100.00)	(67.48)	(6.42)	(35.83)	(9.76)	(15.46)
	[14,686.26]	[8,389.83]	[2,916.33]	[5,727.13]	[2,680.92]	[4,793.42]
Gini	-3.44	-5.88	-1.18	-2.24	-0.63	-1.84
	(100.00)	(171.17)	(34.20)	(65.15)	(18.19)	(53.64)
	[4.82]	[2.14]	[0.68]	[1.49]	[0.47]	[1.20]
% Nonpos.	-6.21	-5.05	-2.33	-3.04	-0.09	0.41
	(100.00)	(81.24)	(37.44)	(48.93)	(1.44)	(-6.57)
	[3.15]	[1.86]	[0.63]	[1.04]	[0.50]	[1.13]
p90p50	-10.72	-7.73	-2.69	-3.12	-0.56	-1.36
	(100.00)	(72.09)	(25.14)	(29.06)	(5.19)	(12.70)
	[6.09]	[4.65]	[1.69]	[2.13]	[0.85]	[1.45]

Table 3.19: Detailed Decomposition for Young Cohort. Figures in brackets represent contribution of the respective category to the overall wealth gap in percent. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

Two potential reasons for these results come to mind: younger Germans might be better comparable as they have more similar attitudes and face an identical institutional framework. Thus, it could be easier to attribute parts of the observed wealth gap to differences in wealth determinants. On the other hand, the higher share of Δ_x might be due to the fact that at lower levels of wealth the covariates are generally associated with a larger part of the wealth gap - as one has seen in Section 4.5.1. Therefore, one might very well conclude that the large relative composition effects for the young can simply be ascribed to the smaller size of the wealth gap at any point in the distribution.

This argument does not really stand up to scrutiny, however. If I compare wealth gaps with similar magnitudes, I find that the relative share of Δ_x is usually higher for the younger cohort even at relatively high levels of wealth. Taking the 95 % quantile of the young cohort as an example, I find that 72.52 % of the 54,650 € gap is attributable to differences in the covariates. For the quantitatively similar wealth gap (56,600 €) at the 70 % quantile of the overall distribution the corresponding share is less than 60 %. As one could see before, the share of the mean wealth gap for the overall distribution (53,000 €) was even as low as 38.74 %.

To assess how these large composition effects come about, I examine the effects of each category as given in Table 3.19. I find that the effects for the socio-demographic background (Δ_d) and the educational attainment (Δ_e) are both positive for individuals in young households in contrast to the effects for the overall population. This might reflect, among others, that the educational qualifications of younger Germans are usually directly comparable - something that is not necessarily the case for older Germans. The contribution of the socio-demographic background to the wealth gap is usually larger than that of the education level. For the upper half of the distribution it is even larger than the income effect. However, neither Δ_e nor Δ_d are usually significant. The largest effect by far is the contribution of the labor market situation of an individual. This effect is usually much larger than the effect of permanent income and lies mostly between 30 % and 40 % of the overall composition effect. This could be due to the fact that the income distribution is more compressed for younger individuals. Thus, income differentials in this group are usually less pronounced than for the population as a whole. Success at the labor market, on the other hand, is much more important during the early stages of one's career. Therefore, times of unemployment, which are much more common among young East Germans, are likely to have a much more detrimental effect on the financial situation of young adults. The income effect, while smaller than for the general population, still contributes a substantive share to the wealth gap and is statistically significant for most parts of the distribution.

All in all, I find that the wealth gap between younger Germans is smaller and more closely associated with differentials in the observable characteristics. The labor market situation seems to play the largest role in this respect.

4.6 Conclusion

In this chapter, I have investigated the gap in net wealth that still exists between East and West Germany today. Using data from the German Socio-Economic Panel, one can see that even in 2012, the latest year available, this gap is quite substantial and proportional to the level of wealth at different points of the wealth distribution. By employing decomposition estimation via counterfactual reweighting procedures, I find that observable differences in potential wealth determinants such as income or labor market situation are associated with varying levels of the observed wealth gap. I observe that for the lower part of the distribution, most of the gap can be attributed to the wealth determinants. However, that share declines quickly for higher wealth levels and accounts for only about a third. Moreover, I find that of the four categories of wealth determinants considered (permanent income, labor market outcomes, education attainment and socio-demographic background), income differentials and differences in labor market situation are associated with the largest part of the wealth gap. Educational attainment and social background, on the other hand, have small negative effects - meaning aligning them between the two parts of the country would even widen the observed gap. This is likely due to factors which are actually more favorable for East Germany such as its low share of foreign nationals and its generally higher level of education.

I also scrutinize the role of owner occupied housing wealth in this context. West Germans have a much higher propensity to own a home compared to their eastern compatriots. As this type of asset constitutes the largest part of wealth of the average investor, it seems natural to assume that this difference in home-ownership rates is partly responsible for the higher level of wealth in West Germany. I indeed find a sizable effect of home-ownership on the wealth gap. However, the effect of home-ownership is quantitatively much smaller than those of income or labor market differentials. This suggests that other factors, such as the generally lower level of housing prices in East Germany, play an important role in this regard.

Finally, I ascertain whether younger Germans are affected differently by the German reunification. I observe much lower wealth differences for the younger cohort compared to the overall population. I also find that the relative composition effect is usually much larger than the one I found for the entire population. This could simply be due to the smaller magnitude of the wealth gap in this group as it is generally easier to explain lower levels of debt. However, I observe that similar magnitudes of wealth differentials are associated with higher shares of the composition effect compared to the general population. Especially labor market outcomes play a most important role here, probably because success at the labor market is much more important at early stages of one's career.

Appendix B: Additional Tables

Variable	Mean	Std. Dev.	Min.	Max.
Permanent Income	1,306.66	759.6	98	73,363
Exp. Ft	17.78	11.63	0	60
Exp. Pt	2.83	4.54	0	46
Exp. Ue	0.88	1.92	0	37
High Job	0.02	0.11	0	1
Selfemp	0.05	0.18	0	1
Retired	0.27	0.42	0	1
Middle Vocation	0.48	0.41	0	1
High Vocation	0.12	0.27	0	1
College	0.17	0.33	0	1
Years Schooling	11.89	2.34	7	18
Age	49.54	16.78	17	102
Male	0.47	0.29	0	1
Married	0.55	0.45	0	1
Foreign	0.08	0.25	0	1
Number of Kids	0.41	0.79	0	8
HH Size	2.48	1.25	1	13
Health Problems	0.19	0.33	0	1
Father College	0.13	0.34	0	1
Mother College	0.07	0.26	0	1
Ever Inheritance	0.26	0.44	0	1
High Inheritance	0.09	0.28	0	1
N	61,692			

Table B.1: Wealth Determinants Pooled. Source: Own computations using GSOEP weights.

Stats	Δ_o	Δ_x	Δ_y	Δ_l	Δ_e	Δ_d
Mean	50,724.50	22,826.46	14,023.30	8,991.29	-1,881.09	1,692.95
	[2,670.30]	[2,696.97]	[1,420.69]	[1,864.32]	[619.37]	[1,278.84]
p5	782.07	571.03	408.91	548.75	-504.91	118.29
	[707.59]	[822.56]	[384.80]	[435.49]	[291.96]	[420.51]
p10	0	0	0	0	0	0.00
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
p20	500	500	450.5	340.31	-129.64	-161.17
	[172.52]	[166.78]	[199.12]	[110.12]	[81.62]	[109.20]
p25	1,606.67	2,155.83	1,834.84	1,038.71	-312.15	-405.57
	[415.77]	[470.39]	[295.71]	[336.27]	[134.52]	[233.15]
p30	3,040.00	3,266.67	2,582.11	1,570.59	-419.57	-466.47
	[623.60]	[657.76]	[305.96]	[472.71]	[169.21]	[400.44]
p40	10,495.00	8,429.05	5,561.17	4,302.81	-826.12	-608.81
	[1,437.70]	[1,418.30]	[862.96]	[1,048.56]	[325.60]	[609.82]
p50	26,768.20	18,684.90	10,309.01	8,894.04	-621.94	103.79
	[2,255.00]	[3,026.99]	[1,523.94]	[2,132.51]	[743.93]	[1,257.53]
p60	42,428.18	20,254.85	11,333.79	8,739.19	-1,203.98	1,385.84
	[2,181.82]	[3,696.91]	[1,863.65]	[2,650.44]	[777.70]	[1,508.55]
p70	55,242.60	20,522.00	12,257.60	8,775.41	-1,981.67	1,470.66
	[2,508.32]	[4,018.41]	[1,961.13]	[2,245.25]	[960.33]	[1,975.26]
p75	64,071.27	23,453.33	15,194.69	8,801.53	-2,524.50	1,981.62
	[3,199.94]	[4,801.77]	[2,305.58]	[2,577.38]	[1,014.19]	[2,079.22]
p80	77,440.97	28,395.17	21,468.29	7,858.56	-3,522.60	2,590.92
	[3,897.86]	[6,982.75]	[3,872.37]	[3,960.22]	[1,404.77]	[2,628.14]
p90	113,735.87	46,867.16	31,014.31	12,622.66	-4,415.93	7,646.11
	[5,638.65]	[7,794.05]	[4,627.99]	[6,068.76]	[1,953.67]	[3,429.97]
p95	177,730.70	74,260.00	49,234.38	19,292.02	-6,533.96	12,267.56
	[10,536.31]	[12,602.80]	[9,714.72]	[10,619.04]	[3,252.75]	[5,350.50]
Gini	-4.63	-2.62	-1.64	-1.35	0.24	0.13
	[1.32]	[1.20]	[0.56]	[0.70]	[0.22]	[0.48]
% Nonpos.	-2.33	-4.96	-3.94	-2.05	0.56	0.46
	[1.13]	[1.03]	[0.55]	[0.67]	[0.24]	[0.45]
% Neg.	-0.9	-0.83	-0.4	-0.62	0.26	-0.06
	[0.76]	[0.62]	[0.28]	[0.37]	[0.14]	[0.25]
% Zero	-1.42	-4.14	-3.53	-1.43	0.31	0.52
	[0.87]	[0.93]	[0.48]	[0.61]	[0.22]	[0.37]
p75p50	-1.05	-1.49	-0.73	-0.78	-0.03	0.05
	[0.41]	[0.45]	[0.21]	[0.28]	[0.10]	[0.13]
p90p50	-2.75	-2.76	-1.34	-1.65	-0.02	0.25
	[0.86]	[1.04]	[0.43]	[0.63]	[0.19]	[0.29]

Table B.2: Detailed Decomposition for 2002. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

Stats	Δ_o	Δ_x	Δ_y	Δ_l	Δ_e	Δ_d
Mean	59,437.98	25,687.37	18,174.46	10,089.90	-4,083.28	1,506.29
	[3,542.45]	[3,393.63]	[1,964.12]	[2,517.85]	[923.53]	[1,702.69]
p5	401.3	978.8	293.74	1,031.40	131.39	-477.73
	[1,105.68]	[4,997.59]	[1,457.36]	[1,578.16]	[1,036.24]	[1,331.06]
p10	0	0	0	0	0	0.00
	[67.97]	[305.64]	[101.66]	[112.32]	[49.73]	[60.31]
p20	118.67	118.67	303.21	287.57	-162.5	-309.61
	[232.76]	[232.76]	[179.05]	[160.96]	[56.44]	[109.49]
p25	1,795.70	2,269.00	1,723.48	1,620.18	-385.51	-689.15
	[585.12]	[533.71]	[289.90]	[373.44]	[112.96]	[212.96]
p30	3,051.00	3,237.00	2,796.66	2,447.96	-707.76	-1,299.85
	[522.20]	[783.61]	[362.28]	[543.21]	[189.09]	[356.72]
p40	9,981.37	8,138.53	6,048.08	5,449.13	-1,642.18	-1,716.50
	[1,222.31]	[1,583.51]	[814.75]	[1,035.83]	[424.50]	[770.68]
p50	22,921.20	14,538.20	9,726.62	8,772.84	-2,819.76	-1,141.49
	[2,165.59]	[2,970.61]	[1,420.76]	[1,772.96]	[882.82]	[1,121.23]
p60	40,898.46	21,470.52	12,850.93	12,199.63	-3,516.52	-63.53
	[2,494.45]	[3,864.14]	[1,902.75]	[2,323.60]	[1,299.11]	[1,400.13]
p70	59,968.04	22,717.80	14,897.96	11,967.31	-4,575.24	427.78
	[2,970.61]	[4,309.68]	[2,266.33]	[2,834.80]	[1,387.23]	[1,810.82]
p75	71,331.44	25,167.00	17,971.98	11,637.37	-5,102.47	660.12
	[3,582.86]	[5,582.26]	[2,596.32]	[3,071.79]	[1,135.48]	[2,030.26]
p80	86,318.20	28,699.40	21,705.37	10,935.63	-5,379.96	1,438.36
	[3,965.31]	[6,226.56]	[4,036.00]	[3,262.58]	[1,672.52]	[2,023.27]
p90	142,513.70	46,639.90	36,671.19	13,173.91	-9,609.16	6,403.97
	[9,852.27]	[11,139.69]	[7,327.62]	[5,493.03]	[3,197.77]	[4,504.10]
p95	211,722.13	77,714.93	55,622.71	19,532.35	-11,287.94	13,847.81
	[10,609.94]	[17,880.59]	[13,464.97]	[11,504.86]	[4,038.56]	[7,485.90]
Gini	-2.71	-3.97	-1.71	-3.07	0.25	0.56
	[1.72]	[2.78]	[1.12]	[1.19]	[0.39]	[0.75]
% Nonpos.	-3.49	-5.66	-4.1	-4.28	0.96	1.76
	[1.16]	[1.72]	[0.63]	[0.90]	[0.29]	[0.53]
% Neg.	-0.99	-0.46	-0.22	-1	0.23	0.53
	[0.86]	[1.32]	[0.47]	[0.61]	[0.20]	[0.38]
% Zero	-2.5	-5.21	-3.88	-3.28	0.73	1.23
	[0.86]	[1.26]	[0.49]	[0.64]	[0.23]	[0.39]
p75p50	-0.43	-1.18	-0.75	-0.86	0.21	0.22
	[0.34]	[0.59]	[0.26]	[0.30]	[0.13]	[0.15]
p90p50	-0.91	-2.57	-1.52	-2.14	0.47	0.63
	[0.78]	[1.25]	[0.63]	[0.71]	[0.29]	[0.38]

Table B.3: Detailed Decomposition for 2007. Figures in square brackets present the standard error from 500 bootstrap samples. Source: Own computations using GSOEP weights.

Chapter 5

Dissertation Summary and Conclusion

The past twenty years have seen a surge in interest in research on the financial situation of private households and the distribution of personal wealth in developed countries. The orientation towards this topic can be understood in the light of an increasingly complex financial environment faced by many households and a shift of longevity risks upon the individual. Owing to these developments, personal investment decisions have become ever more important for the financial well-being of individuals. Moreover, the events of the global financial crisis have highlighted the significance of household finance for the economy as a whole. This doctoral thesis deals with various aspects of the financial situation of German households since the turn of the millennium. The approach of the dissertation is an empirical one, with a heavy emphasis on applied microeconometric methods.

Chapter 2 looks into the causes of the so-called ”stock market participation puzzle” for Germany, i.e. the issue why so few private investors hold risky financial assets. We are particularly interested in the effect of labor income uncertainty on the propensity for stock investment. It stands to reason that households which are exposed to higher levels of this type of risk are more reluctant to take on financial risk. Our identification strategy rests on variation over time. By employing binary panel data methods, we aim at identifying the effects of interest. In a pooled cross-sectional framework, i.e. without controlling for unobserved characteristics, we find plausible results for most covariates such as a positive relationship between stock holding and the levels of household income and risk tolerance. At the same time, we observe that the subjective perceptions and expectations of the future play a role in this context. More importantly, we find that higher labor income risk, as measured by the coefficient of variation of a household's labor income history, is indeed associated with a significantly lower probability of owning risky assets. However, once we control for unobserved heterogeneity via panel dimension variation, the effect of labor income risk diminishes profoundly and is no longer significant. We therefore conclude that the apparent relationship between labor income uncertainty and the propensity to invest in risky equities is partly driven by characteristics that are unobservable to the researcher.

The second study in Chapter 3 of this doctoral thesis is concerned with the overall composition of household financial portfolios. More specifically, I analyze determinants of the risk profiles of such portfolios. To that end, I estimate the shares accruing to three broad risk classes in a joint fractional regression framework as employed by Mullahy (2011). I allow for unobserved household heterogeneity via a correlated random effects approach, similar to the course of action taken by Papke and Wooldridge (2008) for the univariate case. Using pooled regressions, I find that, among others, higher levels of wealth and risk tolerance are associated with considerably riskier financial portfolios. Another noteworthy finding is that risk profiles exhibit a hump-shaped pattern over the life cycle. Once I control for unobserved heterogeneity, many previously significant effects vanish. The contribution of wealth levels and subjective risk appetite, however, remain highly significant and economically meaningful. This suggests that these effects are not primarily driven by joint correlations with unobservable factors. In combination with the findings of Chapter 2, these results emphasize the importance of controlling for unobserved characteristics when analyzing the investment decisions of private households.

Finally, Chapter 4 of the dissertation presents an inquiry into the determinants of the gap in per capita net wealth between East and West Germany. Almost 25 years after German reunification tremendous wealth differences still remain between these two regions. I conduct counterfactual decomposition analyses via the reweighting approach by DiNardo et al. (1996) in order to identify potential explanatory factors for this gap. I find that potential wealth determinants such as permanent income and socio-demographic characteristics help to explain large parts of the wealth gap at the lower to middle part of the wealth distribution but can explain relatively little of the wealth differentials for high net-worth positions. The largest contributing factors in this respect are observable income differentials as well as differences in the labor market situation of Germans residing in the East and the West. Moreover, I find that the wealth gap for younger Germans is considerably smaller and can be explained much better by observable factors. The different labor market situations of young Germans play a particularly important role concerning this matter. I also notice that a substantial gap in home-ownership rates between East and West Germany does matter in this regard. The counterfactual distributions suggest that equalizing the propensities for home equity is associated with a significant reduction in the wealth gap. Yet, this reduction is smaller than the observed gap in home equity. Thus, it is reasonable to assume that geographical differences in housing prices are a non-negligible factor for this issue.

There are many potential areas for future research on the financial situation of households in Germany. One particularly interesting issue that could not be addressed in this thesis are possible consequences of a prolonged zero interest rate environment in the euro zone. It remains to be seen how Germans will respond to these changes in financial conditions and how this will impact their wealth situation. Will the current monetary policy lead to higher home ownership rates, more household debt or more people participating in the stock market? And how will this in turn affect the risk profile of household portfolios and the distribution of personal wealth? These are indeed questions of the utmost importance for each individual household as well as Germany as a whole.

Bibliography

Ackerman, R. A., Fries, G., and Windle, R. A. (2012). Changes in US family finances from 2007 to 2010: Evidence from the survey of consumer finances. *Federal Reserve Bulletin*, *100*, 1–80.

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 L. Guiso, M. Haliassos, and T. Jappelli (Eds.), *Household Portfolios* (pp. 341–388). The MIT Press.

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

Angerer, X. and Lam, P.-S. (2009). Income risk and portfolio choice: An empirical study. *The Journal of Finance*, *64*, 1037–1055.

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 and Finance*, *36*(5), 1320–1335.

Banks, J. and Tanner, S. (2002). Household portfolios in the United Kingdom. In L. Guiso, M. Haliassos, and T. Jappelli (Eds.), *Household Portfolios* (pp. 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.

Barsky, R., Bound, J., Charles, K. K., and Lupton, J. P. (2002). Accounting for the black–white wealth gap: A nonparametric approach. *Journal of the American Statistical Association*, *97*(459), 663–673.

Bauer, T. K., Cobb-Clark, D. h. A., Hildebrand, V. A., and Sinning, M. G. (2011). A comparative analysis of the nativity wealth gap. *Economic Inquiry*, *49*(4), 989–1007.

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

Betermier, S., Jansson, T., Parlour, C., and Walden, J. (2012). Hedging labor income risk. *Journal of Financial Economics*, *105*(3), 622–639.

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.

Biewen, M. (2001). Measuring the effects of socio-economic variables on the income distribution: An application to the East German transition process. *The Review of Economics and Statistics*, *83*(1), 185–190.

Biewen, M. (2014). A general decomposition formula with interaction effects. *Applied Economics Letters*, *21*(9), 636–642.

Bilias, Y., Georgarakos, D., and Haliassos, M. (2013). Equity culture and the distribution of wealth. Technical report, CFS Working Paper.

Bogan, V. L. (2014). *Investor Behavior: The Psychology of Financial Planning and Investing*, chapter Household Investment Decisions, (pp. 83–98). John Wiley and Sons, Inc.

Börsch-Supan, A., Coppola, M., Essig, L., Eymann, A., and Schunk, D. (2008). The German SAVE study: Design and results. Technical report, Mannheim Research Institute for the Economics of Aging (MEA).

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*. MEA discussion papers. Munich Center for the Economics of Aging (MEA).

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

Borsch-Supan, A. H., Bucher-Koenen, T., Coppola, M., and Lamla, B. (2014). Savings in times of demographic change: Lessons from the german experience. *Max Planck Institute for Social Law and Social Policy Discussion Paper*, (18-2014).

Bover, O. (2010). Wealth inequality and household structure: US vs. Spain. *Review of Income and Wealth*, *56*(2), 259–290.

Breuer, W., Riesener, M., and Salzmann, A. J. (2014). Risk aversion vs. individualism: What drives risk taking in household finance? *European Journal of Finance*, *20*(5), 446–462.

Bucher-Koenen, T. and Ziegelmeyer, M. (2014). Once burned, twice shy? Financial literacy and wealth losses during the financial crisis. *Review of Finance*, *18*(6), 2215–2246.

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 and Finance*, *33*(5), 850–860.

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

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

Christelis, D., Georgarakos, D., and Haliassos, M. (2011). Stockholding: Participation, location, and spillovers. *Journal of Banking and Finance*, *35*(8), 1918–1930.

Cobb-Clark, D. A. and Hildebrand, V. A. (2006). The wealth of Mexican Americans. *Journal of Human Resources*, *41*(4), 841–868.

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.

Cowell, F. A., Karagiannaki, E., and McKnight, A. (2013). Accounting for cross-country differences in wealth inequality. CASE Papers 168, Centre for Analysis of Social Exclusion, LSE.

Davies, J. B. and Shorrocks, A. F. (2000). The distribution of wealth. *Handbook of Income Distribution*, *1*, 605–675.

Deutsche Bundesbank (2014a). House prices in 2013 in Germany . Monthly Report 2, Deutsche Bundesbank.

Deutsche Bundesbank (2014b). Ownership structure in the German equity market: General trends and changes in the financial crisis . Monthly Report 9, Deutsche Bundesbank.

Deutsches Aktieninstitut (2015). Aktionärszahlen des Deutschen Aktieninstituts 2014 . Technical report, Deutsches Aktieninstitut (DAI).

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

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, *64*(5), 1001–44.

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.

Eurosystem Household Finance and Consumption Network (2013). The eurosystem household finance and consumption survey: Results from the first wave. Statistics Paper Series 2, ECB.

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.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. *Handbook of Labor Economics*, *4*, 1–102.

Fratantoni, M. C. (1998). Homeownership and investment in risky assets. *Journal of Urban Economics*, *44*(1), 27–42.

Frick, J., Grabka, M., and Hauser, R. (2010). *Die Verteilung der Vermögen in Deutschland: empirische Analysen für Personen und Haushalte*. Forschung aus der Hans-Böckler-Stiftung. Edition Sigma.

Frick, J. R. and Grabka, M. M. (2009a). Wealth inequality on the rise in Germany. *Weekly Report*, *5*(10), 62–73.

Frick, J. R. and Grabka, M. M. (2009b). Zur Entwicklung der Vermögensungleichheit in Deutschland. *Berliner Journal für Soziologie*, *19*(4), 577–600.

Frick, J. R. and Grabka, M. M. (2010). Die personelle Vermögensverteilung in Ost- und Westdeutschland nach dem Mauerfall. In P. Krause and I. Ostner (Eds.), *Leben in Ost- und Westdeutschland* chapter 24, (pp. 493–511). Campus Verl.

Fuchs-Schündeln, N., Krueger, D., and Sommer, M. (2010). Inequality trends for Germany in the last two decades: A tale of two countries. *Review of Economic Dynamics*, *13*(1), 103–132.

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

Grabka, M. M. (2014). Private Vermögen in Ost- und Westdeutschland gleichen sich nur langsam an. *DIW Wochenbericht*, *81*(40), 959–966.

Grabka, M. M. and Frick, J. R. (2007). Vermögen in Deutschland wesentlich ungleicher verteilt als Einkommen. *DIW Wochenbericht*, *74*(45), 665–672.

Grabka, M. M. and Westermeier, C. (2014). Anhaltend hohe Vermögensungleichheit in Deutschland. *DIW Wochenbericht*, *81*(9), 151–164.

Grande, G. and Ventura, L. (2002). Labor income and risky assets under market incompleteness: Evidence from Italian data. *Journal of Banking and Finance*, *26*(2-3), 597–620.

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., Sapienza, P., and Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, *63*(6), 2557–2600.

Guiso, L. and Sodini, P. (2013). Household finance: An emerging field. In G. M. Constantinides, M. Harris, and R. M. Stulz (Eds.), *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.

Haisken-DeNew, J. P. and Frick, J. R. (2005). *DTC Desktop Companion to the German Socio-Economic Panel (SOEP)* (8 ed.).

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 and Organization*, *83*(1), 66–81.

Hauser, R. (2009). Die Entwicklung der Einkommens- und Vermögensverteilung in Deutschland in den letzten Dekaden. In T. Druyen, W. Lauterbach, and M. Grundmann (Eds.), *Reichtum und Vermögen: Zur gesellschaftlichen Bedeutung der Reichtums- und Vermögensforschung* (pp. 54–68). Springer.

Hauser, R. (2010). "Nahblick" und "Weitblick": Erste Schritte zur Erforschung des sozialen und politischen Wandels in den neuen Bundesländern und frühe Prognosen. In P. Krause and I. Ostner (Eds.), *Leben in Ost- und Westdeutschland* chapter 3, (pp. 57–81). Campus Verl.

Hauser, R. and Stein, H. (2006). Inequality of the distribution of personal wealth in Germany, 1973-98. In E. Wolff (Ed.), *International Perspectives on Household Wealth* (pp. 195–224). Edward Elgar Publishing.

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.

HFCS (2013). The Eurosystem Household Finance and Consumption Survey - Results from the first wave. Statistics Paper Series 2, Eurosystem Household Finance and Consumption Network, ECB.

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.

Hochguertel, S. (2003). Precautionary motives and portfolio decisions. *Journal of Applied Econometrics*, *18*(1), 61–77.

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

Hong, H., Kubik, J. D., and Stein, J. C. (2004). Social interaction and stock market participation. *Journal of Finance*, *59*(1), 137–163.

Jenkins, S. P. and Jantti, M. (2005). Methods for summarizing and comparing wealth distributions. ISER Working Paper Series 2005-05.

Kaustia, M. and Torstila, S. (2011). Stock market aversion? Political preferences and stock market participation. *Journal of Financial Economics*, *100*, 98–112.

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 L. Guiso, M. Haliassos, and T. Jappelli (Eds.), *Household Portfolios* (pp. 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. (2014). Regression analysis of multivariate fractional data. *Econometric Reviews*, (ahead-of-print), 1–38.

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.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

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.

Schupp, J., Frick, J. R., Goebel, J., Grabka, M. M., Groh-Samberg, O., and Wagner, G. G. (2009). Zur verbesserten Erfassung von Nettohaushaltseinkommen und Vermögen in Haushaltssurveys. In *Reichtum und Vermögen* (pp. 85–96). Springer.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: A unified framework based on the shapley value. *Journal of Economic Inequality*, *11*(1), 99–126.

Sierminska, E. M., Frick, J. R., and Grabka, M. M. (2010). Examining the gender wealth gap. *Oxford Economic Papers*, *62*(4), 669–690.

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. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), 449–472.

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.

Wagner, J. (2008). Exports and firm characteristics: First evidence from fractional probit panel estimates. Technical report, University of Lüneburg Working Paper Series in Economics.

Wolff, E. N. (2012). The asset price meltdown and the wealth of the middle class. NBER Working Paper 18559, National Bureau of Economic Research.

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