THE DISTRIBUTION OF PENSION WEALTH AND THE PROCESS OF PENSION BUILDING

AUGMENTING SURVEY DATA WITH ADMINISTRATIVE PENSION RECORDS BY STATISTICAL MATCHING

vorgelegt von

Anika Rasner

aus Korbach

Von der Fakultät VII Wirtschaft und Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
Dr. rer. oec.

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Axel Werwatz
Berichter: Prof. Dr. Gert G. Wagner
Berichter: Prof. Dr. James W. Vaupel

Tag der wissenschaftlichen Aussprache: 30. November 2011

Berlin 2012

D 83

Acknowledgments

Completing a dissertation feels like a marathon, and I would not have been able to complete this journey without the aid of countless people over the past years. I must first express my gratitude towards my advisors Prof. Dr. Gert G. Wagner and Prof. Dr. James W. Vaupel for their support, supervision and guidance.

Gratitude is also expressed to Dr. Joachim R. Frick and Dr. Markus Grabka for this priceless collaboration in order to get the matches right from the beginning to the end of this project.

Many thanks go to Dr. Jutta Gampe who became a great mentor and always had the best piece of advice and Dr. Olaf-Groh Samberg for his patience in helping me with loops and other seemingly insurmountable obstacles. Thank you to Silvia Leek who is responsible for all the nice graphs in this dissertation. I gratefully thank Prof. Don Taylor at the Sanford School of Public Policy at Duke University for giving me the opportunity to work with restricted HRS data and Anne Fletcher for her great support during my research stays in the summer of 2009 and 2010. My stays at Sanford greatly benefited my studies that were made possible because of the generosity and hospitality of Prof. Sunny Ladd and Prof. Marjorie McElroy who are always a source of inspiration and great role models for me. Special thanks go to Edgar Kruse and Dr. Ralf K. Himmelreicher who provided me access to administrative pension records and supported the statistical matching project.

Thank you to all my friends and colleagues at the DIW Berlin and the Max Planck Institute for Demographic Research for all your help and support.

Lastly, and most importantly, I wish to thank my parents and brother for their confidence and support. To them, I dedicate this thesis.

Table of Content

List of Figures

List of Figures (Appendix)

List of Tables

List of Tables (Appendix)

List of Abbreviations

AA	Actuarial Adjustment
AF	Adjustment Factor
AHEAD	Asset and Health Dynamics Among the Oldest Old
AIME	Average Indexed Monthly Earnings
ASID	Alterssicherung in Deutschland (*Old-Age Pension Schemes in Germany*)
AVID	Altersvorsorge in Deutschland (*Retirement Pension Provision Schemes in Germany*)
CIA	Conditional Independence Assumption
DB	Defined-benefit
DC	Defined-contribution
DEÜV	Datenerfassungs- und Übermittlungsverfahren (*Electronic submission of earnings data*)
EP	Earning points
FDZ-RV	Forschungsdatenzentrum der Rentenversicherung (*Research Data Center of the statutory pension insurance*)
GED	General Education Development
HRS	Health and Retirement Study
IABS	IAB Beschäftigtenstichprobe (*IAB employment subsample*)

OASDI	Old-Age, Survivors and Disability Insurance
OLS	Ordinary Least Squares
OM	Optimal Matching
PAYG	Pay-as-you-go
PIA	Primary Insurance Amount
PV	Pension Value
RATSWD	Rat für Sozial- und Wirtschaftsdaten (*German Data Forum*)
SSA	Social Security Administration
SAPA	Sample of Active Pension Accounts
SES	Social Employment Situation
SGB	Sozialgesetzbuch (*Social Code*)
SHARE	Study of Health, Ageing and Retirement in Europe
SOEP	Socio-Economic Panel
SSN	Social Security Number
SSW	Social Security Wealth
SUF	Scientific Use File
UVIS	Univariate Imputation Sampling
VVL	Vollendete Versichertenleben (*Completed Insurance Biographies*)

1 Introduction

1.1 Motivation

Over the last decade, the German system of old-age provision has undergone a profound transition. Its overhaul not only has considerable effects for the system's solvency, but also for the distribution of pension entitlements of future retirees.

One source of these changes has been fundamental demographic shifts experienced by all developed countries in recent decades (Barr 2006). They result from an unprecedented rise in life-expectancy (Christensen et al. 2009) as well as the seemingly irreversible downward trend in fertility rates (Kreyenfeld 2009). In these rapidly aging societies, a decreasing number of people of working-age supports an increasing number of retirees (OECD 2011). Everything else being equal, the growing life expectancy implies that the period of pension receipt becomes longer and longer. To secure the overall financial viability of the German public pension system – especially with the baby boomers approaching retirement - policymakers had to halt the trend toward early retirement through the introduction of actuarial adjustments and gradual raises of the statutory retirement age (Brugiavini 2001; Gruber and Wise 1999). Pension types with special rules of eligibility such as the old-age pension due to unemployment will cease to exist. Less stable family lives are another demographic trend of concern, especially for women. Because of their comparatively low own pension benefits, many of them were best protected from old-age poverty through the high pension benefits received by their husbands or in case the husband dies through the receipt of survivor's benefits (Rasner 2006; Stegmann 2009). In case of a divorce, this protection ceases to exist (Smock et al. 1999). More than ever, women are forced to make pension provisions on their own.

A second obvious source of these changes with corresponding repercussions for the individual's pension rights has been a series of legislated reforms governing both public and private pensions intended to obtain a financially sustainable public pension program (Bonin 2009). These reforms will result in marked cuts of net replacement levels and slow down the future growth of pension benefits. The sustainability factor, introduced in 2004, reduces the average net replacement level according to the changing ratio of contributors and beneficiaries within the statutory pension insurance. Because of the modified indexation formula, benefits no longer grow in line with gross earnings (Börsch-Supan et al. 2003). With inflation being high, the slower growth in pension benefits implies falling purchasing power of the retirement income of the elderly. When compared to today's pensioners, future retirees with the same life cycle labor supply will render less pension benefits. Policymakers expect them to compensate for these cutbacks through investments in private pensions. Because these investments are voluntary, the government tries to promote them through public subsidies and tax breaks (Berner 2006). Policymakers also promote easier access and more coverage with occupational pensions, given that this type of retirement provision is strongly segregated along gender and occupational lines (Rabe 2007). Taken together, these changes in the public/private mix result in a paradigmatic shift moving from a strong reliance on the public pension scheme to a multi-pillar system of old-age provision in which private and occupational benefits supposedly complement benefits from the public pillar (Arza 2008). Further distributional consequences will arise from the deferred taxation for all types of retirement income introduced with the Old-Age Income Law (*Alterseinkünftegesetz*) in 2005 (Fehr and Jess 2007).

A third source of these changes that affects the individual's ability to accumulate pension rights has been the trend towards more heterogeneous working patterns that started in the early 1980s, but accelerated after German reunification. Legislation aimed at reducing the persistently high unemployment rates by promoting employer and employee flexibility in labor markets (Eichhorst and Marx 2011). These changes involve greater job insecurity, less social protection and lower incomes. More flexibility in the labor market but also personal preferences for a better work-life balance result in a higher prevalence of (marginal) part-time employment (Holst 2010).[1] Taken together these changes lead to an erosion of regular work (Diekmann and Jann 2004). At the same time, the school-to-work transition has become precarious for some young adults, whereas a still increasing number of adolescents enter higher education, thus postponing their transition into the labor market (Brzinsky-Fay 2007). An adverse effect of these changes has been that workers following more flexible paths accumulate less social security wealth[2] and wind up retiring with lower benefits than they would have reaped had they been continuously full-time employed. Therefore, a declining number of individuals meet the paradigm of the *standard pensioner*.[3]

Explaining how these three sources of change are interrelated and understanding the distributional effects of their interplay is of innate importance not only for policymakers and researchers, but also for future retirees. Under the new system, they have to make more informed choices and face higher risks in planning their long-term well-being as compared to today's retirees (Shuey and O'Rand 2006). For policymakers and researchers, the quest for the system's solvency raises concerns about the adequacy and equitable allocation of future benefits. It is to expect that

---

1. Differences in the preference for part-time work persist between women in East and West Germany (cp. to Holst 2010)
2. Social security wealth, also public pension wealth, is the sum of pension rights and individual accumulates over his or her working life. The accumulation of entitlements starts with the first job that is subject to social insurance contributions or payroll taxes and ends with retirement. Among other factors (e.g. age at retirement) the total social security wealth determines the public pension benefit a person receives.
3. The standard pensioner is a statistic, which describes the relative income position of pensioners compared to the average income of the workforce in a given year. It assumes that the pensioner's wage was equal to the average wage earned in Germany each year for the last 45 years. The net replacement level is an indicator for pension system's generosity.

inequalities in social security wealth will become more pronounced with some groups being more vulnerable than others (Schmähl 2011). Therefore, it is quite possible that these changes will undo the successes Germany has had in alleviating poverty amongst the elderly (Goebel and Grabka 2011) unless policymakers enact new measures that counteract the adverse effects of recent policy reforms and changing employment patterns.

1.2 Problem Statement

In the light of this changing policy context, it becomes crucial to gain a thorough understanding of the role social security wealth plays in the individual's total wealth holdings and how work and family choices as well as policies affect the path of pension rights accumulation. Social security wealth is a substantial source of wealth that grows in importance over the adult's working life. Especially with the recent reforms phasing in, it becomes indispensable to study which demographic groups succeed to accumulate sufficient social security wealth and which groups do not and therefore run risk of becoming poor in old age. For those individuals with insufficient social security wealth on their own, it is critical to know whether they are capable to compensate this lack with other types of wealth.

But it is not only the accumulated social security wealth that matters, but also the process of pension building over the individual's life cycle. Under the new multi-pillar pension system, old-age provision is much more individualized. It is therefore critical to take on a dynamic perspective that allows for the identification of individual-level factors or transitions that either increase or limit the person's ability to accumulate pension rights. One such factor could be the birth of a child. Incentives and disincentives resulting from policies also matter for the path of pension building. These policies could be tax rules that promote asymmetric labor supply in couples with corresponding consequences for the accrual of pension rights.

Quantifying the distributional effects of the reforms, identifying demographic groups that run risk of having insufficient retirement provisions and understanding factors that matter for successful paths of pension accumulation are crucial for future policymaking. It is, however, also a complex undertaking that requires the use of suitable micro-data (Frick et al. 2010a). Ideally, the data have to meet the following criteria: First, data should be representative of the total working age and retired population; second, data should provide complete life cycle earnings and employment trajectories as well as information on social security wealth and other wealth holdings; and obviously, the data should be accessible to researchers.

The restructuring of old-age provision goes along with a growing number of data sources that collect information on issues of work, retirement, and aging in Germany (Jürges 2010). These data range from government commissioned studies over population-representative surveys to administrative records. Despite their similar focus, data differ with respect to the population of interest, the information available and their accessibility. Hence, not all data are equally adequate to give answers to the pressing policy challenges ahead.[4]

Starting in the late 1980s, the government launched two studies: Old-Age Pension Schemes in Germany (*Alterssicherung in Deutschland*, henceforth ASID) and Retirement Pension Provision Schemes in Germany (*Altersvorsorge in Deutschland*, henceforth AVID). The ASID study covers a representative sample of individuals age 55 and older and their spouses. The sample in 2007 comprises 28,896 individuals. The survey collects information on the types and level of income available to the elderly popula-

---

[4] Note that the data sources named here are by no means exclusive. There are other data available that also allow for analyses of employment biographies old-age income, among those the German Micro Census, the Survey of Income and Expenditure, SAVE (*Sparen und Altersvorsorge in Deutschland*), EU-SILC, as well as the German Ageing Survey (DEAS).

tion, the accumulation of income from different sources as well as the composition of old-age income from public, private and occupational pensions. The number of contextual variables for the identification of determinants of the individual's income situation is limited (Kortmann and Halbherr 2008), given that only cross-sectional data are available.

The AVID study collects comprehensive biographical information in addition to data on pension entitlements in public, private and occupational pension schemes for a representative sample of German citizens between 40 and 59 years and their spouses (n = 12,536). Individuals can enter the sample population only if they hold a pension account in the public pension program and validated the information kept in the account (Frommert and Heien 2006).[5] A micro-simulation model projects entitlements from all retirement income schemes to age 65, which allows for the analysis of the level, distribution and composition of old-age income for future retirees (and their spouses) in the light of their biographical choices. A major shortcoming of the AVID is that it systematically excludes certain segments of the population, such as civil servants and self-employed that are by definition not protected under the umbrella of the statutory pension insurance. But also migrants are not part of the AVID sample population.

Both studies collect valuable data that address many relevant policy issues concerning the material well-being of the elderly and the retirement provisions of future retirees. However, data are not available for the scientific community outside the statutory pension insurance (*Deutsche Rentenversicherung*) or the Federal Ministry of Labor and Social Affairs (*Bundesministerium für Arbeit und Soziales*) or they are released with a considerable time lag.

---

5 Married spouses are by definition part of the sample population, irrespective of their age or citizenship. Married spouses also don't have to hold a pension account.

Hence, data are greatly underused.[6]

Unlike ASID and AVID that only collect cross-sectional information, panel studies like the German Socio-Economic Panel (*Sozio-ökonomisches Panel,* henceforth SOEP) and the Study of Health, Aging and Retirement in Europe (SHARE) have a longitudinal focus. With a total of 26 waves, more than 11,000 households and 20,000 individuals interviewed, the SOEP is the largest and longest running panel study that provides representative data of private households in Germany (Schupp 2009; Wagner et al. 2007). Employment and retirement are just some among the multitude of topics in the survey. The SOEP collects retrospective data on individual's family and employment histories, but also detailed information on earnings, pension benefits or other transfer income for each year a person participates in the study. Special modules ask extensive questions about certain topics, such as wealth at the individual and household level. A shortcoming of the survey is that it lacks information on earnings prior to the first interview as well as information on social security wealth.

SHARE is a European-wide comparative panel study representative for the non-institutionalized population aged 50 and older (Börsch-Supan et al. 2011). In 2006/07 the sample covers a total of 2,568 individuals. The data provides detailed information on the employment behavior and economic situation of the elderly, but also on their social networks and health status. The harmonized study design allows for cross-country analyses across fifteen nations. The prospective study design in the first two SHARE waves goes along with a lack of retrospective information on life-histories. Because early life events give a better understanding of late-life outcomes, the study extension SHARELIFE collects retrospective life-

[6] Hauser stresses that the fact that data are not available for independent scientific research outside these institutions contradicts the recommendations of the Commission for the Improvement of the Data Infrastructure (2011).

history data and link SHARE data with administrative pension accounts (Schröder 2011). SHARE provides a detailed picture of the living conditions of the elderly population, but not of younger birth cohorts who are the first to experience the fully phased in reforms of the system of old-age provision.

During the last decade, the statutory pension insurance opened up access to administrative pension records. The Sample of Active Pension Accounts (*Versicherungskontenstichprobe*, henceforth SAPA) and the Sample of Completed Insurance Biographies (*Vollendete Versichertenleben*, henceforth VVL) are datasets maintained by the statutory pension insurance (Stegmann 2006a; Stegmann 2008). The SAPA data covers a representative sample of all individuals holding a pension account, which can be very young individuals with only few entitlements, but also individuals who are close to retirement or recently retired. The SAPA raw data covers a one percent random sample of all pension account holders. The Scientific Use Files (SUF) of SAPA draws a 25 percent subsample of this raw data. The SUF VVL covers a 25 percent random sample of all retirement inflows in a given year. Both datasets provide longitudinal information on various kinds of pension-relevant activities, such as employment or child care, monthly information on accumulated pension entitlements as well as information on the individual's social security wealth. Not all groups hold a pension account in the statutory pension insurance. Hence, certain segments of the population are systematically excluded. The number of covariates that go beyond the administrative purpose of the statutory pension insurance is largely limited. For example, there is no information on the household context or other types of income.

This quick glance illustrates that all data have their respective merits and drawbacks, but none of them meets all criteria necessary for distributional analyses. *Table 1* gives a brief synopsis of the relevant data sources in retirement research in Germany. It shows that data either are not representative of the total population because they systematically exclude certain demographic or occupational groups or they restrict the target population to a specific age range. Or data lack relevant covariates for example complete life cycle information on earnings or employment trajectories. Finally, some datasets are simply not accessible for researchers.

Table 1 Synopsis of Available Data in Retirement Research

	Old-Age Pension Schemes in Germany (ASID)	Retirement-Pension-Provision in Germany (AVID)	Socio-Economic Panel (SOEP)	Study of Health, Aging and Retirement in Europe (SHARE)	Sample of Active Pension Accounts (SAPA)	Completed Insurance Biographies (VVL)
Target Population	Individuals aged 55 years and older (plus spouses)	Individuals between age 40 and 59 (plus spouses)	Private households in Germany	Individuals 50 years+ in Europe	Individuals holding a pension account	Individuals that retired within a given year
Years of Observation	1986, 1992, 1995, 1999, 2003, 2007	1996, 2005	Each year since 1984 in West Germany and 1990 in East Germany	2004, 2006, 2008	2005, 2006, 2007, 2008, 2009	2004, 2005, 2007
Study Design	Cross-Section	Cross-Section	Panel Study	Panel Study	Cross-Section	Cross-Section
Primary Objective	Income situation of the elderly	Projection of retirement-pension-provision of the elderly	Monitor living conditions of individuals & households in Germany over time	Monitor employment behavior, material well-being, health, and social networks of the elderly	Administrative purpose; monitoring of life cycle entitlements over time	Administrative purpose; monitoring of public pension benefits and life cycle entitlements over time
Advantages	Various types of income, transfer payments, pension benefits	Comprehensive information on income from different pension schemes Life-history data	Longitudinal information on work & family choices Earnings and transfer income for years of interview Detailed wealth data	Information on economic situation and employment behavior Retrospective life-history data; link with pension data (not yet available)	Complete life cycle pension-relevant earnings and information on other pension-relevant periods Information on social security wealth	Complete life cycle pension-relevant earnings and information on other pension-relevant periods Information for pension benefit calculation
Disadvantages	Only cross-section Only elderly population Limited number of context variables	Limited age range of target population Systematic exclusion of migrants, civil servants, and self-employed	No information for years prior to first interview or in case of refused interview No social security wealth	Only population aged 50 and above Linked data not yet available	Lack of relevant contextual information (other income, household context)	Lack of relevant contextual information (other income, household context)
Availability of Scientific Use Files	Not immediately, but with delay (most recent 2003)	No	Yes	Yes (except for linked data with pension records)	Yes	Yes

Source: Author's Illustration

This thesis seeks to overcome the limitations of existing data through the combination of administrative records maintained by the statutory pension insurance with population representative survey data from the SOEP. Ideally, these data could be linked over a common identifier, so-called *record linkage* (Winkler 2006). Because of data confidentiality concerns, this procedure is infeasible without the written consent of survey respondents. The requirement for consent adds a potential source of bias given that some individuals might refuse to consent, whereas others provide flawed information that poses an impediment to a successful link (Haines and Greenberg 2005; Jenkins et al. 2006). Statistical matching is a viable and less costly alternative that allows to link statistically similar, but not identical persons (Rässler 2002; Rubin 1986). The matching quality depends on the availability of matching variables in both datasets that are measured in comparable ways. For each individual in the *recipient file*, the matching algorithm selects the most similar person in the *donor file* based on the common matching variables. However, some individuals find better matches than other. Compared to record linkage, statistical matching potentially compromises the exactness of matches. Record linkage is no option because it might amplify the risk of panel mortality. Therefore, this thesis opts for statistical matching as a way to link survey and administrative data.

1.3 Contributions

This work serves a twofold purpose: Its methodological contribution is to prepare, test, and implement a statistical matching procedure to link administrative pension records with population representative survey data from the SOEP (Chapter 2, Chapter 3, and Chapter 4). The second purpose is to use these new data in two applications dealing with research questions that were previously infeasible to analyze. One of the applications analyzes the quantitative relevance of social security wealth in the individual's total wealth holdings (Chapter 4). The other application investigates the interdependencies of the process of pension building and marital trajectories in Germany and the United States (Chapter 5).

Chapter 2 serves the preparation of the statistical matching providing a systematic account of the qualities and limitations of administrative pension records and survey data. In administrative data, this methodological groundwork is still quite rare and lags behind advances in survey methodology (e.g. multiple imputation for missing information, etc.). Based on the Sample of Active Pension Accounts 2007, the analyses test the representativity of administrative data through a comparison with official population statistics. Furthermore, I try to assess how administrative working routines in the statutory pension insurance affect data quality and addresses whether these routines raise concerns about accuracy and selectivity in the data. This work also serves the purpose to scrutinize the measurement concepts of variables. Since data are not for research in the first place, but serve the administration of the pension program, they have to be interpreted in the respective legal context. A thorough understanding of what these variables measure is crucial in order to determine whether they are compatible with concepts in survey data and therefore can serve as matching variables. The analyses also show how the lack of important context variables in administrative data possibly leads to flawed conclusions.

In Chapter 3, we test the feasibility of the statistical matching of the Scientific-Use-File Completed Insurance Biographies (SUF VVL 2004) and the Socio-Economic Panel. This joint work with Ralf K. Himmelreicher, Markus M. Grabka and Joachim R. Frick elaborates a blueprint for the preparation of a statistical matching of administrative pension records and population representative survey data. This blueprint involves the identification of matching variables and the specification of matching populations. It compares the statistical distributions of potential matching variables in both datasets and harmonizes them if necessary. Further tests assess whether we find the same underlying relationships between the variables of interest in both datasets and whether it is possible to replicate SOEP results with data from the SUF VVL 2004. This feasibility study is an important perquisite for the actual implementation of the statistical matching. It gives a clear indication of how to specify the sample population, how to deal with the administrative characteristics of the data and which variables work as matching variables and which do not.

Chapter 4 brings the two strands of research in this dissertation together, namely the actual implementation of the statistical matching and the application of the matched data. The first part of Chapter 4 deals with the implementation of the statistical matching of survey data from the Socio-Economic Panel and administrative records from the Sample of Active Pension Accounts. This joint work with Joachim R. Frick and Markus M. Grabka tests a total of nine matching algorithms and four matching and imputation techniques for a sample of retirees. The unique properties of the data allow for a straight control of the quality of matches under each algorithm and under each technique. The public pension benefit retirees report in the SOEP serves as a benchmark to which the matched or imputed benefit is compared. Based on three evaluation criteria, the study decides which technique performs best. The availability of a benchmark is unique and gives this study an advantage over other research applying statistical matching that have no quality control whatsoever.

The successful statistical matching enhances the analytical power of both datasets beyond the research potential each dataset provides alone: The SOEP benefits from pension program details available in administrative records, such as the individual's social security wealth or complete life cycle information on monthly pension entitlements that are otherwise infeasible to collect in survey routines. In turn, the administrative pension records, gain from the rich longitudinal information available in SOEP data – both, at the individual and household level. The result is a unique population representative dataset well suited for distributional analyses. Moreover, the data also allow researchers to trace the consequences of life cycle work and family choices through to outcomes in old age.

The second major contribution of this thesis is to employ the newly matched data in two applications that were previously infeasible based on the existing data. The first application in Chapter 4 presents an extended wealth inequality analysis that considers social security wealth. Despite its quantitative relevance and the growing importance over the individual's life cycle, social security wealth has been thus far omitted from wealth analyses, at least in Germany. In the comparison of wealth holdings across occupational groups or in cross-country analyses across welfare regimes (Frick et al. 2010a) this omission leads to biased estimates. For example, dependent employees and self-employed differ greatly in how they accumulate pension wealth (Wagner 1980). The statistical matching corrects for this shortcoming and makes this information available in SOEP data, therefore allowing an analysis of extended net worth. The analyses shed light on the role social security wealth plays in the individual's total wealth holdings and identify groups that rely strongly on social security wealth and which groups do not. It also illustrates whether the consideration of social security wealth results in a reduction of overall wealth inequality and how social security wealth changes the distribution of net worth along occupational lines comparing civil servants, self-employed and dependent employees. Overall, these analyses provide a less biased and therefore more complete picture of the level and socio-economic structure of the wealth distribution in Germany.

The second application (Chapter 5) analyzes the impact marital trajectories have on retirement outcomes and the process of pension building of women in Germany and the United States. So far, research documents the strong ties between marital status and labor supply as well as the short- and medium-term effects of marital transitions, whereas evidence on how marital trajectories affect the accumulation of pension rights over the life cycle is limited. These studies were infeasible because of the lack of adequate and comparable longitudinal data. For Germany, the analysis uses the matched *SOEP-SAPA* dataset. For the U.S., the reference data comes from the *Health and Retirement Study* linked with *2004 Permissions: Wage and Self-Employment Income (W2)*, an administrative dataset maintained by the Social

Security Administration. Comparing Germany and the United States is valuable because both countries have mature public pension programs with a rather similar rationale, but exhibit considerable variation in their welfare state conceptions. Using sequence analysis and optimal matching techniques, the study clusters similar marital trajectories in Germany and the U.S. It then compares the retirement outcomes, namely pension benefits from the public pension program and the total retirement income, as well as the paths of pension building across marital clusters. Cross-country studies can therefore help to identify pervasive incentives in welfare states that perpetuate gender-specific employment patterns, economic dependencies in couples, and insufficient financial resources in later life depending on their marital choices.

1.4 Outline of the Thesis

Chapter 2 sets out to give a systematic account of the strengths and weaknesses of administrative pension records and discusses how these limitations might affect the quality of research on the material well-being of (future) retirees. Chapter 3 introduces the idea of a statistical matching of administrative pension records and survey data. It gives a generic blueprint for the preparation of a statistical matching, which involves the identification and harmonization of matching variables and matching populations in both data sets as well as a replication exercise that tests the feasibility of a statistical matching. The aim of Chapter 4 is to identify the best statistical matching or imputation technique for the data at hand and to implement the best performing technique to augment survey data with administrative pension records. Section 4 also presents the results of the first application that uses the newly assembled data for a wealth inequality analysis. The second application in Chapter 5 uses the matched data in a comparative study that analyzes the interdependencies between marital trajectories and retirement outcomes as well as the process of pension building of women aged 50 to 80 years in Germany and the U.S. Section 6 concludes and provides on outlook on future research needs.

2 Analysis within Limits: The Power of Administrative Data

2.1 Introduction

In recent years, researchers from various fields gained access to a large array of administrative data. Typically, these data are appreciated for the large number of observations that exceed cases in survey data by far. The large number of observations also allows for subgroup analyses or the focus on a specific geographic area. Since the collection of data is tied to the working routines of the respective agencies and serve administrative purposes such as the determination of benefit eligibility, the data presumably are reliable and precise.[7] These data also don't suffer from problems inherent in survey data, such as measurement and recall errors or item- and unit non-response because individuals are not directly involved in data collection (Wirth and Müller 2006). Panel attrition – another issue in surveys - doesn't concern administrative data because individuals are by definition part of the sample population or not and they can't opt out of the sample. By and large, administrative data expand research opportunities a great deal.

Despite the reputedly high accuracy and completeness of administrative data, their use in empirical research in the social sciences and economics was not well established in Germany, until recently. The limited utilization was partly due to legal barriers that long time restricted access to micro data for research purposes (Heese 2004).[8] Further complications

---

[7] Because data are the result of working processes of government agencies, they are also called process-produced data. In this paper, we use the terms administrative data and process-produced data interchangeably.

[8] The Social Security Data Protection Act (*Sozialdatenschutzgesetz*) provides rules for the collection, processing and use of highly sensitive personal and private data in the branches of the social insurance system. The use of the data is considered to be an in-

emerged from technical difficulties in accessing and the high costs involved in providing data to researchers (Rolf et al. 2008). Starting in 1999, concerted efforts of the German Federal Ministry of Education and Research (*Bundesministerium für Bildung und Forschung*), the Commission for the Improvement of the Statistical Infrastructure (*Kommission zur Verbesserung der informationellen Infrastruktur*), and the German Data Forum (*Rat für Sozial- und Wirtschaftsdaten*) advanced the data infrastructure. They advocated for the branches of the social insurance system (e.g. the statutory pension insurance) to provide an institutionalized access to anonymized micro data given these institutions keep datasets to administer large-scale government programs (Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik 2001).[9]

Researchers appreciate administrative data with the number of projects based on these data being on the rise and the availability of new datasets increasing year by year (Bender et al. 2008). Most studies point out the advantages administrative data have over survey data, but touch only marginally on their methodological limitations (Rat für Sozial- und Wirtschaftsdaten 2010).[10] Our knowledge is particularly limited as to how these restrictions affect research results (Bender et al. 2008; Fitzenberger et al. 2006). Assessing the quality of these data, however, is crucial and justified because they build a strong foundation for policy planning and deci-

---

fringement of personal rights, in particular the right of informal self-determination. Laws that are safeguarding the use of social security data are laid down in the provisions on social security data confidentiality in Section 35, Book I of the Social Code (SGB I), on social security data protection in Sections 37-85a, Chapter 2, Book X of the Social Code (SGB X) and supplementary data protection provisions in other sections of the Social Code (Bundesministerium für Arbeit und Soziales 2006).

9 Data come in form of Scientific Use Files for which the legal protection through the Social Code doesn't apply they are anonymized. According to the legal definition of Section 67 Chapter 8 of Book X of the Social Code, data are anonymized if they have been perturbed in such a way that the identity of individuals can only be inferred with an unreasonable effort in terms of time, money, and manpower.

10 One of the recommendations of the German Data Forum is to foster methodological research for administrative data (Rat für Sozial- und Wirtschaftsdaten 2010).

sion making and help to quantify the distributional effects of recent reforms or reform proposals.

This paper asks how suitable administrative data are for the analysis of the material well-being of the elderly and future retirees. It provides a systematic account of what we can and cannot do with administrative data, more specifically the Sample of Active Pension Accounts (*Versicherungskontenstichprobe*, henceforth SAPA) a dataset maintained by the German statutory pension insurance (*Deutsche Rentenversicherung Bund*). Like in survey data, there are four potential weaknesses that limit the applicability of the data: lack of representativeness, selection bias, concerns with validity and reliability as well as the lack of relevant covariates. Clearly, these points apply whenever data are collected, but they play out differently in administrative data than they do in survey data. Also, in survey data much methodological groundwork has been done to deal with these shortcomings (e.g. multiple imputations for item non-response) and to improve data quality, whereas our understanding is limited as how to handle them in administrative data. Knowing more about prevalence and scope helps us understand the mechanisms by which these flaws affect research results and therefore enhance data quality and our ability to generalize from the data.

The opening section gives a brief sketch of the German public pension scheme and describes the Sample of Active Pension Accounts. Section 2.4 assesses how SAPA data fare in terms representativeness. With official population statistics from the German Federal Statistical Office, we compare population distributions by age, gender, citizenship status, and region. Section 2.5 discusses how working routines in the administration affect data quality and raise selectivity concerns. Section 2.6 sets out how measurement concepts implemented in administrative data compare to those in survey data and whether information is valid and reliable. In Section 2.7, we discuss whether and if so, how missing variables limit the scope and explanatory power of the data. Section 2.8 draws conclusions about the applicability of administrative data.

2.2 The German Statutory Pension Insurance

A brief sketch of the German statutory pension insurance helps to better appreciate the content of the Sample of Active Pension Accounts and to understand the data's qualities and limitations. The statutory pension insurance is one of the five branches of the German social insurance system.[11] Introduced in 1889, Germany has the oldest social security program in the world. Originally fully funded, the German parliament decided to gradually convert the pension system into a pay-as-you-go (PAYG) scheme (Börsch-Supan 2000b). As the administrator of the largest entitlement program, the statutory pension insurance provides benefits to replace the loss of income due to old age, disability and death (survivor's and orphan's pensions). The agency's coverage is almost universal, because over the life cycle nearly every person living in Germany comes into touch with the system for at least once. However, the public pension scheme excludes certain occupational groups, such as civil servants, certain chambered professions (e.g. lawyers or medical doctors), agricultural workers, and self-employed as well as illegal workers (Breyer and Kifmann 2004). These groups are either covered in special occupational schemes (e.g. civil service pension schemes) or invest in private pension arrangements.

Four main principles characterize the German public pension scheme: First, the system is work-related in that the individual's work history determines the level of benefits upon retirement (Breyer and Kifmann 2004).[12] Second, the scheme is contribution-based. In almost all jobs, the employee pays a certain share of his or her gross earnings into the public

---

[11] The German social insurance system includes the statutory pension insurance, unemployment insurance, health insurance, accident insurance and long-term care insurance.

[12] The work history determines the individual's pension benefits, but also benefits for spouses and children in case of death.

pension scheme. The employer equally matches these contributions. The individual's final pension benefit is roughly equivalent to the joint contributions paid (*principle of equivalency*). Third, benefits are earned entitlements and not subject to a means test. This principle assures that additional income, wealth or other pension benefits do not count against the individual's public pension benefit. Fourth, laws define how pension rights are acquired. These laws set out objective criteria that determine how gainful employment and certain forms of non-employment translate into pension entitlements. Over time, these rules are subject to change.

2.3 The Sample of Active Pension Accounts

For the first pillar, the most important scheme of old-age provision, the statutory pension insurance collects detailed data on the working and beneficiary populations. Today, the agency keeps more than 52 million active pension accounts and pays pension benefits to almost 24.5 million beneficiaries (Deutsche Rentenversicherung 2009).[13] The statutory pension insurance makes excerpts of these records available for research purposes. The Sample of Active Pension Accounts is one of those datasets.[14] The SAPA is a one percent disproportionate stratified random sample covering working and recently retired individuals.[15] [16] For 2007, the dataset com-

13 In 2008, approximately 17.4 million individuals received old-age benefits, 1.6 million disability benefits, and 5.8 million survivor's benefits. One beneficiary can receive multiple benefits (e.g. old-age and survivor's benefits). SAPA data include old-age and disability benefits (Deutsche Rentenversicherung 2009).

14 For this paper I have exclusive access to SAPA raw data. Researchers outside of the statutory pension insurance only have access to the smaller Scientific Use File FDZ-RV SUF VSKT 2007 available at www.fdz-rv.de.

15 A disproportionate stratified random sampling procedure was chosen to obtain enough cases allow for the analyses of smaller sub-segments of the population. The data are divided into 12 homogeneous subpopulations (strata) based on the following criteria: gender, citizenship and branch (distinction of blue-collar, white collar employees, and miners). Data are then further subdivided by the year of birth (for more specifics on extrapolation compare to Richter and Himmelreicher 2008).

16 The working population includes actively and passively insured individuals. A passively insured person already accumulated pension entitlements but did not have any

prises 568,865 records providing all information necessary to determine individual eligibility, accumulated entitlements and benefit amounts.

The SAPA is both cross-sectional and longitudinal in nature. The cross-sectional part of the data refers to the year 2007 - the year the sample was drawn. This so-called *fixed file* provides socio-demographic information and variables for a fictitious pension benefit calculation as well as detailed data on the composition and level of benefits. The longitudinal data section provides several files with pension-relevant life cycle information. These longitudinal files give monthly information on various pension-relevant matters of fact. For each observation, information starts at age 14 and ends in the year the sample was drawn (this means a maximum of 624 months for individuals observed at age 67 in the year 2007) (Stegmann 2008).

Among other things, these longitudinal files provide unusually strong lifetime records of earned pension credits (so-called *earning points*) that reflect the individual's monthly earnings. The earnings information is capped at the maximum contribution ceiling (*Beitragsbemessungsgrenze*).[17] Earnings above this limit don't require the payment of social insurance contributions and consequently, don't show in the data (e.g. the data are right-censored). Despite this shortcoming, SAPA data goes far beyond the scope and quality of most surveys in that earnings information are precise and complete with no rounding errors, underreporting or missing values as long as the earnings are relevant for the public pension scheme (e.g. all dependent employees). The earnings of individuals working in occupational groups that are excluded from the public pension scheme are not visible in

---

pension-relevant episodes in the year the sample was drawn.

17 The maximum contribution ceiling is adjusted annually (§§159, 160 SGB VI). In 2010, the maximum contribution ceiling for the statutory pension insurance amounts to 66,000 Euro for West and 55,800 Euro for East Germany (Deutsche Rentenversicherung 2009).

the data - at least for the time they worked in these occupations.[18]

Besides the earnings trajectories, data also provide information on additional pension-relevant periods, namely episodes of unemployment, sickness, child care or long-term care etc. in which case the agency responsible for transfer payments (e.g. the Federal Employment Agency) pays a reduced amount of substitute contributions to the statutory pension insurance. Some pension-relevant periods can overlap with gainful employment (e.g. child care credits), therefore yielding pension entitlements from both sources. SAPA data provides a basis for the analysis of such parallel events (e.g. child caring and employment of mothers).

Another useful SAPA feature is the so-called SES-file that provides longitudinal information on the individual's employment situation.[19] The SES-file covers periods of education, military and civil service[20], employment and unemployment, sickness and disability, child care and long-term care as well as retirement allowing for analysis of individual employment trajectories (Pollock 2007). The SES-File enables researchers to analyze the duration and sequencing of different activities over the life-course.

2.4 Limits to Generalizability: What is the Population to Generalize to?

Empirical research aims at making inference from sampling populations

[18] Individuals who are in these occupational groups throughout their working lives cannot be part of the sample population. Individuals who first worked as dependent employees and then in an occupation that is excluded from the statutory pension insurance can be part of the sample, with missing values for the latter period.

[19] Stegmann introduced the so-called SES concept. The abbreviation SES stands for social employment situation (*soziale Erwerbssituation*). The concept translates complex pension-relevant states into social situations that are easier to interpret and relevant in social science research (for more details on the SES-concept see Stegmann 2008).

[20] In this context, the term civil service refers to individuals who are carrying out civil duties in lieu of compulsory military service.

to higher-order constructs. Evidence found in samples can be extended to the total population, if the target population is adequately represented in the sampling frame.[21] It depends on the research question, whether the sample adequately represents the target population. Instead, analyses of the pension system's liabilities as retirement obligations refer to all individuals eligible for future payments, whether they live in the country or not. Analyses of inequality and the distribution of pension entitlements typically refer to the total living population in a country. This paper focuses on the latter.

In survey data, the sample population is rarely identical to the target population because certain individuals are omitted from the sampling (e.g. institutionalized population) or refuse to participate. These cases are likely to be systematically different from the total population, and their exclusion potentially biases results. In administrative data, we expect the sampling frame to be identical to the target population, because individuals cannot opt out of being part of the sample population. More so, individuals are part of the sample population by legal definition. At some point almost every person legally living in Germany will get in touch with the pension scheme, because of employment that requires the payment of social insurance contributions or for other pension-relevant reasons (e.g. child care).[22] Therefore, data provided by the statutory pension insurance seem well-suited for analyses on the material well-being of today's and future retirees living in Germany. For the total population, Kruse showed that SAPA data represent 96 percent of the German born population living in Germany (2007). Hence, generalizing to the total working age and retiree population seems to be a safe bet.

---

[21] The target population is the population we want to generalize to. The sampling frame is the operational definition of the target population. Ideally, the sampling frame is constructed in such a way that the sampling population adequately represents the target population (Singleton and Straits 2010).

[22] The unit of analysis in SAPA data is individuals with a pension account.

Typically, analyses based on SAPA data are interested in both German and foreign-born individuals living in Germany. To test whether the SAPA sampling population adequately represents the target population, we compare SAPA data to official population counts provided by the Federal Statistical Office (*Statistisches Bundesamt*) for 2007. We break down the population by age, gender, citizenship status and region.[23] *Figure 1* compares the population distributions by year of birth (1940-1992).[24] Data include both Germans and immigrants. It is necessary to stress beforehand that differences in population distributions can be attributed to imprecision in both data sources. The Federal Statistical Office extrapolates population data with the basis for today's projections being the 1987 Census in West Germany and the 1981 Census in East Germany (Statistisches Bundesamt 2008).[25] The basis for population projections dates back more than 20 years. It is therefore likely to expect systematic distortions in the projection.

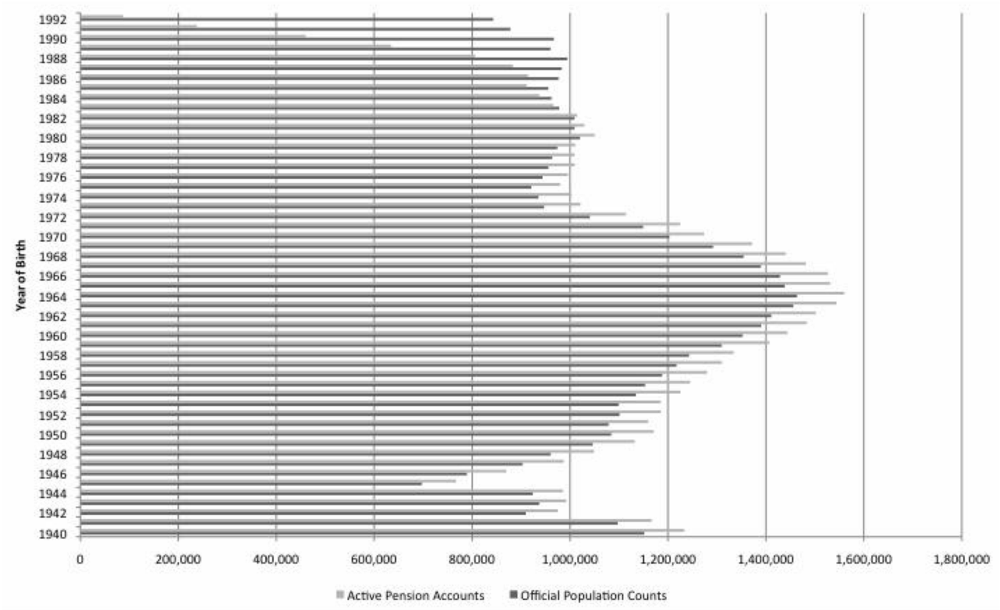
---

23 Individual-level weights apply for SAPA data (variable HRF). This paper presents only selected results. Detailed tables are available upon request from the author.

24 By definition, SAPA data only includes records of recently retired individuals younger than 68 years (Stegmann 2008).

25 To obtain reliable population numbers and to correct for bias in the population projections, the Federal Statistical Office conducts a new Census in 2011. Unlike the U.S., Germany carries out a register-based census that mainly collects data from population registers in municipalities and the Federal Employment Agency. Only a small representative share of the population participates in a survey (Statistische Ämter des Bundes und der Länder 2011).

Figure 1 Comparing Pension Records and Population Counts, Birth Cohorts 1940-1992



Note: Population counts come from the German Federal Statistical Office (Table B15/A1). Source: SAPA 2007 & Federal Statistical Office 2008; Author's calculations

The population distributions are fairly similar in both data sources, but differences are readily apparent. First, population counts (dark grey bars) exceed pension accounts (light grey bars) by far for the youngest birth cohorts (1983-1992), whereas pension accounts are overrepresented for ages 25 and higher. Overall, there is a net over count of pension accounts of almost 550,000 individuals. Separating the population by gender shows that men are overrepresented in pension accounts by 3 percent, whereas women are underrepresented by 1.3 percent.

Breaking down the population by citizenship status and region gives us a better understanding for what causes the observed patterns in the total population. Migrants who work in Germany for some years and then leave the country largely account for excess pension accounts: The number of pension accounts for the foreign-born population exceeds official population counts by 2.7 million (+45 percent) with 2 million excess accounts for men and 750,000 for women.[26] Gender differences are due to higher labor force participation of foreign-born men. These excess accounts remain in the system because pension entitlements are earned rights that individuals can claim even if they no longer live in Germany. Plus, the Statutory Pensions Insurance gets no notification about a person leaving the country. The vast majority of excess accounts are in West Germany. This regional pattern is due to the fact that more than twice as many migrants live in West Germany than in East Germany and that the overall migrant turnover rate is higher in West Germany. These excess accounts impact representativeness because they cannot be easily identified in SAPA data.[27]

In turn, for German citizens, official population statistics outnumber pension accounts by almost 2.2 million observations (-4 percent). Younger birth cohorts mostly drive this pattern, which applies equally to men and women. The reason for the under count of pension accounts for ages 15 to 25 is that having a pension account is conditional on having at least one pension-relevant episode. However, most young adults earn their first pension entitlement not until they complete vocational training or graduate from college in their early to mid-20s (Richter and Himmelreicher 2008). For birth cohorts eligible to collect old-age benefits (1940-1947), pension accounts are underrepresented by more than 200,000 observations. The SAPA sampling design explains this discrepancy: the dataset does not cover the retirement transition of a complete birth cohort (Stegmann 2008). Once retired, individuals remain in the sample popula-

---

[26] *Table A1* & *A2* in the Appendix breaks down the population distribution by birth cohort, gender and citizenship status.

[27] The state variable (*WHOT_BLAND*) indicates the state a person lives in the year 2007. The variable is missing for 14 percent of the foreign-born and less than two percent of the German-born population. No information on current residence might be a proxy for a person no longer living in Germany, but that does not explain the total number of excess accounts.

tion for a year or two and then drop out. However, the transition spreads out over several calendar years because of differences in eligibility for retirement. Hence, not all individuals are observed until age 67. A significant share drops out of the SAPA sample population shortly after retirement – typically between ages 63 and 65.[28]

Pension accounts also fall short of population counts because of certain occupational groups not being covered under the umbrella of the public pension scheme or individuals who never get in touch with the system (Kreyenfeld and Mika 2008). This line of argument applies more to West than to East Germany. First, we observe fewer pension accounts in West Germany because a larger share of the population works in jobs covered in separate occupational pension schemes. Second, women who never get in touch with the public pension scheme because they don't work in jobs subject to social insurance contributions are more common in the Western part of the country. Stay-at-home mothers are one such group that have no active pension account in case they had not worked prior to having children and don't enter the workforce thereafter.

Overall, representativeness of SAPA data is good for German birth cohorts 1982 and older except for retirees, whereas the migrant population is over represented and individuals no longer living in Germany cannot be clearly identified. The problem of excess counts causes a mismatch between the unit of analysis in SAPA data and the target population, namely the total living population in Germany. The distributional analyses focus on today's contributors living in Germany. Moreover, there are no unambiguous criteria to determine which cases to include and which to exclude in the sample population. Possibly, there are certain patterns in the

---

[28] The cross-sectional format of the data and the special feature of first-time old-age pensioners dropping out of the sample shortly after they retired make the study of the retirement transition and inter-cohort trends in retirement behavior infeasible (Fachinger and Himmelreicher 2008).

data that indicate that a person is no longer living in Germany. These patterns could help to specify the sample population in line with the population of interest.

SAPA data distinguishes between pension-relevant and non-relevant episodes. Pension-relevant are all episodes that matter for the calculation of pension benefits or the determination of eligibility. Episodes that are not relevant for pension accumulation are coded as missing. This coding implies that individuals with only one pension relevant episode but the rest of the information missing can be part of the sample population. Hence, observing the majority of the population does not imply, that the sample population is observed for their entire life cycle, but only for this part of life in which pension-relevant episodes occur. This special feature illustrates that there are two sides to completeness in SAPA data. From the perspective of the pension insurance an account is complete, if it covers all pension-relevant episodes that occur over the individual's working life, whereas to researchers accounts might appear incomplete, because they are also interested in the episodes that occur between pension-relevant information.

From the distinction between pension-relevant and missing episodes, we can derive proxies that indicate whether a person is still living in Germany or not. Among others, these patterns are long episodes of missing information in pension accounts or no earned pension entitlements in a long time. The prevalence of these specific patterns is higher in certain groups than in others. To test this assumption, the population is subdivided by gender, citizenship status and region (East vs. West Germany vs. region unknown).[29]

[29] We distinguish ten groups. There is one group that includes migrants in East and West Germany, because of the small number of observations for East German migrants.

Figure 2 compares the distribution of the share of years with missing information for men and women by region.[30] Because of the similar overall pattern for German-born individuals in East and West Germany, *Figure 2* presents the results jointly. The group of migrants is further subdivided into migrants who presumably still live in Germany and those whose current residence is unknown. This distinction is necessary, because migrants with unknown residence are more likely no longer living in Germany. However, this group of migrants does not account for all excess pension accounts. Presumably, also persons coded as living in Germany are no longer in the country.

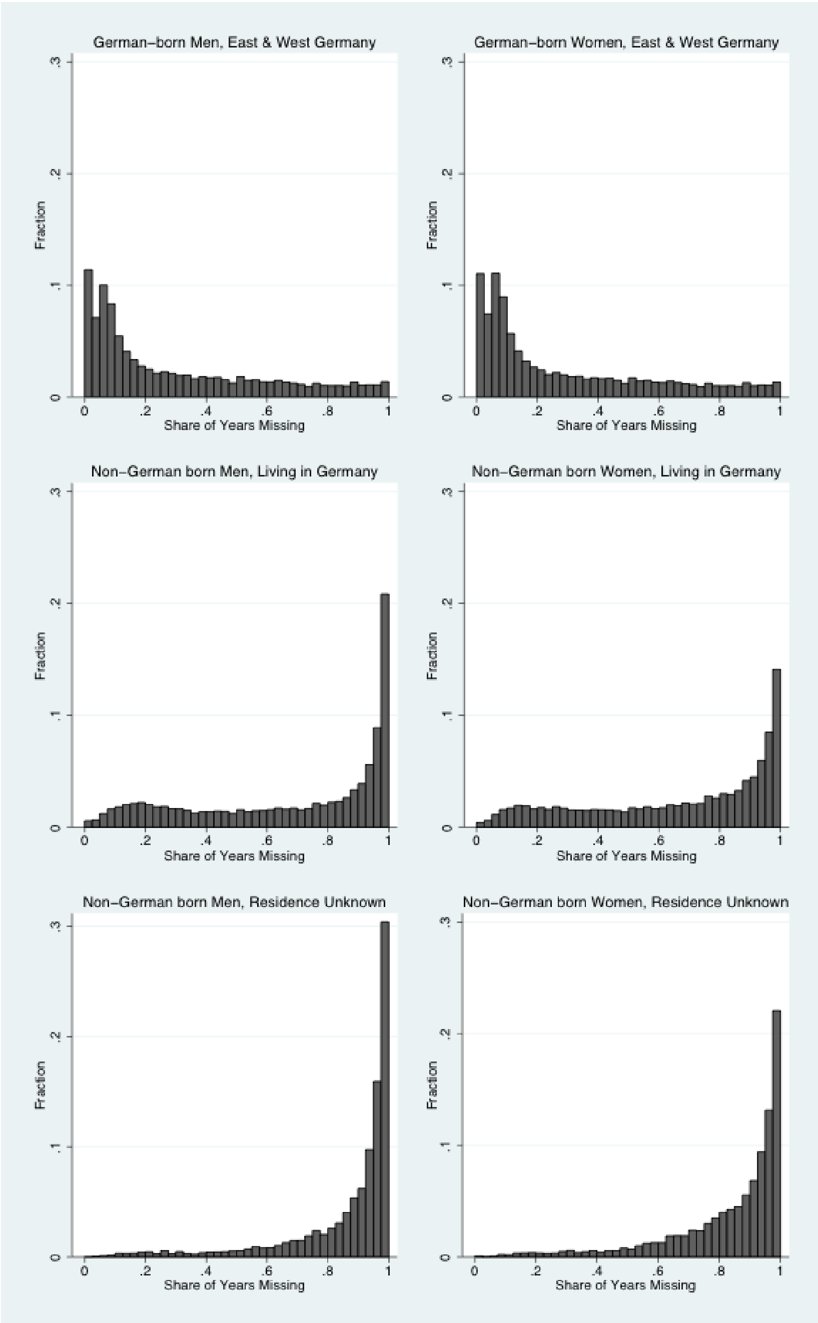
Migrants have significantly higher shares of missing years in SAPA data than the German-born population. The distributions for migrants are the exact opposite of those for East Germans. For half of male migrants, the share of missing years amounts to 0.85 (mean: 0.71), whereas the median and mean for East German men are 0.12 and 0.25, respectively. As expected, the share of missing years is even higher for migrants with unknown residence (median: 0.95; mean: 0.8) than for those who are presumably still living in Germany. Overall, patterns are similar for women. Female migrants have a much larger share of years missing than their German counterparts. Female migrants presumably living in Germany (median: 0.76; mean: 0.67) have lower shares than foreign-born women who no longer live in the country (median: 0.91; mean: 0.83), but clearly higher shares than East German women (median: 0.09; mean: 0.21).[31]

---

[30] The absolute number of years missing is not meaningful because the period of observation varies by individual depending on their year of birth.

[31] The share of years missing is higher for West than for East German women. For females from West Germany, we observe a median of 0.24 and a mean of 0.33. The higher share of missing years is a consequence of the weaker labor market attachment of women in West Germany, whereas East German women have continuous employment careers. Differences between East and West German men are less distinct, but the share of years missing is on average lower for East German men.

Figure 2 Share of Years with Missing Information, by Gender, Region and Migration Status



Source: SAPA 2007; Author's calculations

One explanation for the higher share of missing years is that migrants are older when they first get in touch with the public pension scheme. On average, Germans have their first pension-relevant episode with 18 years compared to migrants who first get in touch at age 25. Significant differences in the number of years passed since the last pension-relevant episode might also indicate that a person is no longer living in Germany. On average, more than 15.6 years (median: 13 years) passed since the last contact with the pension system for persons with unknown residence. This long period of absence applies to both Germans and migrants with no current address, with a mean of 12.6 and 17.7 years, respectively. These individuals have likely moved outside of Germany and therefore lost touch to the public pension scheme. For migrants who are presumably still living in Germany an average of 6.8 years passed since the last-pension-relevant episode (median: 0.0). In comparison, for Germans the last contact dates back an average of 2.5 years with no apparent gender differences.

We can only make educated guesses, about how to interpret these patterns of missing data, because there is no unambiguous cut-off point at which individuals are no longer considered to live in the country. Apart from individuals who left the country, another pattern to look for is persons who pay contributions in the early stages and then have missing data for the rest of their career. Chances are high that we observe a person who started off as a regular dependent employee and then left the public pension system for another occupational pension scheme (e.g. civil servants schemes) or became self-employed. Mothers who do not return to the workforce after having children have a similar pattern. They accumulate pension entitlements early on and have missing information for the rest of their working lives.

The problem of excess accounts seriously limits our ability to generalize to the total population living in Germany.[32] Drawing statistical inference

---

[32] Researchers outside the German statutory pension insurance only have access to a subsample of the data that comes in form of a Scientific Use File (SUF). The SUF restricts the sample population to the German population living in Germany excluding almost 9 percent of foreign-born individuals living in Germany. This specification of the SAPA resolves the issue of over counts; in particular, of migrants who no longer live in Germany. However, dismissing the foreign-born population living in Germany is reason for concern because this group is of special interest to policymakers.

from SAPA data implies making conclusions for the total population of pension account holders living both in and outside of Germany. Plus, all empirical evidence refers exclusively to pension-relevant episodes, whereas non-relevant episodes are coded as missing values. These problems differ from what we typically experience in survey data, with certain groups being underrepresented and information ideally missing at random. In surveys, the sampling design takes the problem of underrepresented subpopulations into account by means of disproportionately increasing the sample size of the respective groups (Singh et al. 1994). In contrast, SAPA sampling design could adjust the sampling probabilities for those subpopulations who are likely no longer living in the country. A possibly more significant threat to our ability to draw inference is potential selectivity caused by the non-random process of account validation.

2.5 Limits to Accuracy and Selectivity Concerns: The Case of Pension Account Validation

Survey data are concerned with measurement error that can occur due to inaccurate information provided by the respondent. Measurement error can also be associated to the survey instrument, the interviewer or the processing of already collected data (Singleton and Straits 2010). One motive for why researchers turn to administrative data is because of their accuracy. Accuracy is presumably high because data collection is tied to the agency's working routines and serves the primary purpose of the respective agency. The same is true for SAPA data, which serve the purpose of administering the old-age and disability pension program.

Today, employers electronically submit information on pension-relevant earnings to the statutory pension insurance. Hence, earnings information

---

Compared to German-born, the migrant population has lower pension entitlements and a higher likelihood of becoming poor in old ages.

in the individual's account is precise and up-to-date. The high precision might not apply to earnings prior to 1972, because the electronic submission of earnings data has not been introduced until then.[33] However, earnings are not the only information necessary to determine eligibility or to calculate benefits, since pension rights can also arise from periods of non-employment. To assure the correctness and completeness of data, the statutory pension insurance asks employees to validate the information recorded in their account (*Kontenklärung*).

Typically, periods of education and child care are missing in the data, since those are not submitted electronically to the statutory pension insurance. But it is also useful to reconfirm earnings data, for example in case a person hold multiple jobs prior to 1972.[34] Hence, earnings information particularly for older birth cohorts might be incomplete. Once validated, pension records provide us with reliable data, whereas non-validated pension accounts leave us with a margin of uncertainty.[35] For non-validated records, we don't know whether information is incomplete or flawed, but the likelihood of errors is higher than in validated accounts. The statutory pension insurance further distinguishes between accounts validated with and without the account holder's help. This distinction is an administrative peculiarity in the data. If an account holder does not react to the request for account validation, a public official can exercise dis-

[33] The electronic submission of earnings was introduced with the Data Collection and Transmission Act (*Datenerfassung und Datenübermittlungsverordnung*) in 1972 (Bundesgesetzblatt I vom 18. Dezember 1972 ; Bundesgesetzblatt I vom 24. November 1972). This Act was modified in 1998 (Bundesgesetzblatt I vom 10. Februar 1998).

[34] Some individuals have two pension accounts. The validation of earnings information helps to solve this issue. Overall, the problem of erroneous earnings information before 1972 is expected to be negligible.

[35] Exceptions are accounts that were validated several years ago, because the period between validation and sampling lacks proper validation. Of all validated accounts, 10 percent were validated in 2003 or earlier.

cretion and declare the account to be validated.[36]

About 46 percent of all SAPA cases are validated involving the account holder, another 24 percent were declared valid by public officials. Thirty percent of the cases are not validated at all. Since account validation is not mandatory, the process is prone to selectivity in that certain persons are more likely to validate their account than others. How do individuals who validate their account differ from those who do not?

The summary statistics compare the socio-demographic characteristics by status of account validation (see *Appendix Table A3*). Women are more likely to validate their information than men and so are Germans compared to migrants. Odds are higher for East Germans to validate their account than for their West German counterparts, whereas individuals with unknown address are more likely to have non-validated accounts. On average, individuals who confirm their information are older and the share of validated accounts increases with age. Women with validated accounts have more children than women who have none. Individuals who confirmed their account information are also more likely to be caregivers, retired and divorced. Account validation without cooperation is more likely for men and West Germans. It becomes also apparent that account validation without cooperation takes place at an earlier age than account validation with the cooperation of the account holder (average age of 41 years vs. 47 years).

Certain patterns in *Table A3* can be readily explained. Women are more likely to validate their account, because they report the birth of a child so

---

[36] Public officials declare accounts validated if a person stayed with the same employer over years and had no status changes, whatsoever. Accounts of men are much more likely to be validated without personal involvement (cp. to *Table A3* in the Appendix), because their employment trajectories are more continuous than those of women and child care credits are rarely assigned to father's accounts.

they can receive child care credits in the statutory pension insurance. These credits are typically assigned to the mother's account.[37] The age pattern is a result of administrative proceedings. Starting at age 30, the agency periodically asks employees to validate their account.[38] However, at early ages there is little reason to verify information and those who do are likely to be selective.[39] Women in their reproductive years are typically the first ones to check their accounts.[40] As individuals approach retirement, it's in their best interest to check the account for accuracy and completeness. At the latest, they need to confirm their information as they claim benefits. Therefore, almost all retirees have validated their accounts. Divorce requires immediate account validation, because pension rights accrued by both spouses during the marriage are split in half as part of the divorce settlement. Presumably, migrants who left Germany don't receive a request for account validation given the statutory pension insurance doesn't know their residential address.

Table 2 reports the results of a multinomial logit model estimating the relationship between the status of account validation (with non-validated accounts being the base category) and individual-level characteristics. *Table 2* is divided into two columns and the coefficients reflect relative risk ratios.[41] The left column holds the comparison of validated accounts with cooperation to non-validated accounts, while the right column compares

---

37 Childcare credits are not credited to the mother's account if she works as a civil servant. By law, civil servants are not entitled to child care credits in the statutory pension insurance (cp. to §56 (4) SGB VI).

38 The statutory pension insurance sends out reminders every five to six years.

39 Young individuals validate their account if they leave the public pension scheme for another occupational pension scheme (e.g. schemes for civil servants, medical doctors or lawyers). In this case, persons validate their account, to reclaim the contributions they already paid into the system.

40 Not every woman reports the birth of a child immediately. Differences in timing explain significant differences in the number of children by status of account validation.

41 Relative risk ratios equal 1 if the independent variable has no significant impact on the decision to clear the account.

validated accounts without cooperation to non-validated accounts.[42]

Overall, the model estimates show that the process of account validation is by no means a random process. The non-randomness holds for both, accounts validated with and without cooperation. The comparison of relative risk ratios in both columns indicates that the regressors predict account validation with cooperation much better than account validation without cooperation.[43] The relative impact of the regressors differs in strength by type of account validation. In line with the summary statistics discussed above, the number of children, the receipt of old-age and disability benefits as well as the categorical age dummies are strong predictors for an account validation with cooperation relative to no account validation. Each additional child increases almost doubles the odds of account validation. Being female decreases the likelihood of account validation with cooperation (-4.9 percent). This negative effect comes as no surprise, because it's not being female but having children that matters in terms of pension entitlements. The large coefficients for the receipt of old-age and disability are self-evident, because the account has to be cleared in order for retirees to collect benefits. The likelihood of account validation increases with age. This age effect holds true for both types of validation, but is even stronger for accounts validated without cooperation, with age being by far the strongest predictor in the right column of *Table 2*. Being a migrant or living in an unknown residence decreases the likelihood of account validation with and without cooperation, but much more so for the latter. The

---

[42] The estimated coefficients in a multinomial logit model compare the base category to each of the two comparison groups. Hence, the coefficients have to be interpreted relative to the base category but not between the two comparison groups. The advantage of a multinomial logit model is that estimates are based on a stable sample size. If we estimate two binary logit models instead, each logit is based on different samples making the comparison of coefficients (or relative risk ratios) - strictly speaking – impossible (Pampel 2000).

[43] Two separate logistic regression models support this finding. The author provides results upon request.

child effect also plays a role in accounts validated without cooperation. However, the effect is less strong compared to account validation with cooperation (increase by 56 percent).

Table 2 Multinomial Logit Regression – Determinants for Account Validation

Base Category (No Account Validation)	Account validation with help	Account validation without help
Female	0.965***	1.024**
	(0.00988)	(0.00982)
Number of Children	2.116***	1.505***
	(0.0227)	(0.0167)
Migrant	0.964***	0.612***
	(0.0108)	(0.00648)
Living in West Germany	0.991	1.041***
	(0.0130)	(0.0132)
Current Residence Unknown	0.789***	0.301***
	(0.0170)	(0.00724)
Care	1.878***	1.854***
	(0.0845)	(0.0819)
Divorce	2.437***	2.050***
	(0.0655)	(0.0556)
Receipt of Old-Age Pension	84.69***	0.280***
	(9.151)	(0.0523)
Receipt of Disability Pension	16.03***	0.202***
	(1.225)	(0.0227)
Age 15 to 24 (reference category)		
Age 25-34	18.85***	64.47***
	(0.717)	(3.414)
Age 35-44	58.30***	256.2***
	(2.499)	(14.44)
Age 45-54	78.97***	353.5***
	(3.980)	(21.93)
Age 55-64	208.9***	673.6***
	(12.79)	(47.76)
Years in School	1.681***	1.189***
	(0.00689)	(0.00522)
Years in Training	1.205***	0.781***
	(0.00552)	(0.00366)
Years in Employment	1.088***	1.056***
	(0.00157)	(0.00149)
Years in Unemployment	1.139***	1.183***
	(0.00389)	(0.00386)
Years in Home production	1.104***	0.977***
	(0.00419)	(0.00391)
Years Missing (no pension-relevant information)	0.952***	0.944***
	(0.00127)	(0.00121)
Constant	0.00952***	0.00874***
	(0.000376)	(0.000467)
Observations	568,586	568,586

Note: Multinomial logit models (unweighted). Odds ratios, standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Source: SAPA 2007; Author's calculations

Interesting differences appear for the duration variables that indicate how much time a person spent in education, employment or unemployment etc. Higher educated individuals (in terms of additional years in school or college) are more likely to validate their account in cooperation. The education effect is positive but much weaker for account validation without cooperation (+64 percent vs. +15 percent). An additional year in training increases the likelihood of account validation with, but decreases the likelihood of account validation without cooperation. These findings are perturbing because if less educated individuals - with likely lower incomes - are less inclined to follow the request of the statutory pension insurance for account validation they possibly forego additional benefits as they retire. Each additional year with missing information lowers the likelihood of account validation by -6.5 percent with cooperation and -9 percent without cooperation. Missing information is a synonym for having no pension-relevant events in a given year, for example if a person becomes a civil servant and is then covered by the civil servant pension scheme.

Researchers face a tradeoff between accuracy and selection using SAPA data. It appears that account validation is a source for sample selection bias. This bias can enter the analysis in two different ways: First, individuals who validate their account differ from those who don't. Second, the researcher's decision on how to specify the sample (validated vs. non-validated accounts) adds bias in much the same fashion (Heckman 1979).

The first source of sample selection bias is comparable to consent bias in survey data. Household surveys increasingly ask respondents to give consent for linking their survey information to additional data sources, mostly administrative data. Individuals who give informed are likely to differ from those individuals who do not. These differences raise concerns as to whether consent-givers are sufficiently representative of the population being studied (Jenkins et al. 2006). Equally, we expect that individuals who validate their account to be different from the rest of the population. Account validation is a selective process that is likely associated to additional unobserved variables, such as the attitude towards government or the trust in the public pension program (Schwarze et al. 2004).

Depending on the research interest, it might be useful to confine the analysis to validated accounts, thereby taking the risk to add selection to the results. For example, an analysis on childbirth and the labor market reentry of mothers should be based on validated accounts only, because the number of children is largely underestimated in not validated accounts, which seriously biases the results. The specification of the sample population is therefore a crucial decision at the beginning of each research project based on SAPA data. And this sample selection decision might compromise the external validity of results.

2.6 Limits to Measurement: Assessing the Validity and Reliability of Variables

The concepts of reliability and validity are criteria to assess the goodness of the operational definitions of variables. Reliability evaluates whether the measurement of a variable is consistent or stable between units of analysis and across time. Validity refers to the accuracy of measurement, namely whether the operational definition of a variable captures what we intend to measure. Intuitively, we expect both reliability and validity to be high in administrative data. Only few studies have challenged these assumptions so far (Fitzenberger et al. 2006; Steiner and Wagner 1998).

Limitations of variables in administrative data might be due to the fact that data were not primarily gathered for research, but for administrative purposes. This fact implies that the Social Code defines how variables are measured. The concept of validity is therefore not as relevant in the assessment of the operational definitions of variables in administrative data, because the law does not give much leeway. Of greater interest is how these definitions deviate from the concepts researchers are typically interested in.

Assessing the reliability of measurement is much more relevant when using administrative data. Two questions matter in the evaluation: First, is the variable of interest vital for the purpose of the agency? Second, is the measurement of the variable time-constant or time-dependent? Concerning the first question, variables that serve the administration of the old-age and disability scheme (e.g. determination of eligibility, calculation of benefits, etc.) are reliable. In turn, variables with no purpose in performing the core tasks of the administration lack reliability and hence, should be used with caution. The question of time-dependency in measurement relates to the issue of consistency in measurement. Changes in law are a threat to consistent measurement, because they eventually change institutional definitions and the operationalization of variables.

For example, since the introduction of the long-term care insurance in 1995 non-professional caregivers accumulate pension entitlements if they provide care to family members (Bundesgesetzblatt I vom 26. Mai 1994). As a pension-relevant form of non-employment, caregiving is included in the data. Before 1995, individuals who provided care to family members did not earn any pension rights. Hence, episodes of caregiving are not visible in the data. Comparisons of the prevalence of certain pension-relevant episodes across time or cohorts might therefore not be meaningful. Any analysis needs to take potential legal changes into account.

Table 3 provides a grid for the assessment of reliability in the measurement of variables in SAPA data. It distinguishes whether variables are relevant for the purpose of the statutory pension insurance or not, and whether the measurement is time-constant or time-varying. We can fit each SAPA variable into one of the three categories. For illustrative purposes, we only discuss the three variables listed in the *Table 3* in detail:[44]

---

[44] *Table A4* in the Appendix provides a synopsis that groups selected variables according

Table 3 Assessing the measurement validity of variables in SAPA data

	Relevant for statutory pension insurance	Not Relevant for statutory pension insurance
Measurement Concept Time-Constant	Earning points	Education
Measurement Concept Time-Varying	Unemployment	

Source: Author's illustration

2.6.1 Earning Points

Earning points are a building block for the administration of the public pension scheme and a key variable in SAPA data. Individuals start to accrue earning points through gainful employment and pension-relevant types of non-employment (e.g. unemployment, child-care, long-term care giving, etc.). The sum of earning points reflects the individual's pension entitlements and is the most important factor in the pension benefit calculation formula. Consequently, many variables in SAPA data are expressed in earning points.

In the cross-sectional part of the data, earning points matter with respect to the calculation of the pension benefit and the total assessment of contributions (*Gesamtleistungsbewertung*). The longitudinal data files provide monthly information on the accrual of earning points from employment and pension-relevant types of non-employment. The data distinguishes between earning points from contribution periods (*Beitragszeiten*) and creditable periods (*Anrechnungszeiten*).

To fully assess the reliability of the earning point information in the data, one must understand how they factor in the pension benefit calculation (Börsch-Supan 2000b; Börsch-Supan and Schnabel 1999). The old-age pension benefit $B_{i,t}$ is based on the following four factors, with each factor

to the above grid.

explained below (compare to Börsch-Supan and Wilke 2004):

$$B_{i,t} = PV_t * EP_i * AA_i$$

where,

$B_{i,t}$ - Benefits of pensioner$_i$ in year$_t$
PV_t - Pension value in year$_t$
EP_i - Sum of individual earning points of pensioner$_i$
AA_i - Actuaril adjunstment depending on retirement age of pensioner$_i$

Earning Points (EP_i): For any year, the earning points describe the earnings position of an individual relative to the average earnings of all other individuals that pay contributions into the public pension system:

$$EP^t_i = \frac{Y^t_i}{\bar{Y}^t}$$

Y stands for the i^{th} individual's earnings in a given year t. For any year t, the earning point (EP) equals 1 if the i^{th} individual earns as much as the average of contributors ($\bar{Y}^t$) in time period t. Earnings points only reflect earnings up to the maximum contribution ceiling. The total sum of earning points, where n is the number of years of employment or equivalent periods of pension credits is then used for the calculation of the final pension benefit:

$$\sum_{t-1}^{n} EP_i$$

Pension Value (PV_t): The pension value is independent of individual-level factors. It is recalculated each year with the help of the benefit indexation formula. For 2010, the pension value amounts to 27.20 Euro (West Germany) and 24.13 Euro (East Germany). The indexation has changed several times during the last years. As of 2011, the indexation of the pension value takes changes in gross earnings, investments in private pensions, and the relation between contributors and beneficiaries in the statutory pension insurance (so-called *sustainability factor*) into account (§68, SGBVI). The pension value is independent of individual-level factors and is therefore not part of SAPA data. Multiplying the pension value with the sum of earning points reflects the actual pension entitlement.

Actuarial Adjustment (AA_i): The actuarial adjustment factor reflects the individual's retirement age. If the individual retires at age 65 (official retirement age), the actuarial adjustment factor equals 1. Each month a person retires prior to the statutory retirement age, decreases the pension benefit by 0.3 percent. The maximum reduction equals 10.8 percent if a person retires three years earlier.

Adjustment Factor (AF_i): The adjustment factor is also considered in the calculation of pension benefits, however only of partial interest for the subject matter. The adjustment factor varies according to the type of pension the individual applies for and lies between 1 (old-age pension) and 0.25 (orphan pension). The adjustment factor for a survivor's pension is 0.55.

The reliability of the earning points in SAPA data is exceptionally high. There were no changes in the operationalization of the earning point concept since its introduction in 1957.[45] Throughout this period, earning points reflect the individual's relative earning position.[46] This straightforward and continuous measurement has many advantages: First, as a price-level adjusted measure it allows for comparisons between individuals and

---

[45] The earning point concept (first named *Werteinheit*) was first introduced with the pension reform of 1957 (Bundesgesetzblatt I vom 26. Februar 1957 ; Bundesgesetzblatt I vom 26. Februar 1957). The calculation of earning points was also applied to earnings prior to 1957.

[46] The only limitations are changes in the minimum and maximum contribution ceiling below and above which individuals don't have to pay social insurance contributions. Furthermore, periods of non-employment and how these periods translate into pension entitlements were acknowledged differently over time (cp. to *Table A5* in the Appendix).

across time. Second, we can express the earning point information in nominal and real earnings if multiplied with the respective average earnings of all contributors to the public pension scheme in those years.

2.6.2 Educational Attainment

In SAPA data, educational attainment is a composite measure of the highest secondary or tertiary schooling degree with information about the completion of vocational training. For each employee and each year (in our case 2007), the employer is asked to report the highest formal degree and vocational training, but not the degree required for the job. The employer sends the information to the employee's health care provider and the health care provider submits the data to the statutory pension insurance and the Federal Employment Agency. Since the introduction of the electronic submission (*DEÜV Verfahren*) of employee information in 1972, educational attainment is part of the data and its measurement has been continuous over time. But despite its continuity in measurement, the variable lacks reliability. This lack is due to the fact that educational attainment does not matter for administrative purposes of the statutory pension insurance. As a consequence, the variable has a high share of missing values.[47] *Table 4* provides the distribution of the variable educational attainment (variable *TTSC3*):[48]

[47] It is likely that higher educated individuals are more likely to have higher earnings than low educated individuals, but they are treated equal in how their earnings translate into pension entitlements.

[48] The limitations of the variable educational attainment also apply to the variables type of occupation (*TTSC1*) and occupational status (*TTSC2*).

Table 4 Distribution of Variable Educational Attainment in SAPA 2007

Value Labels for Variable Educational Attainment	n (Frequency in %)
Missing Information	182,277 (32.1)
Secondary school or higher secondary school without vocational degree	51,381 (9.0)
Secondary school or higher with vocational degree	175,951 (31.0)
High school or technical high school without vocational degree	9,861 (1.7)
High school or technical high school with vocational degree	15,278 (2.7)
Completed degree at university of applied sciences	15,155 (2.7)
Completed degree at university	23,476 (4.1)
No information available/ degree unknown	95,114 (16.7)

Source: SAPA; Author's calculations

SAPA data deviate quite clearly from the measurement of education in survey data. Typically, questions about the secondary and tertiary schooling are separate from questions about the completion of vocational training. This operational definition makes the comparability with empirical evidence from other data sources difficult. But this downside would be negligible, if the variable had not almost 50 percent missing values.

There are several explanations for the high number of missing values and these explanations illustrate why this measure is unreliable. First, information refers to the year 2007. If a person had no earnings subject to social insurance contributions in 2007, the value is missing.[49] Second, even though employers are asked to provide information on educational attainment, there is no sanction if they do not. Plus, there is no way to test, whether the information provided is reliable and whether employers fol-

[49] Educational attainment is not available for the unemployed, self-employed or individuals in atypical employment (400 Euro jobbers).

low consistent rules in reporting information on educational attainment (e.g. is the information based on certificates and transcripts or does the employer check back with the employee directly).[50]

For information on educational attainment, the IAB employment subsample (*IAB Beschäftigtenstichprobe*, henceforth IABS) relies on the same data source as SAPA data, namely the information reported by the employer (Fitzenberger et al. 2006). But unlike SAPA data, the IABS provides longitudinal information on education for each year a person was gainfully employed. Fitzenberger et al. state serious inconsistencies in the reporting of education over time, even though there is little reason to believe that there are large variations in educational attainment for a working person. The authors suggest deductive imputation procedures to reduce the number of missing values in the IABS. However, the imputation procedures cannot be applied to SAPA data, because the education variable is only available in a cross section. A use of the uncorrected variable or case-wise deletion is not recommended, because these actions potentially bias results.

2.6.3 Unemployment

SAPA data provide longitudinal information on unemployment. Starting at age 14 and up to a maximum of age 67, the data report the unemployment status in each month. For the status to show, unemployment has to be acknowledged as a pension-relevant episode in the law - either as a contributory or creditable period. For both types of acknowledgement, the person must have previously worked in a job subject to social insurance contributions and receive unemployment benefits according to the Second

[50] Employers are asked to double-check the information they submit to the branches of the social insurance system, but it is rather unlikely that they do.

and Third Book of the Social Code (SGB II and III).[51] If unemployment is a contribution period, the employment agency pays contributions to the statutory pension insurance and these contributions translate into pension entitlements for the unemployed. If unemployment is a creditable period, then the total time being unemployed counts towards the minimum qualification period (*Wartezeit*) and matter with respect to the total assessment of contributions (*Gesamtleistungsbewertung*). In this case, neither the Federal Employment Agency nor the unemployed pays contributions to the statutory pension insurance.

The accurate measurement of unemployment is relevant for the statutory pension insurance, because periods of unemployment can translate into pension entitlements. It is therefore a valid measure that shows whether a person receives unemployment compensation or unemployment assistance and whether this period of unemployment is a contributory or creditable period. But the administrative data does not measure unemployment consistently over time, because if and how unemployment matters with respect to pension entitlements, changes in the light of pension reforms. *Table 5* illustrates how the treatment of unemployment in the pension scheme changed during the last decades.

51 The population of recipients of unemployment benefits (unemployment compensation [*Arbeitslosengeld*] and unemployment assistance [*Arbeitslosenhilfe/ArbeitslosengeldII*] according to the Second and Third Book of the Social Code is not identical to the definition of unemployment in the Third Book of the Social Code. For example, a person is unemployed but loses eligibility for the receipt of benefits if the total household income is too high and income is credited against the benefit. The same is true for individuals who fail to meet the minimum qualification period for the receipt of benefits. They are no benefit recipients, but officially defined as unemployed (Dannenberg et al. 2008). For obvious reasons, the data does not consider the unemployed that are not officially registered as unemployed – the so-called *hidden labor force* (Holst 2000).

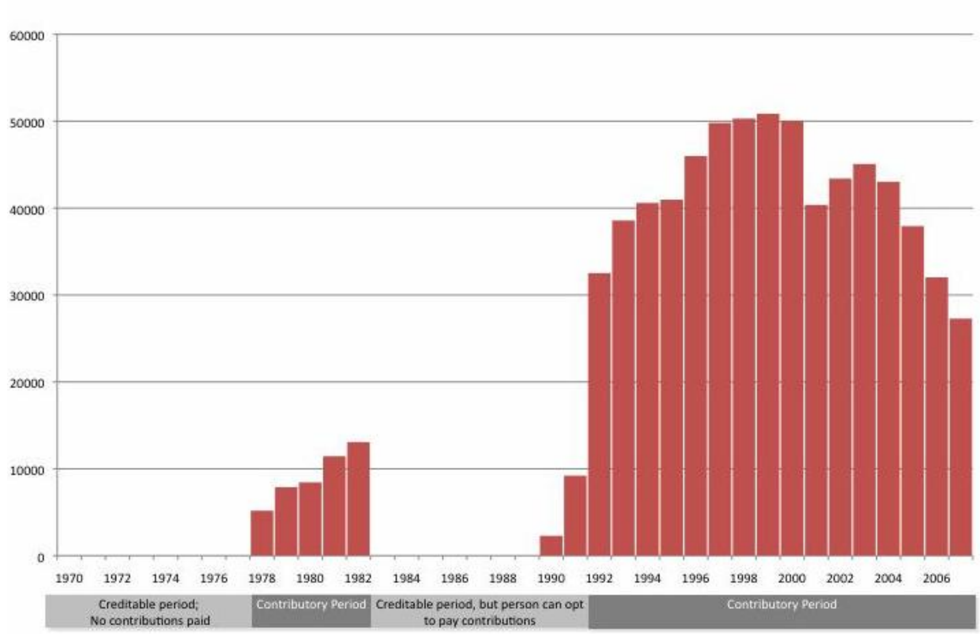
Table 5 *Time-Dependence of Unemployment Compensation in the statutory pension insurance*

Period	Acknowledgement
Current Law	Unemployment compensation (*Arbeitslosengeld*) amounts to 67 percent of the last gross wage for beneficiaries with children and 60 percent for beneficiaries with no children. For all recipients of unemployment compensation, the employment agency pays contributions to the statutory pension insurance.
Period	Treatment of Unemployment Compensation in Pension Law
1957 through 1977	Unemployment is a creditable period, if a person was unemployed for at least one month.
1978 through 1983	Unemployment is a contribution period. The basis for the amount of contributions paid by the Federal Employment Agency is 100 percent of the person's last gross wage.
1984 through 1991	Unemployment is a creditable period, but the insured person can opt to pay contributions (partly or fully) to the statutory pension insurance.
Since 1992	Unemployment is a contribution period. The basis for the amount of contributions paid by the Federal Employment Agency is 80 percent of the person's last gross wage

Source: Excerpt from *Table A5* in the Appendix; Author's illustration

The number of changes in the recognition of unemployment in the statutory pension insurance since 1957 illustrates the relevance of the legal context in the interpretation of the data. *Figure 3* plots the prevalence of the receipt of unemployment compensation between 1970 and 2007.

Figure 3 *Receipt of Unemployment Compensation in SAPA data, 1970 - 2007*



Source: SAPA 2007; Author's calculations

Figure 3 shows significant differences in the prevalence of unemployment over time. It appears as if nobody received unemployment compensation 1970 through 1977, and 1983 through 1989. However, this evidence is misleading, because unemployment compensation was only a creditable period during this time. This comparison illustrates why the interpretation of unemployment in administrative data is not straightforward. It is indispensable to take changes in the pension law into account. The data do not reflect the true number of unemployed individuals, but rather the true number of individuals with pension-relevant times of unemployment. Changes in the prevalence of unemployment cannot be interpreted in absolute terms, but have to be interpreted in the context of legal changes, because these changes largely affect the results of statistical analyses and also compromise the comparability across time or with evidence from survey data (for an interesting application based employment data see Kruppe 2009).[52]

2.7 Limits to Content: The Lack of Relevant Covariates

A thorough assessment of the power of administrative data needs to address the issue of lacking covariates in SAPA data. The data provide exhaustive information about peoples' working lives, their pension-relevant earnings as well as individual eligibility and entitlements. However, information that goes beyond the public pension scheme is missing. Most importantly, these are additional (pension) income and wealth variables –

---

52 Kruppe et al. illustrate the effect of changes in the legal definition of long-term unemployment. Until 1985, persons were considered long-term unemployed if the spell lasted for more than 12 months and the person was registered at the Federal Employment Agency even if unemployment was interrupted by short periods of employment that could last for 13 consecutive weeks. In the years following 1985, every interruption ended the unemployment spell. With every additional spell of unemployment, the count of time unemployed started all over. These changes in definition had a clear impact on unemployment numbers (Kruppe et al. 2008).

not only at the individual level but also for members of the household and needs sharing unit. These variables are essential for research on wealth and inequality. Further multivariate analyses require important explanatory variables such as educational attainment, employment status, occupation or the number of hours worked per week (Rasner et al. 2007). These variables are either not part of the data or they are not reliable. Based on SAPA data alone, it is clearly impossible to estimate the bias that results out of omitted variables.

This paper illustrates the potential bias through a comparison with data from the German Socio-Economic Panel (SOEP). The SOEP is a broad interdisciplinary household panel study that collects data on a representative random sample of the total living population in private households in Germany.[53] Today, 26 waves of data for West Germany (since 1984) and 20 waves for East Germany (since 1990) are available. Each year, the survey collects data for more than 12,000 households and 22,000 individuals interviewing all adult and youth members of the household unit.[54] The SOEP provides particularly useful reference data because it collects not only information on public pension benefits, but also on other types of retirement income. Further, a special topic module on wealth that is launched every five years (the last one in 2007) allows for detailed analyses on the distribution of income and wealth as well as inequality.

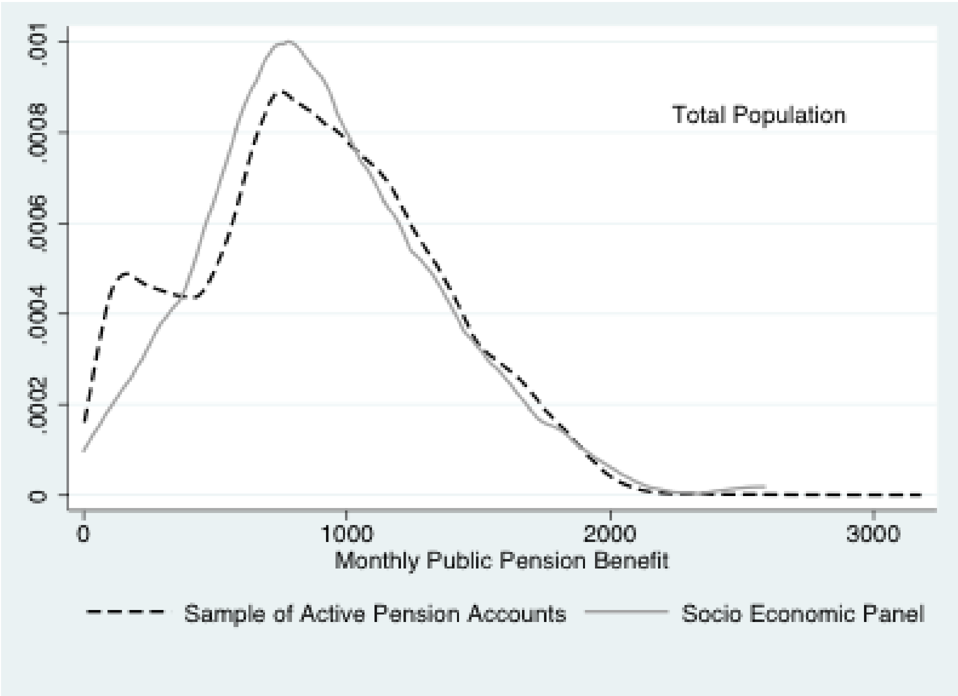
First, we take a look at how public pension benefits in SOEP data compare to SAPA data. We compare the amount of public pension benefits for recently retired individuals in SAPA and SOEP data for the year 2007. The sample population covers retirees aged 60 to 67 who report an old-age

53 This specification implies that certain segments of the population are systematically excluded from the survey, namely institutionalized population, the homeless, emigrants and potential immigrants (Wagner et al. 2007).

54 For more detailed information on the SOEP, go to http://www.diw.de/de/soep or consult further readings, e.g. (Anger et al. 2008; Wagner et al. 2008).

pension benefit. Beneficiaries of disability pensions are excluded, because they cannot be clearly identified in SOEP data. Plus, this group of pensioners significantly differs from old-age retirees.

Figure 4 Distribution of Monthly Public Pension Benefits in SOEP and SAPA data



Source: SAPA 2007 and SOEP 2007; Author's calculations

On average, monthly pension benefits amount to 860 Euro in the SOEP and 878 Euro in SAPA. *Figure 4* shows fairly similar distributions of public pension benefits in SOEP and SAPA data. This similarity is particularly striking for individuals who receive monthly pension benefits of 1,000 Euro and higher. Deviations between the two datasets are apparent for benefits lower than 1,000 Euro, in particular the hump between 100 and 200 Euro in SAPA data. How can we explain the differences in the two distributions? A look at the group-specific distributions in the Appendix (cp. to *Figure A1*) gives answers. We find the largest deviations and the hump in the distribution of West German women. West German women are most likely to receive an own independent public pension benefit and a derived survivor's benefit.[55] Possibly, women don't distinguish between the two benefits and report the sum of both in the survey, whereas survivor's benefits are by definition not part of SAPA data. The differences in the migrant population might be the result of small numbers of observations in the SOEP and selection effects. Overall, the correspondence of the distribution of monthly public pension benefits is sufficiently high.

For the study of the material well-being of the elderly population it is insufficient to look at benefits paid by public pension program. Given the paradigmatic shift in old-age provision going from a strong reliance on the public pension program to a multi-pillar pension system, it is useful to include benefits in the computations. A comprehensive analysis of the distribution of old-age income also considers derived survivor's benefits. The inclusion of survivor's benefits is especially important for women, because a large share of women has relatively small own entitlements and has to rely on either their husband's benefits or the survivor's pension in case he dies. For illustrative purposes, *Table 6* shows how the distribution of old-age income changes depending on what types of old-age income we include.

[55] Independent pension benefits are based on the individual's own accumulated entitlements as opposed to derived pension benefits, such as survivor or orphan's pensions that are based on the entitlements of a relative.

Table 6 Compositional Changes in Old-Age Income Across Different Demographic Groups

	Total (n=1322)	Men East (n=200)	Men West (n=381)	Women East (n=240)	Women West (n=410)
Benefits from statutory pension insurance Only (in €)					
Mean	894	1,128	1,224	850	592
Median	877	1,097	1,203	796	536
Std. Deviation	444	262	424	255	320
Lowest Quintile (in €) Share in Percent	505	0.0	4.0	3.5	44.3
2nd Quintile (in €) Share in Percent	769	4.3	10.6	42.1	29.5
3rd Quintile (in €) Share in Percent	1,025	28.9	18.3	30.5	16.5
4th Quintile (in €) Share in Percent	1,304	50.7	28.4	15.2	7.5
5th Quintile (in €) Share in Percent	2,772	16.1	38.7	8.7	2.2
+ Including Private and Occupational Benefit (in €)					
Mean	1,042	1,139	1,560	854	695
Median	924	1,097	1,404	801	602
Std. Deviation	713	278	901	259	446
Lowest Quintile (in €) Share in Percent	536	0.0	4.2	4.4	44.0
2nd Quintile (in €) Share in Percent	809	6.5	10.6	45.9	23.4
3rd Quintile (in €) Share in Percent	1,097	45.2	15.3	32.5	16.9
4th Quintile (in €) Share in Percent	1,454	38.2	24.3	15.4	10.6
5th Quintile (in €) Share in Percent	7,857	10.0	45.5	1.8	5.1
+ Including Survivor' s and Other Pensions (in €)					
Mean	1,113	1,165	1,569	934	839
Median	1,006	1,108	1,404	876	749
Std. Deviation	710	315	899	341	509
Lowest Quintile (in €) Share in Percent	595	0.0	4.2	14.2	37.1
2nd Quintile (in €) Share in Percent	878	10.4	14.1	38.7	23.6
3rd Quintile (in €) Share in Percent	1,169	54.4	15.9	23.6	17.4
4th Quintile (in €) Share in Percent	1,522	25.6	25.8	17.5	12.5
5th Quintile (in €) Share in Percent	7,857	9.5	40.1	6.0	9.5

Source: SOEP 2007; Author's calculations

Table 6 distinguishes three distributions of old-age income for the total population (left column) and for four demographic groups (columns to the right): West and East German men and women. These groups likely differ in the composition of their retirement income portfolio and benefit to a different extent from the inclusion of additional old-age benefit types.[56] Based on the distribution of old-age income for the total population, we determine the quintile bounds and compare how the groups are distributed across quintiles. The table consists of three panels: the upper panel considers only independent pension benefits from the statutory pension insurance, the middle panel adds old-age income from occupational and private pension schemes on top of public pension benefits, and the bottom panel shows the distribution of the total old-age provision after the inclusion of derived survivor's benefits from various schemes and special old-age benefits (such as accident benefits).

The results show that men in East and West Germany receive significantly higher public pension benefits than women.[57] Men are over-represented in the two top quintiles, whereas women are in the three bottom quintiles. East German women fare better when it comes to the level of public pension benefits than do West German women. Consequently, the gender pension gap is wider in West Germany (52 percent) than it is in East Germany (25 percent).[58] The comparison of proportions in the upper panel to the two bottom panels shows who gains and who loses from the inclusion of additional old-age income benefits. The middle panel includes benefits from occupational and private pension schemes. West German men benefit most from this inclusion. Their average benefit increases by 345 Euro. The benefits of West German women grow by roughly 100 Euro. East German men and women don't gain at all from the consideration

---

[56] Because of the small number of observations for migrants, I exclude this group from the computations.

[57] For the purpose of this paper, we only discuss the results for German men and women in detail.

[58] The gender pension gap describes the percentage distance of the average monthly pension benefit of women to those of men (Rasner 2006).

of occupational and private pensions. The coverage in these complementary pillars of old-age provision seems to be negligible, which results in a lower relative income position for men and women in East Germany. Their share in the top quintile decreases (men: 16.1 vs. 10 percent; women: 8.7 vs. 1.8 percent), whereas the opposite is true for their West German counterparts (men: 38.7 vs. 45.5 percent; women: 2.2 vs. 5.1 percent).

Women in East and West Germany gain most from the inclusion of survivor's benefits.[59] Compared to the middle panel, West German women gain an average of 144 Euro compared to East German women with a gain of 80 Euro. Especially West German women improve their relative income position, which results in a decreasing share of women in the bottom quintile (44.0 vs. 37.1 percent). For obvious reasons, men gain almost nothing from the consideration of survivor's benefits: 1) Husbands are typically older than their wives and have a lower life expectancy. Hence, only a small share of men collects survivor's benefits. 2) Their high average pension benefit counts against their survivor's pension. This deduction can result in a survivor's benefit equal to zero.[60]

Overall, this illustrative analysis stresses why the lack of relevant information in administrative data might lead to flawed conclusions. An analysis that is limited to entitlements from the statutory pension insurance falls short and identifies West German women as the main group of concern for policymakers, whereas the consideration of occupational and private pensions alludes to the problem of lacking coverage in these complementary pillars of old-age provision in East Germany. A thorough

---

[59] This impact is likely to be even higher for older cohorts. *Table 6* considers retirees aged 60 to 67. For these cohorts, the share of widows is still relatively small.

[60] In 2008, almost 5 million women and 873 000 men were eligible for survivor's benefits. Forty percent of the surviving husbands, but only one percent of surviving wives received a survivor's pension equal to zero because they had too high independent pension benefits (Deutsche Rentenversicherung 2009).

analysis of old-age poverty risks has to consider additional income and wealth components – not only at the individual but also at the household level. However, this type of information is missing completely in administrative data. It is not even possible to link the pension accounts of individuals who live in the same household.

2.8 Conclusion

Process-produced administrative data largely improve the data infrastructure in Germany. As a by-product of government agencies' working routines, they provide detailed (and cheap) information on a wide range of government programs. Administrative data supplement survey data that typically can't collect program specifics in such great detail. Since data serve the respective purpose of each agency their high quality and the lack of flaws are often taken for granted. There are several reasons for this assumption: First, the number of observations in administrative data exceeds sample sizes in survey data by far. Second, data are expected to be representative, because almost every person gets into touch with the agencies at some point over the life cycle. Third, sample selection bias seems to be of no concern, because program rules define the sample population and individuals cannot opt out. And fourth, data are presumably reliable and precise because they serve the administration of government programs.

This paper provides a systematic account of whether these assumptions about administrative data hold true. Evidence from the Sample of Active Pension Accounts provided by the German statutory pension insurance shows that analyses based on administrative data alone are limited in scope. The data are concerned with the following issues: 1) coverage errors in the sampling population that limit statistical inference; 2) non-randomness in account validation that raise the issue of sample selection bias; 3) specifics in data collection and legislative changes that compromise the validity and reliability of variables; and 4) the lack of relevant covariates that limit the explanatory power of statistical analyses or make those analyses even infeasible.

The issues sound somewhat familiar to what we experience in survey data, but the paper illustrates that these limitations play out differently in administrative data. They are a direct consequence of the legislative framework government agencies operate in. This framework specifies the process of data production (collection) and gives little leeway for changes. Hauser et al. stress that science can barely exert any influence over this process (Hauser et al. 2008). This shows why opening access to administrative data to the broader scientific community was an important first step, but also why more steps have to follow. Given that the scientific community has little influence on the process of data production, raises the need for a better cooperation with the scientific community in the stage of data preparation, which is one of the recommendations of the German Data Forum (Rat für Sozial- und Wirtschaftsdaten 2010). This cooperation is especially fruitful with institutions that collect their own data (e.g. large-scale household panel studies). Further, the analyses underline the need for proper data documentation, easy to comprehend for individuals who lack a thorough institutional and legal knowledge of the German public pension scheme. This documentation should guide researchers in questions of sample specification and the choice of variables. In particular, it should clearly state, whether variables are reliable or not.

Apart from improvements on the part of data providers, there is need for a better understanding of the advantages and limitations of these data as an essential precondition for more meaningful empirical analyses. A comparison of administrative data with reference statistics from survey data shows how differences in conceptualizations might bias statistical results. Eventually, a more thorough knowledge about the properties of administrative data could motivate methodological innovations comparable to ongoing improvements in survey research methods (e.g. imputation of key variables with missing values etc.).

Moreover, there are additional, more innovative paths to broaden the scope of administrative data. One such approach that holds large research potential is the matching of administrative with survey data (Hauser et al. 2008). Data can be matched over statistical matching or record linkage procedures. These techniques allow us to circumvent the limitations of administrative data and to enhance their power. The linkage of administrative and survey data combines the completeness and accuracy of administrative data with the scope and of population representative survey data – maximizing the strengths and minimizing the limitations associated with each of these two data sources. The next paper presents a feasibility study for a statistical matching of administrative pension records and population representative survey data.

3 Best of Both Worlds: Preparatory Steps in Matching Survey Data with Administrative Pension Records[61]

3.1 Introduction

The previous paper identified statistical matching as one option to augment the power of administrative data. This paper provides a feasibility study for the statistical matching of survey data from the Socio-Economic Panel Study (SOEP) and administrative pension record from the Scientific Use File Complete Insurance Biographies 2004 (*Vollendete Versichertenleben*, henceforth SUF VVL 2004) maintained at the statutory pension insurance. To establish a link between these two micro datasets is appealing because they are good complements. The SUF VVL 2004 provides details necessary for the calculation of the individual's pension benefit as well as complete information on monthly pensionable earnings across the individual's life cycle (Stegmann 2006a).[62] The SOEP gives household context information and other relevant components of income. A successful matching of the two datasets brings together the best of both worlds by combining their respective benefits and circumventing their drawbacks.

Several factors speak for using the SUF VVL 2004 for the feasibility study. First, it is the first cross-sectional dataset with longitudinal information the Research Data Center of the statutory pension insurance (FDZ-RV) provided to researchers. The longitudinal nature makes it compatible with the longitudinal household panel survey data in the SOEP. Second, the dataset covers a representative sample of recently retired individuals in the

---

[61] This paper is joint work with Ralf K. Himmelreicher, Markus M. Grabka, and Joachim Frick. Anika Rasner is the first author of this work.

[62] This is true for all earnings that are subject to social insurance contributions. Certain occupational groups are systematically excluded from the public pension insurance, such as farmers, civil servants, or the self-employed. The SUF VVL 2004 does not provide information about the earnings of these occupational groups.

statutory pension insurance (*gesetzliche Rentenversicherung*). This sample population can also be easily specified in the SOEP, so that we have two comparable matching populations.[63]

Statistical matching does not aim at finding the exact same person in both datasets. Finding the exact same person is impossible, because due to the measures instituted to protect the confidentiality of personal information, no common identifiers are available. Hence, the two datasets cannot be merged easily. However, through statistical matching, cases that are similar in terms of the observed characteristics of a person can be identified and linked. By combining information from different sources, one can obtain a much more comprehensive dataset for the topic of interest (van der Putten et al. 2002). Statistical matching is becoming increasingly popular in economics and social sciences. In particular, it is proving to be a useful tool in the evaluation of public policies. For example, Hujer et al. (2004) and Caliendo (2006) have applied statistical matching methods in the evaluation of the effects of job creation schemes on success in the labor market.

The dataset that would result from matching the SOEP and the SUF VVL 2004 opens a multitude of new research possibilities. First, it would allow us to simulate the old-age income of actual and future cohorts of retirees. On the basis of the available household context information, we would be able to make qualified statements about the distribution of old-age income and quantify the prevalence of old-age poverty among the population of interest. Second, we could analyze how the accumulation of pension entitlements evolves over the life cycle and how certain demographic events (e.g. divorce, birth of a child) affect the individual's ability to accumulate

---

[63] The limitations of the Sample of Active Pension Accounts identified in the previous chapters such as the issue of account validation or coverage errors in the sample population don't pertain to the Scientific Use File *Completed Insurance Biographies* because the dataset covers only retired individuals with validated accounts.

pension rights high enough to lift them above the poverty line. Further, the dataset would also be suitable to approximate the social security wealth of individuals who have not yet retired. Research that addresses the distribution of wealth and income needs to take this wealth component into account. Up to now, these accumulated pension rights have not been considered adequately in distributional analyses, even though it is essential for obtaining unbiased wealth estimates. For example, this becomes relevant when comparing the wealth of individuals who are insured in the public pension insurance scheme with the wealth of those groups who are excluded from public pension insurance (e.g. the self-employed or civil servants). Last but not least, the longitudinal dataset would allow us to evaluate behavioral effects of recent policy reforms.

This paper presents the preparatory steps for a statistical matching of administrative pension records with survey data. The work will not focus on distributional analyses and does not intend to present any results. Its sole purpose is to test the feasibility of a statistical matching of both datasets. It is structured as follows. The opening section sets out issues of confidentiality and gives a brief sketch of the German data protection law and its implications for the use of social data in general and for the statistical matching in particular. The following section describes both datasets followed by an outline on why the two data sources complement each other and pinpointing the potential pitfalls in the statistical matching (section 3.3). The next section specifies the population of interest, presents the key matching variables, and compares their respective distributions in both datasets.[64] Section 3.6 discusses the results of various regression models for different demographic groups and assesses the predictive quality of

[64] It is important to distinguish between the dataset *Completed Insurance Biographies 2004* and the *Scientific Use File Completed Insurance Biographies 2004*. The first refers to the total population of first time retirees in 2004, whereas the latter refers to a sample of the total population.

the model. The following section, presents out-of-sample predictions, which show whether the regression results estimated on the basis of one dataset can be replicated, applying the estimated coefficients to the other dataset. Section 3.9 concludes whether statistical matching is a feasible approach to complement population representative survey data with information from administrative pension records.

3.2 Issues of Data Confidentiality

Data from the statutory pension insurance are social security data that are protected by the Social Security Data Protection Act (*Sozialdatenschutz*), which is part of the Social Code (*Sozialgesetzbuch*). The Social Code establishes rules for the collection, processing, and use of highly sensitive personal and privacy data in the branches of the social insurance system, such as the statutory pension insurance (Bundesministerium für Arbeit und Soziales 2006). Some uses of the data are regarded as an infringement of the individual's personal rights, in particular, the right of informal self-determination. Laws that safeguard the use of social security data are laid out in the provisions on the confidentiality of social security data in § 35 Book I of the Social Code (SGB I), on the protection of social security data in § 67 - 85a, Chapter 2, Book X of the Social Code (SGB X) and supplementary provisions for the protection of data in other sections of the Social Code (Bundesministerium für Arbeit und Soziales 2006).

The articles of the Social Code do not apply if the data have been anonymized. The process of anonymization therefore allows the Research Data Center of the Federal German Pension Insurance to issue Scientific Use Files to researchers who are interested in the empirical analysis of retirement and disability. According to the legal definition of § 67 of SGB X, social security data are anonymized if they have been altered in such a way that the identity of individuals can only be inferred by expending an unreasonable effort in terms of time, money, and manpower. This type of anonymization is called de facto anonymization. In contrast, if it is impossible to infer the identity of the individual from the data, then we speak of absolute anonymization (Heese 2004). The high costs of absolute anonymization outweigh its benefits and furthermore, compromise the research value of the data. Anonymization is a trade-off between the risk of personal information being disclosed and the usability of data for research. De facto anonymization makes it almost impossible to re-identify individuals while still providing analytically valid micro-data to researchers (Hawala et al. 2005).

In order to analyze the factually anonymized Scientific-Use Files provided by the FDZ-RV, researchers have to sign a data use contract. The data transfer from the FDZ-RV to the researcher adheres to the principles of safe harbor. The use of the SOEP data is bound by the strict requirements in Germany for the protection of the confidentiality of data (see *Bundesdatenschutzgesetz*). In order to work with the anonymized micro-data, researchers have to sign a data transfer contract. Further technical and organizational requirements have to be met before access is granted to the data so that the data is protected from unauthorized access. These requirements involve a personal computer or a computer network that is password-protected. Furthermore, persons who work with the data are obliged to protect its confidentiality. The data transfer contract explicitly prohibits any attempt to de-anonymize the data or to re-identify individual respondents in the data.

Despite the above restrictions, the statistical matching of the two datasets, the SUF VVL 2004 and the SOEP, is allowed. According to the data transfer contract of the SOEP group and the data protection representative of the statutory pension insurance, statistical matching is allowed only if the matched datasets are both anonymized. Consequently, statistical matching is not allowed if an anonymized dataset is to be matched with non-anonymized micro-data.

3.3 The Data

3.3.1 Completed Insurance Biographies 2004 (SUF VVL 2004)

The Scientific Use File Completed Insurance Biographies 2004, provided by the Research Data Center of the Federal German Pension Insurance, is based on administrative records or pension accounts of individuals, who are entitled to receive public pension benefits.[65] It is the first longitudinal dataset that the FDZ-RV issues to researchers who are interested in retirement and disability (Stegmann 2006a).[66] The SUF VVL 2004 is a systematic random sample of individuals who received public pension benefits for the first time in 2004.[67] The Scientific Use Files (SUF) was generated in a two-stage sampling procedure. In the first step, a twenty percent sample was drawn from the pool of first-time retirees in 2004. In the second step, a subsample of twenty-five percent was drawn for selected age groups. The final data product, the SUF VVL 2004, is a five percent sample of first time pensioners that contains a total of 39,331 cases (Stegmann 2006a).[68] The sample is selective for several reasons. First, the sample comprises only persons eligible for public pension benefits. This criterion implies that certain subgroups of the population are systematically excluded, e.g. the self-employed, or civil servants in the case that they never accumulated any entitlements within the social security system.[69]

---

[65] In the remainder of the paper, we will use the abbreviation SUF VVL 2004 when speaking of the *Scientific Use File Completed Insurance Biographies 2004*.

[66] The data, as well as more detailed information, can be found at www.fdz-rv.de or in the special issue of Deutsche Rentenversicherung Volume 61, Issue 9-10, which deals exclusively with the SUF VVL 2004 and empirical applications based on the data.

[67] The sample of completed insurance biographies comprises first-time old-age pensioners as well as first-time disability pensioners. The analysis will be confined to old-age pensioners.

[68] Individuals can be part of the sample population conditional on having a validated pension account (*Kontenklärung*).

[69] Other groups are farmers, lawyers, medical doctors, and certain craftsmen, because they are covered by their respective profession-based pension scheme, such as the farmers' pension scheme.

Second, the sample considers only two benefit types, namely old age and disability pensions. Beneficiaries who receive other benefit types, such as educational pensions, or survivor's pensions (i.e. no personal pension entitlements) are not part of the sample. Third, the sample excludes persons eligible for public pension benefits in foreign countries given the respective countries have a social security agreement with Germany.

As a result of these selection criteria, the sample is representative neither of the population as a whole, nor of the group of the elderly. The lack of representativeness is due to the fact that access to public pension benefits depends on whether individuals fulfill the eligibility criteria making intercohort analysis strictly speaking, impossible (Fachinger and Himmelreicher 2008). The SUF VVL 2004 is composed of a cross-sectional and a longitudinal part. The first part contains technical variables (e.g. person ID, year of first-time receipt of pension, etc.) and demographic information (e.g. sex, year of birth, nationality, etc.), as well as aggregate data related to the calculation of the individual's public pension benefit referring to the year 2004. The longitudinal part is divided into several sub files. Ideally, the longitudinal information is available for a maximum of 624 months, starting in January in the year the person turned 14 years up to December in the year the person turned 65 years. A missing value appears in the data if a person was not employed in a job that is subject to social insurance contributions or if no other situation applied that matters in terms of pension entitlements. For our purposes, the longitudinal files on the individual's earning points and the social employment situation (*SES*- file) are most relevant. The social employment situation gives information about the individual's employment trajectory. The *SES*-concept follows a broader definition of employment (Stegmann 2006b). It does not only refer to employment subject to social insurance contributions, but also to other pension-relevant states – relevant in that they translate into pension payments as the person retires.[70]

3.3.2 The Socio-Economic Panel

The Socio-Economic Panel (SOEP) is a household panel study that started in 1984. The SOEP is a broad interdisciplinary survey that covers a representative sample of the total population living in private households in Germany.[71] To date, 26 waves of data for West Germany and 20 waves for East Germany are available. The most recent accessible data was collected in 2010. For this feasibility study, we use data from 2005 (*wave V*).[72]

The micro-data provide information on individuals, households and families, and enable researchers to measure stability and change in living conditions over time. The survey measures a broad variety of objective indicators that cover such topics as demography and population, labor market and occupation or income, taxes and social security. It also contains a large choice of subjective indicators that aim at investigating the individual's perceptions, tastes and preferences, as well as (in more recent years) cognitive abilities and personality traits. The standard components are surveyed year by year, whereas certain special topic modules (e.g. Social Security and Poverty in 2002 or Use of Time and Preferences in 2005) are asked every few years. The richness of the data and continuous extensions attract researchers from various academic disciplines, for example, economics, sociology, statistics, demography, psychology, and geography.

Ideally, information is collected by asking (i) every person in the household above age 16 to complete an individual questionnaire, and (ii) one

---

70 Section 5.2 describes the content of the *SES* file in greater detail.

71 This implies that certain segments of the population, which may be relevant for the analysis at hand, are at least partly excluded from the survey; namely, the institutionalized population, the homeless, emigrants, and potential immigrants (Wagner et al. 2006).

72 For more details on SOEP data see Haisken-DeNew and Frick (2005).

person, typically the head of the household, to complete a household questionnaire. Most relevant for our purpose is the biographical information surveyed, which contains the individual's complete employment history, starting at age 15. The information in the *PBIOSPE* file is gathered through a special biographical questionnaire that is administered only once, in order to obtain information for the time prior to the first interview. The *PBIOSPE* file stores information about the employment history, categorized into different types of activities. The biographical data are then updated year by year on the basis of the ongoing survey. The annual individual questionnaire collects information about the person's occupational status in the previous calendar year and is then aggregated into annual values (Frick and Lohmann 2010). The major advantage of the SOEP data is that all income components, apart from the individual's pension entitlements, are collected in order to obtain a comprehensive measure of the economic well-being of the household. Plus, the SOEP provides extensive information about the household context and changes in its composition.

3.3.3 Perfect Complements: The Best of Both Worlds?

We want to develop a statistical matching procedure in order to obtain a dataset that combines the best of survey and administrative data. The two datasets complement each other perfectly, for several reasons. As outlined above, the dataset SUF VVL 2004 provides high-quality work histories with information about monthly earnings and the employment situation, as well as reliable data for the calculation of the individual's monthly pension benefit. However, other important covariates are missing. First and foremost, the data lacks information about the household context, as well as benefits and transfers from other pension schemes. This information is necessary for investigating issues related to inequality or the distribution of old-age income. Without additional information about income, definite statements about the development of old-age poverty are highly speculative, if not impossible.[73] Statistical matching with the SOEP will enable us to address this shortcoming of the SUF VVL 2004; namely, the lack of contextual information. The SOEP provides very detailed information about income, not only for the individual respondent, but also for the household in which the person lives. The income information ranges from wage and salary income, and private and government transfers, to asset income (Frick et al. 2010a). The data also provides comprehensive demographic information about the birth of children, marital status, and changes in status over the entire life span.[74] The information on marital trajectories in SOEP data is far more detailed than the marital status variable in the administrative pension data. The SUF VVL 2004 distinguishes only married and not married individuals.[75] The information refers to the point in time, when the individual retires.[76] In contrast, the SOEP data measures five different marital status categories (single, married, widowed, divorced, no longer married), which are surveyed year by year.

One shortcoming of the SOEP data is the lack of earnings information for the years prior to the first interview. The SOEP surveys the respondent's occupational status retrospectively, but not the individual's earnings history. This approach reduces the response burden, but it is also motivated by the lack of reliability and accuracy of earnings information that is collected retrospectively (Ferber and Birnbaum 1979). If the SOEP and SUF

---

[73] The old-age poverty rate of women would be highly overestimated if we did not consider additional income information. In the majority of cases, it is the public pension benefit of the husband that lifts women above the poverty threshold or, in the case that the husband dies, the survivor's benefit (Deutsche Rentenversicherung 2006a; Hagen et al. 2007).

[74] The information is available in the so-called *BIOMARSY* file.

[75] The category *married* includes married and remarried persons. The category *not married* covers widowed, divorced, and never married persons.

[76] Obviously, changes in marital status over the life course will explain much more variation in public pension benefits than the marital status at the point of retirement.

VVL 2004 data are matched statistically, this shortcoming can be circumvented, at least with respect to earnings that are subject to social insurance contributions, which are available over the entire life cycle. However, no life cycle information is available for other components of income. Therefore, the statistical matching will also enable earnings information to be taken into account, thus yielding a more comprehensive measure of social security wealth as a share of total household wealth.

3.3.4 Potential Pitfalls: When Worlds Collide

Despite the fact that the two datasets complement each other, there are certain pitfalls that need to be taken into consideration in both the preparation and implementation of the matching procedure. Three major pitfalls have been identified: (i) population sample versus inflow sample, (ii) differences in sampling probabilities, and (iii) differences in sample sizes. We now take on each of these potential pitfalls in the above order:

(i) Population Sample vs. Inflow Sample: The SOEP is a population-representative sample of the total living population in German households. Hence, it is possible to generalize from the sample data to the total population. However, we cannot use the entire sample population, because in this analysis we are interested in first-time pensioners only. Therefore, the sample population must be reduced considerably in order to specify the population of interest. Yet this reduced sample still needs to be large enough to allow us to draw inference from this small sample population to the total population first-time retirees in 2004.

The SUF VVL 2004, on the other hand, is a so-called inflow sample (Fitzenberger and Speckesser 2005). We use the inflows into retirement in the year 2004, more specifically inflows into old-age pensions. Being part of the sample is therefore conditional on the first-time receipt of old-age pension benefits. This sample specification entails that a person must have accumulated some sort of pension entitlements throughout his/her working life. Certain segments of the population can, by definition, never be part of the SUF VVL 2004 sample population.

These differences between a population sample and an inflow sample need to be considered when specifying the sample population. Persons who are part of the SOEP sample population may not be part of the SUF VVL 2004 sample population. The correct specification of the population needs to yield two sample populations that resemble each other in the key dimensions (Section 3.4).

(ii) Differences in Sampling Populations: In a representative sample, the probability for each person to be part of the sampling population is principally the same. In the SOEP, this is only true in theory. There are two reasons for this: First, the institutionalized population was not representatively included in the first wave.[77] Second, certain groups are oversampled deliberately. Oversampling means that the sampling probability for some groups is higher than for others. The purpose of oversampling is to obtain high enough numbers of observations for the analysis of certain subgroups in the total population. For example, East Germans and immigrants have a higher sampling probability than West Germans.[78] Hence, in the SOEP, not every person in the total population has the same sampling probability.[79]

In the SUF VVL 2004, being part of the sample is conditional on the first-time receipt of public pension benefits. As noted above, this specification

---

[77] However, persons of the initial sample population who lived permanently or temporarily in institutions were followed in later waves (Haisken-DeNew and Frick 2005). Individuals who moved from private households to institutional housing will be followed. Nevertheless, the SOEP does not aim at being representative for this population.

[78] The sampling probability for East Germans is 0.0004 and for foreigners it is 0.0008, compared to a sampling probability of 0.0002 for West Germans (Haisken-DeNew and Frick 2005).

[79] However, these differences are corrected for by appropriate weighting factors that explicitly control for the underlying differences in sample design.

implies that certain segments of the population are systematically excluded from the VVL 2004 sample population. If the condition of first-time benefit receipt holds true, the sampling probability is the same for every person. The effects of oversampling and different sampling probabilities in the SOEP for the statistical matching with the SUF VVL 2004 are further illustrated in Section 3.6. We show how the sampling probabilities contribute to differences in the distribution of certain core variables. These differences need to be taken into account when thinking about the appropriate matching technique, for example by the application of analytic weights.

(iii) Different Sample Sizes: Differences in sample sizes come into play when comparing the distribution of certain variables in both datasets. If sample sizes are small, the distribution is much more susceptible to outliers, which in turn impairs the comparability of the two datasets. Section 3.7 illustrates the outlier problem when comparing the variable monthly public pension benefit in the two datasets. The differences in sample size will be addressed in the implementation of the matching procedure, but not in this paper.

3.4 Specification of the Sample Population

For the matching procedure to be successful, the two sample population must be specified correctly. This specification involves a thorough understanding of the structure of the two samples, summary statistics and the distribution of certain core variables (e.g. gender, age, marital status, etc.). A statistical matching requires two populations that resemble each other as closely as possible in relevant ways. Otherwise, unequal populations will be matched to each other, which will impair the reliability of the results.

First, it is necessary to identify the population of interest in both datasets. In our case, the population of interest is first-time old-age retirees. It is much easier to identify this population in the SUF VVL 2004 because the dataset only consists of individuals who retired in 2004. However, in the SOEP, we have to isolate those individuals who retired recently and identify recipients of old-age public pension benefits, which is slightly more complicated. It is important that the pension rules concerning the age at retirement were constant for all individuals in the sample populations. Although plenty of social security reforms were passed between 2000 and 2005, they were directed principally towards future cohorts and only partially affect the public pension benefits and retirement behavior of this recent cohort of retirees. Hence, pension rules may be considered to be constant. In the following sections, we explain in detail how the two sample populations were identified.

3.4.1 Specifying the Analysis Population in the SOEP

Despite a relatively large total sample size of more than 10,000 households and almost 24,000 individual respondents in 2005, the sample population has to be specified in accordance with the respective research question. In our analysis, we focus on the financial well-being of first-time retirees. Therefore, the analysis was confined to a very small segment of the total SOEP population.

In the first step of the analysis, we make no use of the SOEP's panel structure. We base the analysis solely on cross-sectional data of the year 2005 (*wave V*), which comprises 21,097 cases. We use data for 2005 instead of for 2004 because the key information is collected retrospectively, and the majority of questions in the 2005 questionnaire, especially those related to the income situation, refer to the year 2004. The target variable is the individual's independent pension benefit. The interviewer asks the respondent for the benefit type and benefit amount collected from different pension schemes (variables *vp10301-vp13020*) in 2004. The question distinguishes independent benefits from own entitlements and derived benefits, such as survivor or orphan's pensions. Based on this question, we also distinguish retirees (receiving benefits from the statutory pension insurance) from non-retirees. Every person who reports a monthly public pension benefit from the public pension insurance is coded as a *retiree in 2005*. Using this criterion, we identify a total of 4,518 persons who receive public pension benefits in 2004.

Since the population of interest is first-time old-age pensioners, the population has to be further specified. Persons who received public pension benefits, but who were below age 60 in the year 2005, were coded as disability pensioners.[80] The group of disability pensioners cannot be identified over a specific variable in the SOEP questionnaire. Therefore, we had to work around this difficulty by using plausible assumptions. Current pension rules do not allow for the receipt of old-age pension benefits prior to age 60. Hence, by definition, any public pension benefits paid before age 60 presumably are disability benefits. Using the *PBIOSPE* data as a basis, we identified those individuals who retired between 2000 and 2004.[81] If a person who received public pension benefits reported that he or she had retired (*spelltype* equals 8) and that this period started later than 1999 (*beginy* > 1999), the person was coded as first-time old-age pensioner between 2000 and 2004.[82] Altogether, 949 persons belong to the population of interest.

---

80 A total of 447 persons received public pension benefits and retired prior to age 60 and were therefore coded as disability/invalidity pensioners in the data. Due to the young age of some respondents coded as disability pensioners, we assume that some might also have received orphan's pensions; however, this is very difficult to ascertain.

81 It is impossible to base the analysis solely on first-time retirees in 2004 because of small case numbers. We therefore prolong the time frame and consider first-time pensioners who retired in the years from 2000 to 2004.

82 Additional plausibility checks have shown that some respondents, who reported being retired, did not report any public pension benefits. We double-checked whether these people receive public pension benefits from other pension schemes. If this was not the case, the individuals were excluded from the population of first-time old-age pensioners from 2000 to 2004.

3.4.2 Specifying the Analysis Population in the SUF VVL 2004

The specification of the population of interest for the VVL data is less complicated than for the SOEP data. The original dataset consists of 39,331 cases. From the outset, retirees who also accumulated pension rights in other countries than Germany (so-called *Vertragsrentner*) were excluded from the Scientific Use File VVL 2004 (Stegmann 2006a).[83] The SUF VVL 2004 distinguishes only two types of public pension benefit: old-age pensions and disability pensions. Given that old-age pensioners are the focus of our research question, we exclude all recipients of disability pensions.[84] We consider the following benefit types of old-age pensions in the analysis: the regular old-age pension, old-age pensions due to unemployment or partial employment in old age, old-age pension for women, old-age pension for persons with disabilities, and the old-age public pension benefit for persons with long insurance periods.[85] A total of 7,730 persons receive other public pension benefits and were therefore excluded from the sample.

Furthermore, we excluded retirees who receive German public pension benefits while living in a foreign country. This group has to be left out

---

83 Persons qualify for the payment of a so-called *Vertragsrente* if the two countries the person worked in have a bilateral social security agreement, also called a *totalization agreement*. A *totalization agreement* governs the payment of benefits between the two countries (Social Security Administration 2007). The monthly public pension benefits of *Vertragsrentner* depend on the rules of the totalization agreement and therefore need to be interpreted in the light of these rules. For *Vertragsrentner,* a straightforward interpretation of the impact of the employment history on the level of public pension benefits is no longer possible. Plus, these persons cannot be identified in the SOEP.

84 Old-age pensioners were identified over the variable *leat*, which classifies the individuals according to the type of public pension benefit they receive.

85 Originally, these public pension benefits differed in terms of the eligibility criteria and the retirement age. The eligibility criteria (e.g. statutory retirement age & earliest possible age limit for the receipt of public pension benefits) were harmonized in the course of several reforms. For all benefit types, except for the old-age pension for person with disabilities, the statutory retirement age was raised to 65. Early retirement is penalized by permanent benefit reductions.

from the VVL because they are by definition not part of the SOEP sample. In the SOEP, a person drops out of the sample if he or she is no longer living in Germany. The same applies to persons who fall under the regulations of the Foreign Pension Law (*Fremdrentengesetz*).[86] A total of 446 persons fall under the regulations of the Foreign Pension Law. It was necessary to exclude this group of individuals, because we do not have any information about labor supply before they came to Germany. If these persons have been employed abroad, the SUF VVL 2004 data will not contain information about these periods, whereas SOEP data has information about these periods. Due to this discrepancy in the two datasets, the group was left out of the analysis.

In addition, we excluded partially retired individuals (n=67). In the SOEP, we specified the population on the basis of whether a person reports being retired in a given year and receives a monthly public pension benefit. If both conditions apply, the person is coded as retired. It is not possible to control whether a person receives only partial public pension benefits. Therefore, we excluded the group of partial social security recipients from the SUF VVL 2004. After the specification, the total sample population covers 30,829 individuals.

3.5 Finding Matching Variables

For the statistical matching procedure to be successful, the datasets need to share a set of common variables measured in comparable ways. It is useful to choose the set of common variables on the basis of theoretical consider-

[86] The Foreign Pension Law was enacted in 1960. Public pension benefits were paid to individuals of German ancestry who lived in areas outside of Germany and were forced to flee their homelands due to adverse political conditions. For individuals who fall under the regulations of the Foreign Pension Law, public pension entitlements earned in Eastern Europe are taken into account when calculating the German public pension benefit (Himmelreicher 2005).

ations and the research question that is addressed. In our analysis, we focus on the impact of the individual's employment history on the level of public pension benefits. In the following paragraphs, we identify potential matching variables in both datasets, contrast their measurement concepts and compare the summary statistics.

3.5.1 Monthly Public Pension Benefit

The target variable for the statistical matching is the monthly pension benefit individuals collect from the statutory pension insurance. The SOEP collects information on the monthly public pension benefit by directly asking respondents which agency pays their pension and how high the monthly payments were they received in 2004. Respondents are expected to report their gross pension benefit paid by the statutory pension insurance each months. For the preparatory steps of the statistical matching we use the value reported by the respondents.

It is important to be aware of one important detail pertaining to the interplay of public pension benefits and health insurance contributions and how the interplay affects the accuracy of our dependent variable. Depending on the individual's earnings before retirement, the recipients of public pension benefit can either be insured in the statutory health insurance or hold a private health insurance plan.[87] The type of health insurance coverage determines the amount of the monthly public pension benefit payment. Health insurance contributions of persons covered by the statutory health insurance are deducted from the public pension benefit before it is paid out to the beneficiary. In contrast, persons covered by private health insurance or those who are insured voluntarily in the statutory health in-

[87] Persons with earnings below the maximum contribution ceiling are automatically insured in the compulsory health insurance scheme, whereas persons with earnings above this margin can opt for a private health care provider.

surance receive a higher social security payment, but have to pay their health care premiums out of the effective social security payment.

For illustration, let us assume that a person covered by the statutory health insurance has the same gross public pension benefit of 980 Euro as a person who is privately or voluntarily insured. For the person covered in the statutory health care scheme, the statutory pension insurance directly pays one half of the health and long-term care contributions from the total gross public pension benefit of 980 Euro into the statutory health insurance. This direct deduction reduces the amount paid out to the beneficiary to 955 Euro. For a person covered by a private health insurance carrier, the statutory pension insurance does not pay the health care and long-term care contributions directly to the provider but pays out the full amount to the beneficiary, who then has to pay the health care premium. In this case, the amount paid out to the individual is higher than the gross pension benefit; namely, 1120 Euro (Deutsche Rentenversicherung 2007b).

We assume that respondents in the SOEP tend to report the public pension benefit that is transferred to their account every month. Even though respondents are explicitly asked to report the gross public pension benefit, it needs to be questioned whether respondents distinguish between their gross and net public pension benefit in the interview situation. For income from the statutory pension insurance, the comparison of income aggregates in the SOEP with official statistics shows that respondents in the SOEP tend to report a slightly higher public pension benefit, relative to the benefit they actually receive according to the official statistics (Grabka 2004).

Table 7 presents the summary statistics for the target variable, namely the monthly public pension benefit for the population of first-time retirees between 2000 and 2004 in the SOEP.[88] *Table 7* distinguishes four demographic groups: men and women in East and West Germany.

Table 7 Average Public Pension Benefits for First-Time Pensioners, 2000 and 2004

	West	East
Men	Mean: 1,268 Euro Standard Deviation: 487 Median: 1,300 Euro n=304	Mean: 1,048 Euro Standard Deviation: 267 Median: 1,000 Euro n=139
Women	Mean: 537 Euro Standard Deviation: 366 Median: 429 Euro n= 358	Mean: 732 Euro Standard Deviation: 306 Median: 687 Euro n=148

Source: SOEP 2005; Author's calculation

The results in the above table are in line with our expectations: West German men receive the highest average public pension benefit (1,268 Euro) followed by East German men (1,048 Euro). East German women have a considerably higher average pension (732 Euro) than West German women, whose average public pension benefit is 537 Euro. *Table 8* provides results for first-time retirees from 2003 to 2004, a subsample that comes even closer to that of the SUF VVL 2004. We identified a total of 351 first-time pensioners from 2003 to 2004 in the SOEP.

Table 8 Average Public Pension Benefits for First-Time Pensioners, 2003 and 2004

	West	East
Men	Mean: 1,290 Euro Standard Deviation: 518 Median: 1,280 Euro n=102	Mean: 1,013 Euro Standard Deviation: 237 Median: 990 Euro n=55
Women	Mean: 567 Euro Standard Deviation: 397 Median: 469 Euro n= 134	Mean: 603 Euro Standard Deviation: 211 Median: 600 Euro n=60

Source: SOEP 2005; Author's calculations

88 The variable *monthly public pension benefit* was topcoded at 2,500 Euro, because the SOEP data contains some implausible cases. Section 3.6.1 explains the reason for the topcoding in more detail.

Table 8 shows that the average public pension benefits have fallen for East German men (minus 35 Euro) and even more so for East German women (minus 129 Euro), whereas they have increased slightly for West German men (plus 22 Euro) and women (plus 29 Euro). However, the apparent changes in average monthly public pension benefits for the group of first-time pensioners between 2000 and 2004 compared to the group of first-time pensioners between 2003 and 2004 might be an indication of the negative impact of longer periods of unemployment as a result of the troubled economic situation in East Germany in the early 20th century. Furthermore, *Table 8* illustrates that the number of cases is quite small for this sample specification. For these two reasons, we choose the group of first-time pensioners between 2000 and 2004 to be a sample population of reasonable size.

The SUF VVL 2004 lacks explicit information about the individual's public pension benefit. However, the data includes all variables necessary for the calculation of the public pension benefit. The data only contains information for the independent public pension benefits, which are benefits based on the individual's own entitlements as opposed to derived pension benefits, such as survivor's or orphan's pensions. The SUF VVL 2004 does not include explicit information about the individual's public pension benefit, because this information is a potential source for the re-identification of persons in the sample.[89] The calculation of the benefit is based on the variable *sum of individual earning points* (*PSEGPT90*). Roughly speaking, this variable summarizes all full contribution periods, reduced contribution periods, and non-contributory periods (Himmelreicher and

[89] The decision to exclude the variable *individual's monthly public pension benefit* is worth reconsidering, because it is *the* variable of interest in the data for most of the researchers. For the matching, the variable is particularly useful because it plays such a central role in the matching procedure. According to information from the *Research Data Center*, the variable will be included in future Scientific Use Files.

Mai 2006).[90] In addition to these contribution periods, the variable *sum of individual earning points* takes the pension type factor and the actuarial adjustment in case of early or late retirement into account.[91]

Despite the consideration of the pension type factor and the actuarial adjustment, it is not possible to derive the individual's monthly public pension benefit directly from the sum of individual earning points. Due to the different actual pension values in East and West Germany, it is necessary to consider the share of earning points that a person accumulated in East and West Germany, respectively. For 2004, the actual pension value for West Germany amounted to 26.13 Euro and for East Germany to 22.97 Euro (Deutsche Rentenversicherung 2005). In the SUF VVL 2004, it is possible to adjust for the share of earning points accumulated in each East and West Germany using the variable *anteilos*, which describes the share of earning points accumulated in East Germany. *Table 9* illustrates the calculation of the individual's monthly pension benefit in the SUF VVL 2004 data:

---

90 Additional components go into the variable *sum of earning points*. However, their relative importance is negligible (Himmelreicher and Mai 2006).

91 The pension type factor varies with the type of pension a person receives and lies between 1 (for old-age pensions) and 0.25 (for an orphan's pension). Given that our analysis is bound to old-age pensioners, the pension type factor equals 1 for the entire sample population (Börsch-Supan and Wilke 2004). In contrast, the actuarial adjustment factor varies from person to person depending on the respective retirement age. If the person retires at the statutory retirement age, the factor equals 1. In the case of early retirement, the factor is reduced by 0.3% per month up to a maximum of 18% (Börsch-Supan 2000a). Late retirement increases the adjustment factor accordingly.

Table 9 Calculation of Individuals' Public Pension Benefit in the SUF VVL 2004

Pension_{EAST} = PSEGPT90 * ANTEILOS * Pension Value $_{EAST}$

+

Pension_{WEST} = *PSEGPT90* * (1 – *ANTEILOS*) * Pension Value $_{WEST}$

=

Monthly Public Pension Benefit

, where

PSEGPT90 = sum of individual earning points
ANTEILOS = share of earning points accumulated in East Germany
(1 – ANTEILOS) = share of earning points accumulated in West Germany
Pension Value $_{EAST}$ = 22.97 Euro in the year 2004 for East Germany
Pension Value $_{WEST}$ = 26.13 Euro in the year 2004 for West Germany

Source: Author's illustration

Table 10 provides the summary statistics for the monthly public pension benefit in the SUF VVL 2004 with the number of observations being significantly higher than in the SOEP.

Table 10 Average Public Pension Benefits for First-Time Pensioners in SUF VVL 2004

	West	East
Men	Mean: 1,064 Euro Standard Deviation: 498 Median: 1,136 Euro n=10,463	Mean: 1,000 Euro Standard Deviation: 307 Median: 966 Euro n=3,520
Women	Mean: 474 Euro Standard Deviation: 331 Median: 384 Euro n= 13,193	Mean: 723 Euro Standard Deviation: 276 Median: 689 Euro n=3,653

Source: FDZ-RV - SUFVVL2004, Author's calculations

The comparison of the summary statistics for East and West German men and women in the SOEP and VVL data shows that the distribution of public pension benefits is quite similar in both datasets, with the exception of West German men. Furthermore, it is noticeable that for all four demographic groups, the average public pension benefit in the SOEP is higher than in the SUF VVL 2004. Potential explanations for this pattern might be the result of over reporting of the retirement income by SOEP respondents or rounding effects. Hence, earnings tend to cluster at 50 Euro or 100 Euro steps. The over-reporting in survey data is systematic in such a way that respondents tend to report earnings of either 1,500 Euro or 1,450 Euro, rather than earnings of 1,435 Euro, whereas administrative data supposedly provides exact data (Hanisch 2005).[92]

For East German men and women, the fit between SOEP and VVL data is exceptionally good. In the SUF VVL 2004, East German men receive an average public pension benefit of 1,000 Euro compared to 1,048 Euro in the SOEP (Δ = 48 Euro). For East German women, the fit is even better. In the SUF VVL 2004, East German women receive an average public pension benefit of 723 Euro compared to 732 Euro in the SOEP (Δ = 9 Euro). The standard deviation for East German women confirms the similarity of the distribution of public pension benefits (SUF VVL 2004: 277; SOEP: 306). The results for West German women also lie within a tolerable margin. In the SUF VVL 2004, West German women receive an average public pension benefit of 474 Euro compared to 537 Euro in the SOEP (Δ = 64 Euro).

The largest discrepancy between the two datasets is found for the group of West German men. In the SUF VVL 2004, West German men receive an average public pension benefit of 1,064 Euro, whereas in the SOEP they receive an average benefit of 1,268 Euro (Δ = 205 Euro). One explanation for the large discrepancy might be that West German men are a very heterogeneous group (standard deviation of 487). Compared to the other groups, they are much more often self-employed or work as civil servants. Hence, they receive public pension benefits from different pension schemes (e.g. private or civil servant pensions). It is therefore possible that men simply report their total retirement income when they are asked to state their social security benefit from the statutory pension insurance. We

92 Administrative data is generally expected to represent the truth, whereas survey data is assumed to be prone to over- or underreporting (Kapteyn and Ypma 2007). However, the author's show in their comparison of administrative data and survey data that measurement error is also an issue in administrative data.

address this problem in Section 3.6.2.

3.5.2 Time Spent in Different Pension-Relevant States

3.5.2.1 Data Preparation

The focus of the analysis is on the effects employment histories have on the level of old-age income. In the SOEP, the file *PBIOSPE* covers information on employment histories. For the statistical matching, we need aggregate information indicating how much time a person spent in a certain activity. *PBIOSPE* distinguishes nine types of employment/activities listed in *Table 11*, plus the category missing if none of the nine types of employment applies:[93]

Table 11 Activities Distinguished in the SOEP Data

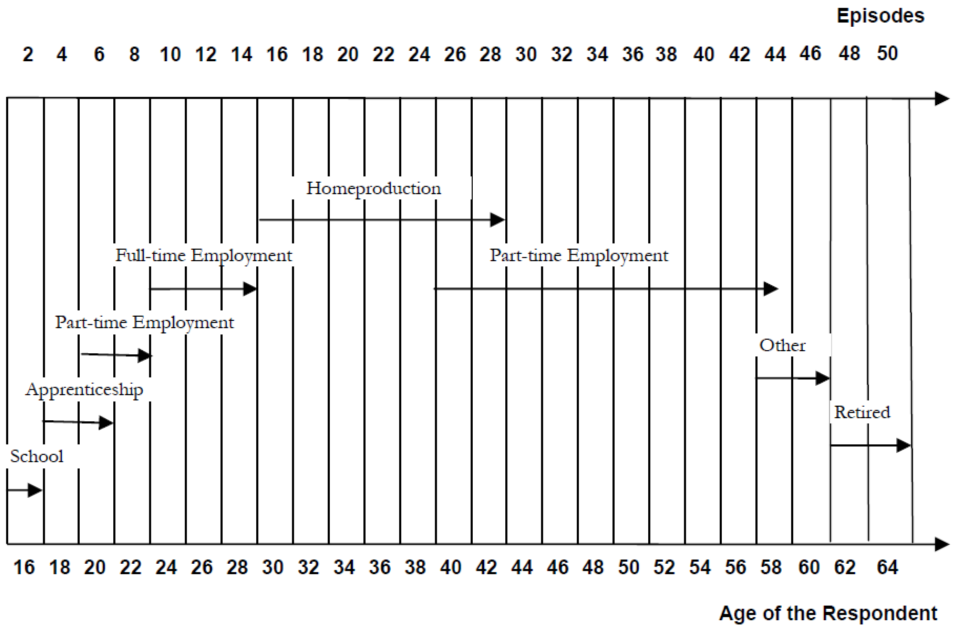
	Activity
A 1	school/university
A 2	training/apprenticeship
A 3	military/civilian Service
A 4	full-time employment
A 5	part-time employment
A 6	unemployment
A 7	home production
A 8	retirement
A 9	other activities
A 10	missing

Source: Frick and Lohmann (2010)

In the ideal case, we have information for 51 years. Between ages 15 to 65, the individual i spends his/her time in different activities a. Activities can overlap, which means that a person can report more than one activity in a given year y. *Figure 5* illustrates a fictitious employment history of a person between ages 15 to 65.

[93] In the remainder of the paper the terms *types of employment* or *types of activities* will be used as equivalents.

Figure 5 Fictitious Employment History



Source: Himmelreicher and Viebrok (2004)

In *Figure 5* periods of apprenticeship and part-time employment overlap at age 19 and 20, home production and part-time employment overlap between age 39 and 43, and part-time employment and other activities overlap at age 57 and 58. SOEP data has to be recoded so that it fits the information available in the SUF VVL 2004. In the case of overlapping periods in a given year, activities were weighted according to the number of activities reported in that year. We apply an *equal distribution assumption*, which means that every full year is divided by the number of activities reported in that year. We need to use this simplifying assumption because information is only available on an annual basis for the majority of years.[94] Based on the fictitious employment history in *Figure 5* the *equal distribution assumption* implies that at age 19 the year is divided by two activities. Hence, six months are credited towards each category: apprenticeship

94 Monthly information is available for the years an interview was given. In the ideal case, we have monthly information on the occupational status if a person participated in all 22 waves of the SOEP. For the time prior to the first interview, information is only collected on an annual basis in the employment history questionnaire.

and part-time employment.

If a person reports the activity home production, we deviate from the above assumption. Home production does not count if other types of employment are reported in that same year. This deviation is necessary because some women are likely to report that they are in home production while they are working full-time, whereas others are not likely to report being in home production. If we applied the *equal distribution assumption*, the time women spend in home production would be overstated relative to the time spent in other types of employment. For example, in *Figure 5* part-time employment and home production overlap between the ages 39 and 43. In this situation, we count four years in part-time employment and dismiss the time spent in home production.

For each person, the time spent in the nine different types of employment is summed up over the years 15 to 65. If the person reports no type of employment in a given year, the year is coded as missing. Even if there are gaps in the employment history, the number of years should add up to 51 years for every retired person. *Table 12* shows how we translated information from the example employment history for our purposes.

Table 12 Translating a Hypothetical Employment History in the SOEP

Age	Episode	Number of activities	Number of Years Counted
15 -17	school	1	2 years school
17-19	apprenticeship	1	2 years apprenticeship
19-21	apprenticeship/part-time	2	1 year apprenticeship/1 year part-time
21-23	part-time	1	2 years in part-time
23-29	full-time	1	6 years full-time
29-39	home production	1	10 years home production
39-43	home production/part-time	2	4 years part-time
43-57	part-time	1	14 years part-time
57-58	part-time/other	2	6 months part-time & 6 months other
58-61	other	1	3 years other
61-65	retired	1	5 years retired
	Total		51 years

Source: Author’s illustration

The SUF VVL 2004 has an equivalent file to *PBIOSPE* in the SOEP. The so-called *SES*-file covers various pension-relevant states that count in terms of pension entitlements. Unlike *PBIOSPE*, the *SES*-file distinguishes thirteen types of employment, which are listed in *Table 13*.

Table 13 Activities in the SUF VVL 2004

	Activity
SES 1	school/university
SES 2	apprenticeship/training
SES 3	home production
SES 4	Unemployment
SES 5	military/civilian service
SES 6	other activities
SES 7	care giving
SES 8	invalidity/sickness
SES 9	employment subject to social insurance contributions
SES 10	marginal employment
SES 11	self-employment
SES 12	Disability pension
SES 13	old-age pension

Source: Stegmann (2006a)

Variables pertaining to the social employment situation are non-missing if the episode is relevant for a person's pension entitlements. For example, the self-employed can opt to pay social insurance contributions on a voluntary basis. Under these circumstances, the employment situation *self-employed* applies. However, if a self-employed person does not pay voluntary contributions in the social security system but instead invests in a private pension scheme, this type of employment does not fall under the social employment situation *self-employed*. If none of the above types of employment applies in a given month, a missing value appears.

In the SUF VVL 2004, information is available on a monthly basis. Hence, the time a person spent in each employment situation can be summed up more precisely in the SUF VVL 2004 than in the SOEP. The SES file starts in January the year a person turned 14 and ends in December the year a person turns 65 (Stegmann 2006a). In the ideal case, information is available for 624 months (52 years times 12 months). *Table 14* gives a simplified illustration of the *SES*-file.

Table 14 Structure of the Longitudinal File Social Employment Situation

	SES001	...	...	SES312	...	...	SES624
Activity	School			Employment Subject to Social Insurance Contributions			Retired

Source: Stegmann (2006a); modified for own purposes

In the case of the SES file, we do not have to use the simplifying *equal distribution assumption*, because even if types of employment overlap, only one type of employment is recorded. In the case of overlapping types of employment, the decision as to which type of employment to record depends on a set of priority rules. The priority rules are already applied in the process of data preparation and serve the purposes of anonymization and the reduction of complexity of pension data (Stegmann 2006a). The priority rules depend on the type of contributions that are paid into the system. Employment that is subject to social insurance contributions is prioritized against all other types of employment. Then follow voluntary contributions (*freiwillige Beitragszeiten*), creditable periods (*Anrechnungszeiten*), credited substituted periods (*Ersatzzeiten*), receipt of public pension benefits (*Rentenbezug*), child care credits and raising several children (*Kindererziehungszeit und Erziehung mehrerer Kinder*), as well as child care periods and credits (*Kinderberücksichtigungszeit und Gutschrift*) (for further details see Stegmann 2006a). Due to these priority rules, the time spent in the different pension-relevant states can easily be summed up over the life cycle.

3.5.2.2 Summary Statistics: Average Time Spent in ...

In order to get a better understanding of the data, we first calculate the average time spent in various types of employment in the age group 15 to 65 in SOEP and SUF VVL 2004. These calculations are a first step in order to find out, whether the two datasets measure similar concepts.

There are two different ways to calculate the average time spent in each pension-relevant state: The first option is to consider all persons in the denominator, independent of whether or not individuals have spent time in a certain pension-relevant state. For example, if a person spent no time in home production, he/she will still be counted in the denominator. An average value of five years spent on home production therefore needs to be interpreted as follows: for all persons in the defined subsample, the average duration spent in home production amounts to five years. The alternative option is to consider only non-zero observations, which means only those individuals that have actually spent time in a certain type pension-relevant state. If a person did not spend any time in home production, then this observation is not considered in the denominator. A person that spent five years in home production is considered in the denominator. An average value of 12 years in home production therefore needs to be interpreted as follows: for those persons who have spent time in home production, the average duration spent in home production amounts to 12 years.

We distinguish different demographic groups when calculating the average time spent in the nine types of activities. In the first set of calculations, we only distinguish between men and women. In the next step, we distinguish between men and women in East and West Germany.[95] Furthermore, we distinguish between Germans and persons with a history of migration.[96] The average time spent on different types of employment

---

[95] The East-West distinction is based on the variable *vbula*. The variable *vbula* distinguishes the 16 different states (*Bundesländer*) of the Federal Republic of Germany. The variable *East* captures the following five states: Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, and Thüringen. The variable *West* captures the following 11 states: Baden-Württemberg, Bayern, Berlin, Bremen, Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, and Schleswig-Holstein. Given that it is not possible to distinguish between East and West Berlin in the VVL data, we subsume Berlin under the *West* category.

[96] To identify a person with a history of migration in SOEP data, we first checked whether a person had German citizenship in the year 2005 (variable: *nation05*) and whether the person had German citizenship since birth, or whether it was acquired later (variable: *vp137*). The variable *germborn* indicates whether a person was born in Germany or immigrated after 1948. If a person reports that he/she immigrated after 1948, then the variable migration history was coded with 1. The construct validity of our migration variable was double-checked with the variable *immiyear*, which indicates the year of immigration. If a person reports a year of immigration, the person was

were first calculated for Germans and persons with a history of migration together and then calculated separately for the two groups. This step helps to illustrate how the group of migrants differs from Germans and how to handle the group of persons with a history of migration in the multivariate analyses and the actual statistical matching.[97]

For the SOEP, we calculate the average time spent for three different populations: (1) all retirees in 2005; (2) first-time pensioners from 2000 to 2004; and (3) first-time pensioners in 2003 and 2004. In the results, we distinguish two different categories of home production. The first category sums up all periods of home production, independently of whether they overlap with other types of employment. The second category considers periods of home production only if no other activities apply. The activity type missing applies if no activity was reported in a given year. The variable sum adds up the time spent over all types of activities.

In principle, the calculation of the average time spent in the thirteen social employment situations of the SUF VVL 2004 follows the same rules. It only differs in some respects from the SOEP calculations: First, the average time spent on the VVL types of employment was only calculated for the population of first-time old age pensioners in 2004. Furthermore, no analytic weights were considered in the calculations for the VVL, because no such weights exist.[98]

3.5.2.3 Making Results Comparable

Aggregating the time individuals spent in different types of activities enables us to use the variables for the statistical matching procedure. Howev-

---

expected to have a migration history; hence, the variable migration equals 1.

[97] For the calculation of the average time spent in various activities, we apply the analytic weights attached to each observation in the SOEP to control for the different sampling probabilities.

[98] For the results, see *Table A6* to *Table A29* in the Appendix.

er, the two datasets differ in terms of the number and kinds of activities they distinguish. We therefore have to align the activities in both datasets applying plausible assumptions. *Table 15* illustrates the proceeding (also note the shaded cells). The alignment of activities happens in two steps: First, we align the fourteen categories in the SUF VVL 2004 (*Column 1*) with the ten SOEP categories (*Column 2*). In the SUF VVL 2004, we summarize the following categories: employment subject to social insurance contributions, marginal employment, and self-employment. This summary measure should capture the same types of employment as the SOEP categories full-time and part-time employment, respectively. For the statistical matching, we will have one category employment. The categories care, other, and invalidity & sickness in the SUF VVL 2004 correspond to the other category in SOEP data. The third column lists the final nine categories that are relevant for the statistical matching procedure.

Table 15 Streamlining Types of Employment from SUF VVL 2004 and SOEP

Column 1		Column 2		Column 3
SUF VVL 2004		SOEP		Final Categories
school/ university	→	school/university	→	school/university
apprenticeship/training	→	apprenticeship/training	→	apprenticeship/training
home production	→	home production (only exclusive spells)	→	home production
unemployment	→	unemployment	→	unemployment
military/civilian service	→	military/civilian service	→	military/ civilian service
other caregiving invalidity and sickness	→	other	→	other
employment subject to social insurance contributions marginal employment self-employed	→	full-time (also self-employment) part-time	→	employment
disability pension old-age pension	→	retirement	→	retirement
Years Missing	→	Years Missing	→	Years Missing

Source: Author's illustration

Table A30 in the Appendix compares the average time spent in the different types of activities after the alignment of categories in the SOEP and the SUF VVL 2004. We compare the SUF VVL 2004 results with the SOEP results for the group of first-time pensioners between 2000 and 2004 as well as for the group of first-time pensioners in the years 2003 and 2004.

3.5.3 Gender

In addition to the time each individual spent in the different types of employment, we need additional variables for the statistical matching of both datasets. One of the most important variables is gender. The employment histories of women differ to a great extent from those of men, with corresponding consequences for the public pension benefits. Therefore, gender is one of the most important variables for the matching procedure. *Table 16* shows the distribution of gender in both samples.

Table 16 Distribution of Variable Gender in SOEP and SUF VVL 2004, weighted [99]

Gender	SOEP 2000 – 2004		SUF VVL 2004	
	N	Percent	n	Percent
Male	443	44.0	13,983	45.4
Female	506	56.0	16,846	54.6
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

The distribution of gender is quite similar in both datasets with the share of females being slightly higher (male to female ratio of 1 to 1.17).

3.5.4 Region

The variable region distinguishes whether a person lives in East or West Germany in 2004.[100] We opt for the East-West differentiation rather than less aggregated state dummies, because of the greater explanatory power of

[99] For the following cross-tabulations, analytic weights were applied.
[100] For details on the coding of the *region* dummy see footnote 92.

the East-West distinction. This variable best captures the geopolitical, institutional, and economic differences between the former GDR and FDR. The distinction between these two parts of Germany is necessary, despite the German reunification in 1990. This study focuses on the cohort of retirees that spent most of its working life under one or the other regime, which in turn strongly affected their respective employment histories. For example, the average employment history of an East German woman is more similar to that of a West German man than it is to the employment history of a West German woman. *Table 17* shows how the variable *region* is distributed in the two datasets.

Table 17   Distribution of Variable Region in SOEP & SUF VVL 2004

Region	SOEP 2000 - 2004		SUF VVL 2004	
	N	Percent	n	Percent
West	662	74.6	23,656	76.7
East	287	25.4	7,173	23.3
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

The share of individuals living in East and West Germany is quite similar in both datasets, with the proportion of East Germans being slightly higher in the SOEP.

3.5.5 Marital Status

The marital status of a person is another important predictor for the level of public pension benefits. Both datasets contain information on the individual's marital status, but the information differs in two respects. First, the SUF VVL 2004 measures marital status only at retirement. Hence, there is no information about changes in marital status over the life-course. Second, the administrative data distinguishes only two categories: married (capturing married and remarried) and not married (capturing widowed, divorced, never married).

The SOEP measures marital status longitudinally; hence, changes in status can be followed over time. Furthermore, the SOEP distinguishes five categories of marital status: married and living together, married but living apart, never married, divorced, or widowed. For the matching procedure, we align the SOEP information with the VVL data. For each person, we use the marital status information at retirement. The new marital status category *married* subsumes the two categories married and living together and married but living apart. The other new category *not married* includes the categories never married, divorced, and widowed. After the successful matching, we can return to the more detailed and longitudinal marital status information in the SOEP. *Table 18* compares marital status in both datasets.

Table 18 Distribution of Variable Marital Status in SOEP and SUF VVL 2004

Marital Status	SOEP 2000 – 2004		SUF VVL 2004	
	N	Percent	n	Percent
Not Married	169	23.4	7,173	22.4
Married	780	76.6	23,656	77.6
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

We can attribute the small differences in the distribution of marital status in SOEP and SUF VVL 2004 to differences in measurement. Despite these differences, the distribution is fairly similar in both datasets.

3.5.6 Number of Children

Information about the number of children and their respective birth dates are available in both datasets. However, in the SUF VVL 2004, this information is only available for women, but not for men. In pension data, information on children is typically assigned to the mother. Specifics in pension law explain this pattern, because the statutory pension insurance gives pension credits to parents during the first three years of the child's life. These credits go to the pension account of the primary caregiver, typically the mother. Exceptions to the rule occur either if the mother dies or if the mother works as a civil servant. Consider a situation in which the mother works as a civil servant and the spouse is gainfully employed and obligated to pay contributions into the statutory pension insurance. In this situation, child care credits go to the account of the spouse.[101] In the SUF VVL 2004 only one percent of all children are assigned to the records of the father (n=180), whereas 99 percent (n=15,178) are assigned to the mother's pension accounts (Himmelreicher and Mai 2006).

The SOEP provides information on children for fathers and mothers. The file *BIOBIRTH* and *BIOBRTHM* contains information about the birth history of men and women (Frick and Lohmann 2010). They provide information about the birth history in the biographical questionnaire asked in the first interview. The birth information is then updated each year on the basis of data collected in the individual questionnaire. This procedure assures that the complete birth history is available for all SOEP respondents. In order to assure comparability, we only consider information on the number of children for women in the statistical matching. *Table 19* provides the results of the distribution.

Table 19 Distribution of the Variable Number of Children

Number Of Children	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
No children	50	13.0	2,003	11.9
One Child	109	26.1	4,000	23.7
Two Children	205	37.6	6,311	37.5
Three Children	90	14.0	2,918	17.3
Four Children	35	7.0	1,010	6.0
Five+ Children	17	2.3	604	3.6
Total	506	100.0	16,846	100.0

Source: FDZ-RV – SUFVVL2004 & SOEP 2005; Author's calculations

Results in *Table 19* illustrate that the congruence between the two datasets

[101] According to §56, (4) SGB VI civil servants are not eligible for child care credits from the statutory pension insurance.

is better for high-parity mothers with four and more children and for mothers with two children. Differences are larger for women with no children, one child or three children. These discrepancies could be due to the fact that children are only visible in pension data if mother's report them.

3.5.7 Retirement Age

Both datasets provide information on the individual's age at retirement. Due to the exclusion of disability pensioners from the sample population, the earliest possible retirement age is 60 years.[102] There is also an upper limit of 65 years, which is due to the sampling design of the SUF VVL 2004 that does not cover individuals older than 65. The SOEP questionnaire does not include a direct question, but can be reconstructed using the biography questionnaire *PBIOSPE*.[103] *Table 20* summarizes the distribution of retirement age in both datasets, with the mean retirement age of the sample given at the bottom.

[102] This is in line with our expectations. According to current pension rules, it is impossible to receive any kind of old-age public pension benefit (e.g. old-age pensions for women, old-age pensions due to unemployment, etc.) before age 60.

[103] Some persons report repeated episodes of retirement, some of which start before age 60. Persons with repeated episodes of retirement presumably received disability pensions. We solved the problem by taking the maximum starting age of the episode retirement.

Table 20 Distribution of the Variable Retirement Age

Retirement Age	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
Age 60	299	31.3	7,998	25.9
Age 61	169	17.8	2,287	7.4
Age 62	86	9.4	3,152	10.2
Age 63	119	13.0	3,911	12.7
Age 64	82	9.8	1,979	6.4
Age 65	194	18.7	11,502	37.3
Total	949	100.0	30,829	100.0
Mean Retirement Age	62.10		62.77	

Source: FDZ-RV – SUFVVL2004 & SOEP 2005; Author's calculations

There are apparent differences in the distribution of the retirement age in the data, in particular at age 61 and 65. In the group of SOEP first-time old-age pensioners, 17.8 percent retired at age 61, compared to only 7.4 percent in the SUF VVL 2004. According to the SOEP, only 19 percent retire at age 65, compared to more than 37 percent in the administrative records.

A comparison to the official statistics of the statutory pension insurance supports the distribution found in the administrative records (cp. to *Table 21*), whereas *PBIOSPE* seems to be less adequate to determine the individual's exact retirement age.[104] In the official statistics, the retirement age spikes at age 60, 63, and 65 (Verband Deutscher Rentenversicherungsträger 2005, multiple volumes). The comparison of *Table 20* and *Table 21* illustrates the expectedly good correspondence of the SUF VVL 2004 with official statistics. Column 6 provides aggregate data from the official statistics for the retirement inflows of the years 2000 to 2004. These are pooled inflows providing a measure of comparison for the SOEP data. Apart from the large deviation for individuals who retired at

[104] The better fit between the SUF VVL 2004 and official statistics does not come as a surprise, because data come from the same source, namely the statutory pension insurance.

age 61, the distribution of the retirement age is fairly similar between the group of first-time pensioners in the SOEP and those in the official statistics. Despite these deviations, the mean retirement age is almost exactly the same in both datasets with about 62 years.

Table 21 Distribution of Retirement Age for First Time Old-Age Pensioners between 2000-2004

Retirement Age	Retired in 2000	Retired in 2001	Retired in 2002	Retired in 2003	Retired in 2004	Retired 2000-2004
Age 60	46.10	40.09	30.93	26.24	25.90	33.85
Age 61	5.64	7.08	10.92	10.80	6.42	8.17
Age 62	3.43	5.58	6.12	6.48	9.10	6.14
Age 63	10.90	11.21	12.15	13.51	13.19	12.19
Age 64	1.48	1.77	2.74	2.78	2.80	2.31
Age 65	30.19	32.04	34.88	37.33	39.87	34.86
Age 66+	2.26	2.23	2.26	2.87	2.72	2.46

Source: Deutsche Rentenversicherung Bund Rentenzugang 2000-2004; Author's calculations

A possible explanation for the differences between the SOEP and SUF VVL 2004 might be the interaction of age, cohort, and period effects that result from the pooling of first-time retirees in the SOEP in the years from 2000 to 2004 (Fachinger and Himmelreicher 2008). In addition, the small number of observations in the SOEP might contribute to differences in the distribution of the retirement age in both datasets.

3.5.8 Migration History

When comparing the average time individuals spent in different activities in the two datasets, we applied a broad indicator for individuals with a migration history, because we expect individuals with a migration history to differ from German natives. This broad indicator identified migrants based on citizenship status, changes in citizenship status and immigration year. For the statistical matching, the migration indicator in the SOEP has to be aligned with the citizenship variable available in the SUF VVL 2004.

In the administrative data, persons with a history of migration were identified using the variable indicating the person's citizenship (variable *SA*).

The variable SA only discriminates between German citizenship and non-German citizenship. Hence, the SUF VVL 2004 defines a much narrower construct for migration history.[105] *Table 22* illustrates the distribution of the variable migration history in both datasets.

Table 22 Distribution of the Variable Migration History in the SOEP and SUF VVL

Migration History	SOEP 2000 – 2004		SUF VVL 2004	
	N	Percent	n	Percent
Yes	103	4.1	660	2.2
No	846	95.9	30,169	97.8
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

With more than 4 percent, the share of persons with a history of migration is larger in the SOEP than in the VVL (~2.2 percent). The exclusion of persons who fall under the Foreign Pension Law (*Fremdrentner*) in the initial specification of the sample is one explanation for the lower share of individuals with a migration history. This group cannot be identified in the SOEP, because it is impossible to tell pension benefits from the German public pension scheme and a foreign pension scheme apart. Another explanation is that not every person with a migration history is eligible for pension benefits from the statutory pension insurance. These individuals are not part of the SUF VVL 2004 but part of the SOEP sample population.

3.5.9 Type of Health Insurance

Retirees can either hold a public or private health insurance plan. In the statutory health insurance, the payment of contributions by members is

[105] In the specification of the SUF VVL 2004 sample population, we decided to forego a broader definition of the variable *migration history* by excluding persons who fall under the regulations of the Foreign Pension Law. Persons whose pension is subject to a bilateral social security agreement were excluded from the sample completely.

either mandatory or voluntary. The SOEP asks respondents for the type of health insurance they hold. The interviewer further asks for the type of membership, namely whether the respondent is a mandatory or voluntary paying member of health insurance premiums. In the SUF VVL 2004, the variable *AT* provides information about the type of health insurance. The SUF VVL 2004 summarizes voluntary paying members *and* members of a private health insurance in one category. Another category is the group of mandatory paying members. The third category is the group of persons that are not insured according to German law.

In order to harmonize the variable type of health insurance in both datasets, we combine the SOEP information from the variables asking respondents about their type of health insurance and about what type of members they are. We assign persons who report to be privately insured to the category voluntary paying members or members of a private health insurance. The same applies to persons who report to be voluntary paying members. All others fall in the group of mandatory paying members. The category persons not insured according to German law is an administrative peculiarity in the SUF VVL 2004. According to information from the statutory pension insurance, the majority of observations in this category are persons whose health insurance status has not been validated at the point of data preparation. We classify these observations as mandatory paying members in order to obtain a comparable measure for the variable type of health insurance and accept the slight inaccuracy of this procedure.

Table 23 Distribution of the Variable Type of Health Insurance

Health Insurance	SOEP 2000 – 2004		SUF VVL 2004	
	N	Percent	n	Percent
Statutory	863	90.9	28,613	92.8
Private	86	9.1	2,216	7.2
Total	949	100.0	30.829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

The inconsistencies in the frequency distributions are the result of differences in the operationalization of the variable type of health insurance in both datasets. In particular, the category *not insured according to German law* brings uncertainty in the alignment of variables in both datasets that cannot be easily resolved.

3.5.10 Educational Attainment

Educational attainment is a crucial variable with high explanatory power for the individual's lifetime earnings and consequently, the level of public pension benefits the person receives as he/she retires. Variables that describe the educational attainment of a person are available in both datasets. However, these variables differ considerably in terms of their measurement. The SUF VVL 2004 combines the highest secondary or tertiary schooling degree with information about the completion of vocational training (Fitzenberger et al. 2006). However, the measure lacks reliability. The lack of reliability is due to the fact that the variable has no relevance whatsoever for the calculation of the individual's public pension benefit. Hence, there is no incentive for employers to invest much time and manpower in submitting accurate information to the branches of the social insurance system. As a consequence, the variable in the SUF VVL 2004 has a high number of missing values. For the statistical matching, we need to determine whether the variables for educational attainment are comparable in both datasets and hence useful matching variables. To give a definite answer to this question, we align the operationalization in both datasets. We modify the SOEP variables so they fit the information provided in the SUF VVL 2004. The second step compares the distribution of the modified variable educational attainment in both datasets and analyze whether we find a positive education gradient in the level of public pension benefits.[106] *Table 24* displays the operationalization of educational attainment in

[106] Educational attainment has been found to have a positive effect on the level of public pension benefits in other studies based on data from the statutory pension insurance (Rehfeld et al. 2007).

the SUF VVL 2004.

Table 24 Distribution of the Variable Educational Attainment in the SUF VVL 2004

Value Labels for Different Categories of Educational Attainment (based on TTSC3)	Value	Share in % (n)
Missing Information	-9	49.8 (15,347)
Secondary school or higher secondary school without vocational training *(Hauptschule/Realschule ohne abgeschlossene Berufsausbildung)*	1	6.4 (1,967)
Secondary school or higher secondary school with completed vocational training *(Hauptschule/Realschule ohne abgeschlossene Berufsausbildung)*	2	27.1 (8,355)
High school or technical high school without vocational training (Abitur oder Fachhochschule ohne abgeschlossenen Berufsausbildung)	3	0.2 (57)
High school or technical high school with completed vocational training (Abitur oder Fachhochschule ohne abgeschlossenen Berufsausbildung)	4	0.9 (278)
Completed degree at *Fachhochschule* (Abschluss einer Fachhochschule)	5	2.1 (647)
Completed degree at a university or technical university (Hochschul-/Universitätsabschluss)	6	2.4 (746)
No information available/ degree unknown	7	11.1 (3,432)

Source: FDZ-RV - SUFVVL2004, own calculation

To obtain a comparable measure in the SOEP data, we had to restructure the information. *Table 25* illustrates the approach that we used.

Table 25 Educational Attainment Information in the SOEP and Alignment with SUF VVL 2004 Categories

SOEP	
$PSBIL School Education	
Missing	-1
Secondary School	1
Higher Secondary School	2
Fachhochschulreife	3
High school	4
Other degree	5
No completed degree	6

SOEP	
$PBBIL01 Vocational Training	
Does not apply	-2
Missing	-1
Apprenticeship	1
Full-time vocational school	2
School for health care professions	3
Trade and technical school for vocational education	4
Training for public employees	5
Other training	6

SOEP	
$PBBIL02 Tertiary Education	
Does not apply	-2
Missing	-1
University of Applied Sciences	1
University, Technical University	2
University in a foreign country	3
Engineering School and School of Applied Sciences of former GDR	4
University of former GDR	5

SOEP	
$PBBIL03 Completed Degree	
Does not apply	-1
Missing	-2
No completed degree	1
College Degree	2

COMBINATION OF SOEP VARIABLES					**VVL CATEGORIES FOR EDUCATIONAL ATTAINMENT**	**NEW EDUCATION VARIABLE**
PSBIL	**PBBIL01**	**PBBIL02**	**PBBIL03**		**TTSC3**	
5					No information available/ Degree unknown	Unknown (-2)
- 1					Missing Information	Missing (-1)
6					(School Drop-Out: no degree)	Low
1 or 2			1		Secondary school or higher secondary school without vocational training	Low
1 or 2	>0			→	Secondary school or higher secondary school with completed vocational training	Medium
3 or 4			1		High school or technical high school without vocational training	Medium
3 or 4	>0				High school or technical high school with completed vocational training	Medium
3 or 4		1 or 4			Completed degree at university of applied science	High
3 or 4		2 or 3 or 5			Completed degree at a university or technical university	High

Notes: If information was missing for the variable *PSBIL*, we combined information from *PBBIL01*, *PBBIL02* and *PBBIL03* to obtain a comparable measure. Source: SOEP 2005; Author's Illustration

The four upper boxes present the four educational attainment variables in the SOEP. At the bottom of *Table 25*, the first four columns show how these variables were combined in order to match the measure in the SUF VVL 2004. Following Haak (2006), we then constructed a new education variable that differentiates between low, medium, and high educational attainment. The category of school dropouts, which is a category in the SOEP, but not in the SUF VVL, was grouped under low educational attainment.[107] The distribution of the new variable for educational attainment is almost identical in both datasets. *Table 26* illustrates the distribution in the two datasets, considering only valid values.[108]

Table 26 Distribution of New Variable Educational Attainment

Educational Attainment	SOEP		SUF VVL 2004	
	N	Percent	n	Percent
Low	156	17.8	1,967	16.6
Medium	590	66.0	8,690	72.1
High	175	16.3	1,393	11.6
Total	921	100.00	12,050	100.00

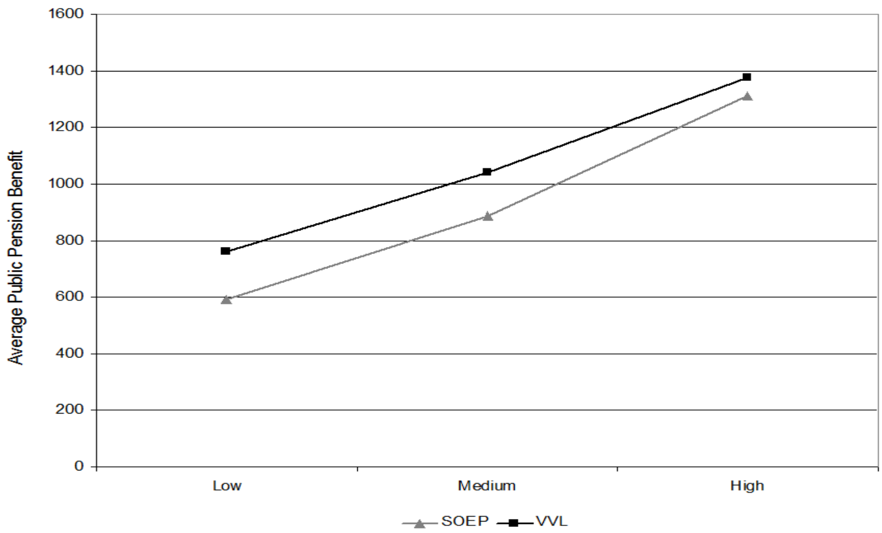
Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

Figure 6 confirms the positive education gradient with respect to average pension benefits. Higher educational attainment goes along with a higher public pension benefit. However, the large share of missing values requires further investigation. The statistical matching does not allow for missing values and dropping all observations that lack information on educational attainment is no option. The multivariate analyses following in the next sections might provide answers on whether and how to include the variable in the statistical matching.

107 In contrast to Haak (2006), persons who have completed high school or technical high school but have not completed vocational training were categorized in the group of medium educational training.

108 Not that considering only valid values implies that a large share of the SUF VVL 2004 sample drops out of the calculation

Figure 6 Returns to Education based on SOEP and SUF VVL



Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

3.6 Estimating Regression Equations

3.6.1 Which Variables Enter Which Model?

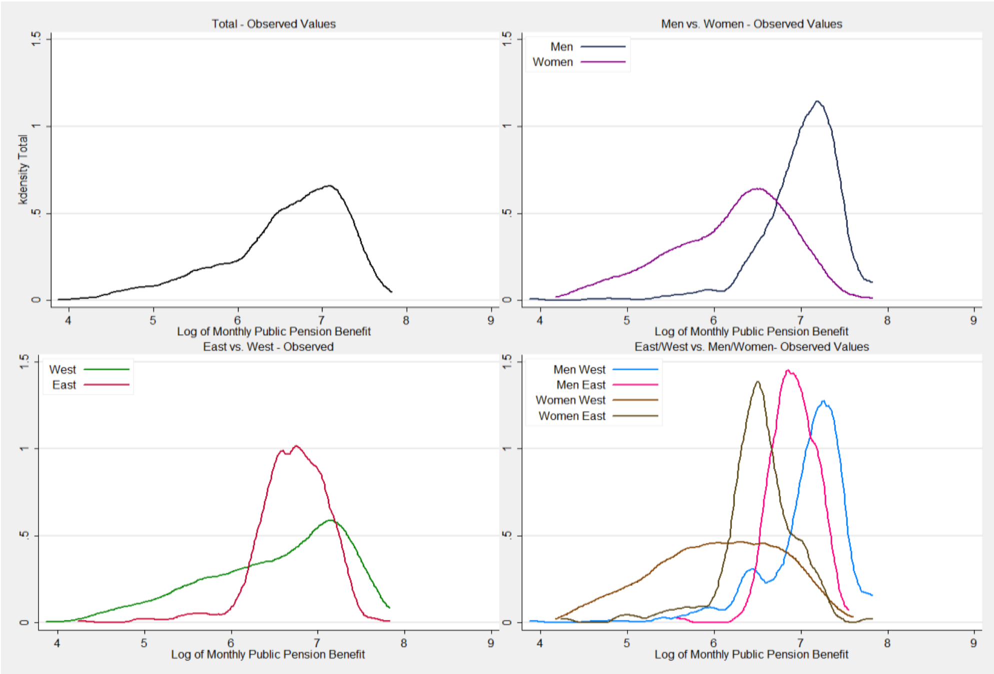
After specifying the two populations of interest and ensuring that the distributions of the core variables are similar, we test how the SOEP compares to the SUF VVL 2004 in a multivariate analysis. This step is necessary because the preparation of the statistical matching involves a regression procedure that still has to be determined. This section compares the estimates of various regression models in terms of strength and direction for both datasets. A total correspondence is rather unlikely, due to differences in the measurement of certain variables and considerable differences in sample sizes.

We run multivariate analyses based on the SUF VVL 2004 population and the group of first-time pensioners that retired between 2000 and 2004 in the SOEP. We opt for this SOEP population – instead of first-time retirees between 2003 and 2004 – because it comes closest to the SUF VVL 2004 with respect to the distribution of the core matching variables. Plus, the pooled SOEP population has a reasonable sample size, which allows for the estimation of regression models for different demographic groups, namely East and West German men and women. The dependent variable is the logged monthly pension benefit.

Figure 7 and *Figure 8* show the distributions of logged pension benefits in SOEP and SUF VVL 2004 data. In both figures, the upper left panel shows the logged monthly pension benefit for the total population. The upper right panel compares logged benefits for men and women. The lower left panel compares the distribution of benefits in East and West Germany and the lower right panel the distributions for West and East German men and women. Between datasets, distributions across demographic groups are fairly similar. There are no apparent differences – neither for the total population (upper top panel) nor for smaller subsamples (e.g. the distribution for men and women in the upper right panel).

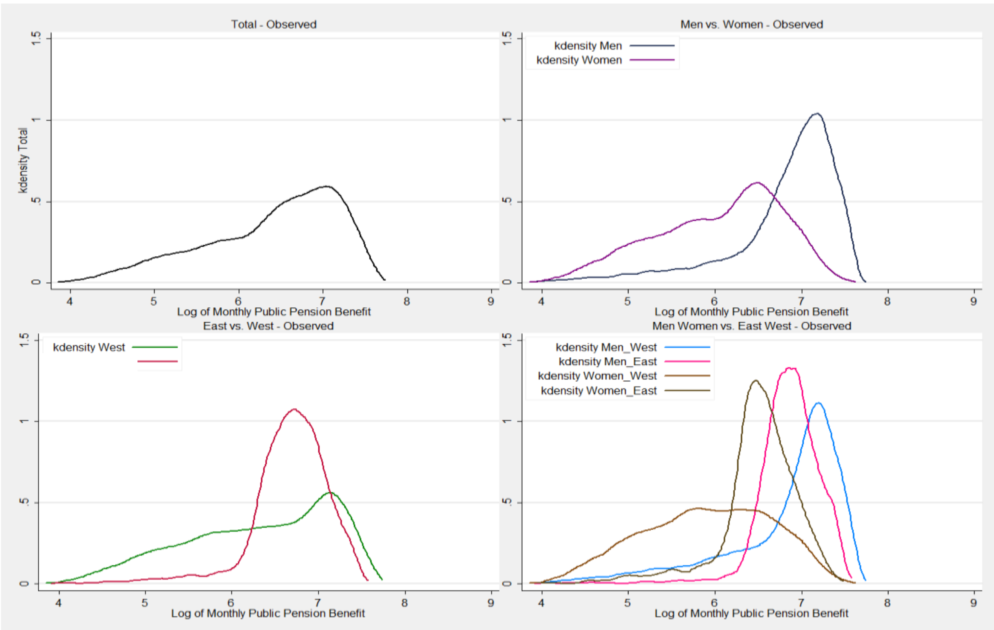
However, differences in the distribution of pension benefits across the four demographic groups illustrate why there is no alternative to separate regression estimations because of considerable differences in the distribution of logged benefits. In particular, the distribution for West German women deviates quite clearly from the rest of the population. West German stand out, not only because of the - on average - lower pension benefits, but also with respect to the average time spent in various activities (cp. to Section 3.5.2 and *Table A6 to Table A29*). In order to assess which model fits our purpose best, we go from a very general model based on the total sample population to subsamples separated by gender and region.

Figure 7 The Distribution of Monthly Public Pension Benefit for Different Demographic Groups in the SOEP



Source: SOEP 2005; Author's calculations

Figure 8 The Distribution of Monthly Public Pension Benefit for Different Demographic Groups in the SUF VVL 2004



Source: FDZ-RV - SUFVVL2004; Author's calculations

Table 27 summarizes the models we estimated based on nine different subsamples. Column 2 briefly describes each subsample. Column 3 lists the abbreviation we use for each subsample in the remainder of the paper. Columns 4 and 5 compare the case numbers per subsample in each dataset.

Table 27 Subsamples within the Sample Population and Case Numbers

Mod-el	Population Specification	Model Abbreviation	SOEP N	SUF VVL N
I	Total sample population	Total	949	31,744
II	Only West Germany, men & women	Total West	662	24,213
III	Only East Germany, men & women	Total East	289	7,261
IV	Only men, East & West Germany	Total Men	443	12,274
V	Only women, East & West Germany	Total Women	506	17,200
VI	Only men, only West Germany	West Men	304	10,727
VII	Only men, only East Germany	East Men	139	3,547
VIII	Only women, only West Germany	West Women	358	13,486
IX	Only Women, only East Germany	East Women	148	3,714

Source: Author's illustration

The first set of regressions includes variables that describe the average time spent in different activities, plus some basic controls for sex and region. The second set of regressions expands the number of controls including variables, such as migration history, family status, type of health insurance, retirement age, education, number of children, and educational attainment. These variables are potential variables for a statistical matching.

The dependent variable of our regression models is the monthly public pension benefit. The number of independent variables varies by subsample, e.g. not all regression models consider the same independent variables. The variable years in the military is excluded for women. For men, we do not consider the variable years in home production. Even though some women and men have valid values in the respective categories, we opt to exclude them from the estimation, because their inclusion potentially biases the estimates. The extended models do not consider the variable number of children, because information on the birth history is only available for women.

3.6.2 Regressions Diagnostics

3.6.2.1 Dependent Variable: Monthly Public Pension Benefit

The first set of regressions based on both datasets show clear differences in the results. The regressions based on SOEP data have throughout a highly negative constant, which is contrary to the first intuition and also different from the results based on the SUF VVL 2004. A closer look at the distribution of the dependent variable reveals some striking outliers in SOEP data. *Table 28* displays the summary statistics and the largest values of the distribution of the dependent variable.

Table 28 Summary Statistics of Public Pension Benefit in SOEP and SUF VVL 2004

Summary Statistics	SOEP	SUF VVL 2004
1st Percentile	97	76
1st Quartile	500	363
Median	811	773
3rd Quartile	1,200	1,184
99th Percentile	3,000	1,913
Mean	895.97	813,04
Standard Deviation	611.13	500,29
Largest Values	3,780	2,077
	4,500	2,088
	5,800	2,166
	8,500	2,294

Source: FDZ-RV - SUFVVL2004 and SOEP 2005; Author's calculations

The four highest values in the SOEP data range from 3780 Euro to 8500 Euro, which is far above the maximum possible monthly pension benefit within the current pension legislation. Because of the maximum contribution ceiling, a person can accumulate a maximum of two earning points per year.[109] Using a hypothetical earnings profile, we determine the max-

---

109 In the year 2004, the ceiling was set at monthly earnings of 5,150 Euro. No social insurance contributions are paid or earnings above this ceiling. Monthly earnings of 5,150 Euro roughly correspond to two earning points per year (Deutsche

imum possible monthly public pension benefit for any person to reach within the rules and regulations of the statutory pension insurance. We assume that a hypothetical person accumulates two earning points per year, each point being worth the actual pension value of 26.13 Euro.[110] In addition, the person works continuously with no interruptions for 45 years. Plugging these numbers into the simplified pension benefit formula (cp. to Section 3.5.1), our hypothetical person would receive a maximum monthly public pension benefit of 2,351 Euro. This value falls within the range of values in the SUF VVL 2004, as illustrated in the right column of *Table 28*.

The outliers in the SOEP data are cases of non-sampling errors, such as respondents who misinterpret the question and therefore report the annual instead of the monthly public pension benefit or respondents who report their total old age income adding up income from various pension schemes. Another explanation for non-sampling errors might be error on part of the interviewer. Instead of noting a monthly public pension benefit of 850 Euro, the interviewer might have noted a monthly public pension benefit of 8,500 Euro. It is impossible to tell which kind of error applies. We therefore top code the monthly public pension benefit at 2,500 Euro and opt against the case wise deletion of implausible cases, because of the already small number of observations in the SOEP sample. A total of thirty-four cases were top coded.[111] *Figure 9* compares different percentiles of the distribution of pension benefits in both datasets after the top coding. The distribution of the dependent variable in both datasets appears to be nearly congruent in the lowest decile. Between the second and the fifth decile the distribution disperses, but become very similar again in the further course. The large deviation at the 99th percentile is a

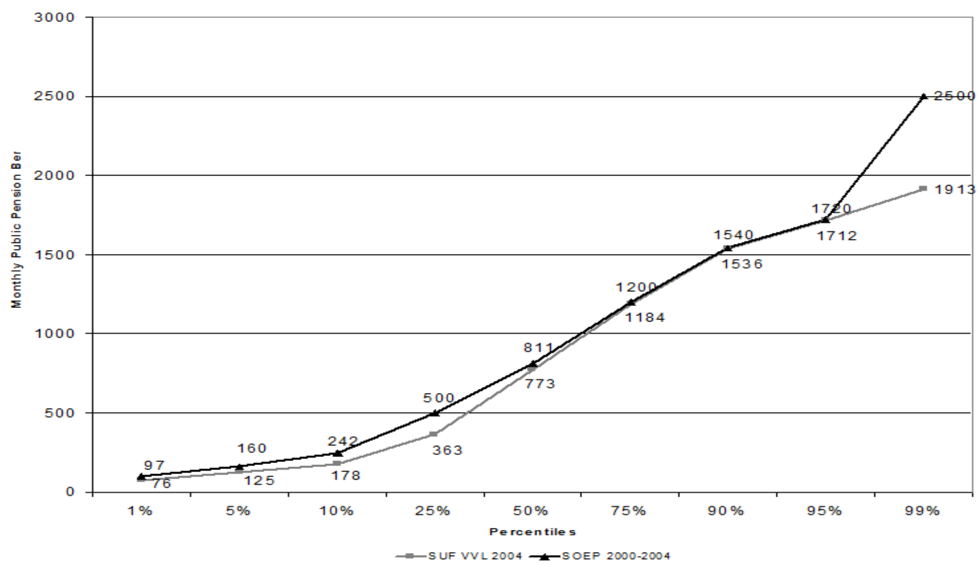
Rentenversicherung 2006b).

110 This is the actual pension value for West Germany.

111 The summary statistics in Section 3.5.1 already consider the topcoding in the SOEP.

result of the topcoding.

Figure 9 Distribution of Public Pension Benefit in SOEP & SUF VVL 2004 across Deciles



Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

3.6.2.2 The Relationship between Public Pension Benefits and Years of Employment

In an employment-centered public pension scheme - like the German system is – we expect to find a strong positive relationship between public pension benefits and years of employment subject to social insurance contributions. In this kind of pension system previous periods of employment are directly linked to the final public pension benefits. For periods of employment to count, they have to be subject to the payment of social insurance contributions.[112] With the payment of these contributions, the individual accumulates entitlements that later qualify for the receipt of public pension benefits.

112 Certain occupational groups are exceptions to the rule in that they are not obligated to pay social insurance contributions, e.g. the self-employed who can opt to pay voluntary contributions into the public pension insurance or pay money into a private pension plan.

Roughly speaking, individuals with long periods in employment typically receive high public pension benefits. Based on empirical evidence of the SUF VVL 2004, we find a positive relationship between public pension benefits and years of employment. Surprisingly, we do not find this expected relationship in the multivariate regressions based on SOEP data. It is to expect that these discrepancies between SOEP and SUF VVL 2004 are due to differences in the two sample populations. We know that the SUF VVL 2004 only covers individuals who retired in the year 2004 and that all earnings in the data are pension-relevant earnings. Hence, finding anything other than a strong relationship between years in employment and the monthly public pension benefit would have been implausible. All other forms of employment that have no relevance in terms of pension entitlements are missing. These are periods of self-employment, employment as a civil servant or other forms of employment that are not pension-relevant (e.g. illegal employment).

In contrast, SOEP respondents report periods of employment, irrespective of whether or not these periods are pension-relevant. Hence, we do not have any way of telling these different forms of employment apart. This explains why we did not find a clear-cut positive relationship between years of employment and public pension benefits. For example, the data shows respondents who report more than forty years of employment with very low public pension benefits. It is possible that these are observations in which the person worked for a few years in employment subject to social insurance contributions, but then became self-employed and no longer accumulated pension entitlements in the statutory pension insurance.

To address this problem, we control for the occupational status of a person for the years 1995 to 2004 (variable *stib$*). Two additional dummy variables enter the regression equation, indicating whether a person is self-employed or works as a civil servant. The data shows that in this cohort of retirees, very few people worked as civil servants in the years prior to retirement (n = 14), whereas the number of self-employed is slightly higher (n = 76). When we incorporate the two dummy variables into the regression models, the coefficients are more robust in the *Total* models (Model I through V).[113] In these models, the coefficients are intuitive in terms of strength and direction. The coefficients are strongly negative and significant. This finding is plausible, because self-employed and civil servants are, by definition, excluded from the public pension system unless they pay voluntary contributions. The coefficients are also more robust in the *Total Men* model than in the *Total Women* model. The difference is that there are more men than women working as self-employed or civil servants. This gender difference is at least true for the cohort of retirees we are interested in. If we differentiate the sample populations by region and gender (e.g. *West Men* Model), the results are less robust. Due to the small number of self-employed and civil servants in the sample population, we summarize the variables self-employed and civil servant to one dummy variable. We expect the coefficient to be negative. It is not possible to identify self-employed persons or civil servants in the SUF VVL 2004 data. Hence, these modifications only apply to SOEP data.

3.6.2.3 Years in Schooling and Years in Training

The regression results call for some additional, but minor modifications in both datasets. We apply a topcoding to the variables years in school and years in training, because we observe significant differences in the average time spent in these activities in both datasets. On average, respondents in the SOEP report approximately two years of schooling[114] and 2.3 years of training.[115] In the SUF VVL 2004, respondents report an average of 0.7

---

113 *Table 27* provides the abbreviations and their respective sample specification for the regression models.

114 Considering only non-zero values, respondents report on average 3.5 years in schooling.

115 Considering only non-zero values, respondents report on average 3 years of training.

years of schooling[116] and 1.35 years of training.[117] However, some very high and therefore implausible values distort the distribution. In SOEP data, the maximum value reported for years of schooling is 22 years and 28.5 years for training. In the SUF VVL 2004 in turn, the maximum value reported for schooling is 24.6 years and 12.8 years for training. For our analysis, only times relevant for the calculation of monthly pension benefits are of interest. These times either add to the minimum qualifying period or actually add pension entitlements (*creditable* or *contributory periods*). Therefore, we top code the variables years in school and years in training at a maximum of 10 years. *Table 29* lists how many cases are affected in both sample populations.

3.6.2.4 Years in Other Activities, Years Retired, and Years Missing

The distributions of the variables years in other activities, years retired and years missing also reveal a large variance. However, top coding is not an appropriate way to handle these variables. The question is rather whether the variables should enter the model on a continuous scale. The interpretation of the coefficients is as follows: for example, one additional year in other activities increases/decreases the monthly public pension benefit by some x amount. This interpretation does not make sense given that it is not at all obvious what type of activities fall in the category years in other. As an alternative, the variables can enter the regression equations as dummy variables. The new variable other equals one if the number of years in other activities exceeds three. The new variable retired equals 1 if the number of years retired exceeds four. And the new variable missing equals 1 if number of years missing exceeds three. In the multivariate

---

116 Considering only non-zero values, respondents report on average 2.9 years in schooling.

117 Considering only non-zero values, respondents report on average 2.7 years of training.

analysis, the coefficients of these new dummy variables indicate whether persons that have high values in the three original variables are systematically different from others, everything else being constant. *Table 29* summarizes the modifications and lists the number of cases affected by each modification.

One last word on the striking differences concerning the number of years missing in both datasets: In the SUF VVL 2004 all months equal missing that are not pension-relevant. If none of the 13 social employment situations apply (cp. to *Table 6*), then the month equals missing. In turn, SOEP respondents are free to report any activity they consider as relevant in the biography questionnaire.

Table 29 Data Modifications for Regression Analysis

Variable	Modification	Number of Observations affected by Modification in SOEP	Number of Observations affected by Modification in SUF VVL 2004
Years in school	Top Coding: If years in school exceed 10, then top-coding at 10 years of schooling.	35	8
Years in training	Top Coding: If years in training exceed 10, then top-coding at 10 years of training.	2	1
Years in other activities	Dummy Variable if years in other activities exceed three.	23	4,016
Years with missing information	Dummy Variable if years with missing information exceed three.	138	29,018
Years in retirement	Dummy Variable if years in retirement exceed four.	24	119

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

3.6.2.5 Years in Military

Should we apply the same procedure to the variable years in military? In two sets of regressions, the variable enters the equations continuously and as a dummy. The results indicate that it does not make a difference, because coefficients are not significant and weak in strength in both alternatives. Hence, the variable years in military is left out of the modified regression models.

3.6.2.6 Migration History

Surprisingly, the variable indicating a person's migration history lacks significance in all models – except the *Total* and *West* models. The lack of statistical significance in the East models is due to the small number of migrants in East Germany. We further refine the measure for migration history because of the heterogeneity within this group. For example, it is safe to assume that respondents from France have much more in common with Germans than respondents from Ghana. Therefore, we add the distinction between EU and Non-EU migrants. The group of EU-migrants consists of persons that come from the EU-14 countries (EU-15 minus Germany).[118] All others belong to the group of Non-EU migrants. Contrary to our expectations, it does not make a difference whether we include a general measure of migration history or a further refined one that distinguishes between EU- and Non-EU migrants. In both cases, the variables were automatically dropped from the estimation of the *East* models because of the small numbers of observations. The strength and direction of the coefficients for EU- and Non-EU migrants correspond to the migration coefficient. Therefore, we keep the variable in its original form.

3.7 Regression Results

In order to test the feasibility of a statistical matching of administrative pension records and population representative survey data, we now estimate another set of multivariate regression models based on the modified data with the logged monthly public pension benefit as the dependent variable. We assume the pension benefits to be a linear function of the independent explanatory variables discussed above (e.g. time spent in different

---

[118] In the meantime, additional countries entered the European Union. The EU-15 groups included the following countries: Belgium, Denmark, Germany, Finland, France, Greece, Ireland, Italy, Luxemburg, the Netherlands, Austria, Portugal, Sweden, Spain, and the UK.

types of employment, gender, region, etc.) that appear on the right-hand side of the equation. The overall goal is to find the model that best predicts the monthly public pension benefit. We therefore estimate roughly the same regression equation in both datasets. *Table 30* compares the explained variance (r^2) in each of the nine estimated models.

Table 30 Comparison of Explained Variance in SOEP and SUF VVL 2004

	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
SOEP	0.57	0.59	0.57	0.34	0.42	0.31	0.37	0.35	0.56
SUF VVL	0.78	0.80	0.70	0.74	0.70	0.76	0.61	0.67	0.67

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

The results in *Table 30* are in line with the expectations: regression models based on SUF VVL 2004 data explain significantly more variance than those based on SOEP data. These differences are a direct consequence of the SUF VVL 2004 only considering pension-relevant periods, in contrast to the SOEP that captures all activities, irrespective of whether these activities are pension-relevant or not. In both datasets, the model fit is best for the *Total West* model, with 80 percent of the variance explained in VVL 2004 and 59 percent explained in the SOEP. The *Men West* model has the least good fit in SOEP data, with only 31 percent of the variance explained. The *Men East* model has the least good fit in the SUF VVL 2004 data, with 61 percent of the variance explained.[119]

Table 31 boils down the results and only compares the direction and significance levels of the regression coefficients in both datasets. The upper left part of the cell captures the results based on SOEP data, the lower right box those based on the SUF VVL 2004. The boxes are green if the effect of the regression coefficients works in the same direction in both datasets. Boxes are red if the effect of the regression coefficients differs in terms of direction. Some variables only enter the regression equations of one da-

[119] *Table A32* and *Table A33* in the Appendix provide detailed regression results.

taset or they are dropped from the estimation because of small numbers of observation. In this case, the box is white. The color of the box does not depend on the significance level of the coefficients. For example, if the coefficients in the two datasets work in the same direction, but the coefficient in the SOEP is significant at the 10% level and the SUF VVL 2004 coefficient at the 1% level, the box is still highlighted in green. This is because significance levels are largely a matter of case numbers. Almost all coefficients based VVL data are significant at the 1% level, which is a direct consequence of the large number of observations.

Table 31 Comparison of SOEP and SUF VVL 2004 with respect to Direction and Significance of Regression Coefficients

Dependent Variable	(1) Total	(2) Total West	(3) Total East	(4) Total Men	(5) Total Women	(6) Men West	(7) Men East	(8) Women West	(9) Women East
Years in School (topcode)	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+/** +/***	+/*** +/***	+/*** +/***	+ +/***	+/*** +/***
Years in Training (topcode)	+/*** +/***	+/*** +/***	+/*** +/***	+/* +/***	+/*** +/***	+/* +/***	+ +/***	+/** +/***	+/*** +/***
Years in Employment	+ +/***	+ +/***	+/** +/***	+ +/***	- +/***	+ +/***	+/** +/***	- +/***	+ +/***
Years in Unemployment	-/*** +	-/*** -	-/* -/*	-/*** +/***	-/*** +/***	-/*** +/**	-/** -/***	-/*** +	+ +/***
Years in Homeproduction	-/*** +/***	-/*** +/***	-/* +/***		-/*** +/***			-/*** +/**	- +/***
Retired (Dummy)	- +/***	- +/***	+ +/***	-/* +/***	+ +/***	-/* +/***	+ +/***	- +/***	+ +/***
Other (Dummy)	-/** +/***	-/** +/***	- +/***	-/** +/***	-/** +/***	-/* +/***	Dropped +/***	-/** +/***	+ +/***
Missing (Dummy)	-/* -/***	- -/***	-/** -/***	-/* -/***	- -/***	- -/***	- -/***	- -/***	-/*** -/***
Receives Pension for Civil Servants (Dummy)	-/*** n/a	-/*** n/a	- n/a	-/*** n/a	dropped n/a	-/*** n/a	- n/a	dropped n/a	dropped n/a
Receives Private Pension (Dummy)	-/* n/a	-/* n/a	- n/a	-/** n/a	+ n/a	-/** n/a	- n/a	+ n/a	+ n/a
Worked as Civil Servant/ Self Employed (Dummy)	-/*** n/a	-/*** n/a	-/** n/a	-/*** n/a	-/*** n/a	-/*** n/a	-/** n/a	-/** n/a	+ n/a
Educational Attainment: low (Reference Category: medium)	- -/***	- -/***	- -/***	-/* -/***	- -/***	- -/***	- -/***	- -/***	- -/***
Educational Attainment: high (Reference Category: medium)	+/** +/***	+/** +/***	+ +/***	+ +/***	+ +/***	+/* +/***	+ +/***	- +/***	+ +/***
Educational Attainment: missing (Reference Category: medium)	+ -/***	- -/***	+/*** -/***	-/ -/***	+/** -/***	- -/***	+ -/***	- -/***	+/*** -/***
Educational Attainment: unknown (Reference Category: medium)	-/** -/***	- -/***	- -/***	-/*** -/***	+ -/***	-/*** -/***	dropped -/***	- -/***	- -/***
Sex (female=1)	-/*** -/***	-/*** -/***	-/*** -/***						
West (west=1)	+/*** +/***			+/*** +/***	- +/***				

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's illustration. Upper left cells: SOEP results; lower right cells SUF VVL 2004 results. Significance levels: * p< 0.05, ** p < 0.01, *** p < 0.001. + indicates positive regression coefficient; - indicates negative regression coefficient. Red cells: contradiction in regression coefficients; green cells: consistency in regression coefficients.

3.7.1 Discussion

The majority of boxes in *Table 31* are green, indicating that the explanatory variables work in the same direction in both datasets and coefficients point in the intuitive direction. The same is true for the constant that is positive and highly significant across all models. We find pronounced differences for the coefficients of the following variables (red boxes): years in unemployment, years in home production, retired and other, as well as for the educational attainment category missing. In what follows, we discuss the reasons for these inconsistencies and search for better functional equivalents in the two datasets.

3.7.1.1 Years in Home Production

Inconsistencies in the variables home production are the results of variables not measuring the same thing in both datasets. In the SUF VVL 2004, years in home production only refer to pension-relevant periods, such as child-care periods or child-care credits (*Kinderberücksichtigungszeiten* or *Kindererziehungszeiten*). Today, a parent – typically the mother of the child – receives three years of child-care credits and another seven years of creditable periods per child. If a person opts to stay at home thereafter, the variable home production does not cover this time. Instead, if no other pension-relevant state applies, the respective period will be coded as a missing. Furthermore, we need to consider the priority rules that were applied in the process of data preparation. If two pension-relevant activities overlap, the priority rules decide which type of activity actually shows in the data. Given that child-care periods have the lowest overall priority (cp. to Section 3.5.2), we only observe them if no other pension-relevant circumstance applies.

In SOEP data years in home production can cover all those periods in which a person stayed at home to manage the household or care for children, irrespective of whether or not these periods are pension-relevant.

We tried to take this difference into account by considering home production only if a person reported no other activity in a given year. Obviously, this type of alignment did not work. Therefore, we have to find better functional equivalents in both datasets. One promising approach is to combine the variables years missing and years in home production. Since the VVL variable on home production only refers to pension-relevant periods, we apply the same rules to the SOEP variable. The applicable pension rules for the group of first-time retirees determine the rules by which we transform the SOEP variable. Women receive one year of child-care credits for all children born before January 1st 1992 (§ 56 SGB VI). For all children born thereafter, women receive three years of child-care credits. Based on this information, we construct a new variable for home production that depends on the number of children. Consequently, a mother of three receives three years of child-care credits.[120] Equivalently, a mother with one child receives one year of child-care credits. For the difference between the original variable for home production and the new one, we follow the logic in the administrative records. Given that in the SUF VVL 2004 a month equals missing if no pension-relevant activity applies, we do the same in SOEP data and set the difference to missing in order to obtain functional equivalents in both datasets.[121]

3.7.1.2 Years in Unemployment

The regression results also reveal inconsistencies in the variable years in unemployment. In the SUF VVL 2004, the variable only represents peri-

---

[120] We assume that all child-care periods are credited to the pension account of the mother. In this instance, we deviate from the SUF VVL 2004.

[121] It is not feasible to take non-contributory periods (*Berücksichtigungszeiten für Kindererziehungszeiten*) into account. These periods serve to close gaps in the insurance history but do not have an increasing effect on the monthly public pension benefit (§ 57 SGB VI) There is no straightforward solution to how many years of non-contributory periods are considered per child. The maximum is 10 years. However, these non-contributory periods only apply if there is no other pension-relevant circumstance (e.g. periods of employment that are subject to social security contributions).

ods of registered unemployment in which the unemployed person receives benefits according to Book 2 and 3 of the Social Code (§58 Abs. 3 SGB VI). In this case, the Federal Employment Agency pays substitute contributions to the statutory pension insurance that constitute pension entitlements. In turn, SOEP respondents report all different forms of unemployment - whether registered or not. Also periods of unemployment that went unnoticed by the statutory pension insurance.[122] There are no functional equivalents for the variable years in unemployment in both datasets. It is impossible to control for the problem of the hidden labor force. Despite this imperfection the variable years in unemployment enters the statistical matching.

3.7.1.3 Other

It is difficult to find a straightforward explanation for the discrepancies in the variable other. The inconsistencies might indicate that the variable captures completely different circumstances in both datasets. The SUF VVL 2004 measure summarizes three SES activities: care giving, invalidity and sickness as well as other activities (cp. to Table 15). The variable other captures compulsory contributions (in the case of care giving or sickness and invalidity), voluntary contributions and creditable periods, which explain the consistently strong positive coefficient across all models based on the administrative data (Stegmann 2006a, p. 547). Let us illustrate the positive relationship between the variable other and the monthly public pension benefit by means of periods of invalidity and sickness. During the first six weeks, a sick person is eligible for the continuation of payment (*Lohnfortzahlung im Krankheitsfall*) of prior earnings if the person worked in the position for more than four weeks (§ 3 EntgFG).[123] In this case, em-

---

122 Persons who are unemployed but not officially registered as unemployed are often referred to as the *hidden labor force* (or *Stille Reserve*). For an encompassing overview of this phenomenon in the German labor market see (Holst 2000).

123 EntgFG stands for Gesetz über die Zahlung des Arbeitsentgelts an Feiertagen und im

ployers and employees continue to pay contributions to the statutory pension insurance as if the person is regularly employed.[124] If a person continues to be sick after six weeks, he or she receives sickness allowances (*Krankengeld*). In this case, the health insurance provider pays the contributions.[125] These substitute contributions have the same increasing effect in terms of pension benefits as contributions from regular employment.

The SOEP measure for the variable *other* is inherently different from the one in SUF VVL 2004 data, because it covers quite heterogeneous types of activities, such as being on maternity leave, traveling around the world, or being incarcerated. Obviously, these activities have no increasing effect on the level of pension benefits. These differences explain why the direction of coefficients in both datasets diverges. For the sake of finding functional equivalents, we redefine the variable *other*. Because of the increasing effect periods of sickness and invalidity as well as other activities have on the level of pension benefits; we treat them as an equivalent to regular employment. From now on, we assign these activities to the category employment subject to social insurance contributions. The category care remains in the variable other.

3.7.1.4 Retired

The coefficient for the variable retired is consistently positive and highly significant across all VVL models.[126] Intuitively, the coefficient should be negative indicating a decreasing effect for the monthly public pension benefit. This intuition appears plausible because of the emphasis the Ger-

---

Krankheitsfall.

124 The level of contributions to be paid depends on prior earnings.

125 The sick allowance can be paid for up to 78 weeks within a period of three years. The level of contributions equals 80% of the contributions paid when the person received the continuation of payment.

126 As a reminder, the variable *retired* is coded with 1 if a person has more than four years of retirement.

man pay-as-you-go system puts on employment. It is quite likely that observations with longer episodes of retirement previously received disability benefits. In German pension legislation, the time a person spends receiving disability pension benefits counts as a *creditable period* (§ 58 Abs. 1 Ziff. 5 SGB VI). If a disability pension becomes eligible for an old-age public pension benefit (*Altersrente*) these creditable periods translate into contribution periods (§ 71 Abs. 1 & 2 SGB VI). For this translation, the statutory pension insurance extrapolates the employment history to age 65. The extrapolation is based on the previous employment history and earnings or the so-called total evaluation of contributions (*Gesamtleistungsbewertung*). Hence, if the employment history was continuous and earnings were high prior to being disabled, the total evaluation of contributions for a person is quite favorable. In fact, times in disability can then lead to an increase in pension benefits.[127]

In the SOEP, there are several explanations for what is captured in the variable retired. First, it might capture the receipt of disability benefits. Alternatively, it might reflect partial retirement agreements (*Altersteilzeit* or *Vorruhestand*). Elderly employees in partial retirement can negotiate with their employer to work only part-time after reaching a certain age and then slowly phase into retirement.[128] Ideally, the employee should spend the last five years of his career working part-time. However, most employees prefer the so-called block model. They spend 2.5 years working full-time and then 2.5 years in full retirement. In the official statistics, employees in partial retirement count as being employed. We do not know

[127] Persons with more than four years in retirement accumulated on average 42 earning points compared to 31 earnings points for persons who spent less than four years in retirement

[128] Employers and employees have a mutual interest in partial retirement, even though the motives differ quite clearly. For employers, partial retirement is a way to rejuvenate the workforce, whereas for employees, it is an alternative to early retirement that circumvents costly actuarial adjustments (Brenke 2007; Hoffmann 2007).

whether SOEP respondents perceive themselves as retired or employed. The inconsistencies in the results might be due to differences in the underlying concepts of the variable *retired*. Unfortunately, there is no self-evident solution in constructing functional equivalents. Based on the discussed modifications, we run the next set of regressions.[129] *Table 32* illustrates whether the modifications served the purpose of streamlining both datasets.

3.7.2 Effectiveness of Modifications

The reduced number of red cells in *Table 32* indicates that the modifications were partly successful. The discrepancies for the variable retired and missing were more or less resolved, but discrepancies persist in the variables home production, unemployment, and educational attainment missing. The number of years in home production and unemployment are important predictors for the level of pension benefits. They should not be left of the statistical matching. The situation is different for the variable educational attainment. The statistically significant impact of the dummy *educational attainment missing* indicates that the missing values are not missing at random. Plus, it is not clear, whether the non-missing information is reliable. Given that an imputation of missing values in the educational attainment variable is impossible, this variable will no longer be considered in the statistical matching.

---

129 *Table A33, Table 34*, and *Table A35* provide detailed results of the modified regression estimations.

Table 32 Comparison of SOEP and SUF VVL 2004 with respect to Direction and Significance of Modified Regression Coefficients

Dependent Variable Monthly Public Pension Benefit	**(1) Total**	**(2) Total West**	**(3) Total East**	**(4) Total Men**	**(5) Total Women**	**(6) Men West**	**(7) Men East**	**(8) Women West**	**(9) Women East**
Years in School (topcode)	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+*** +/***	+/*** +/***
Years in Training (topcode)	+/*** +/***	+/*** +/***	+/*** +/***	+ +/***	+/*** +/***	+ +/***	+ +/***	+/*** +/***	+/*** +/***
Years in Employment	+/*** +/***	+/*** +/***	+/*** +/***	+ +/***	+/*** +/***	+ +/***	+/** +/***	+/*** +/***	+/*** +/***
Years in Unemployment	-/*** +***	-/*** +/***	- +/***	-/*** +/***	- +/***	-/*** +/***	-/** +/***	- +/***	+ +/***
Years in Homeproduction	-/*** +/***	-/* +/***	+ +/***		-/* +/***			-/*** +/**	+ +
Retired (Dummy)	+ +/***	+ +/***	+/** +/***	-/* +/***	+/** +/***	-/* +/***	+ +/***	+ +/***	+/* +/*
Other (Dummy)	-/** -/***	-/** -/***	- -/***	-/** -	-/** -/***	-/* -/*	Dropped -	-/** -/***	+ -/***
Missing (Dummy)	- -/***	- -/***	-/** -/***	-/* -/***	- -/***	- -/***	- -/***	- -/***	-/* -/***
Worked as Civil Servant/ Self Employed (Dummy)	-/*** n/a	-/*** n/a	-/** n/a	-/*** n/a	-/*** n/a	-/*** n/a	-/** n/a	-/** n/a	+ n/a
Educational Attainment: low (Reference Category: medium)	-/* -/***	- -/***	- -/***	- -/***	- -/***	- -/***	- -/***	- -/***	- -/***
Educational Attainment: high (Reference Category: medium)	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+/*** +/***	+/** +/***	+/*** +/***	+/* +/***	+/*** +/***
Educational Attainment: missing (Reference Category: medium)	- -/***	- -/***	+ -/***	- -/***	+ -/***	- -/***	+ -/***	- -/***	Dropped -/***
Educational Attainment: unknown (Reference Category: medium)	-/* -/***	- -/***	- -/***	- -/***	- -/***	- -/***	dropped -/***	- -/***	Dropped -/***
Sex (female=1)	-/*** -/***	-/*** -/***	-/*** -/***						
West (west=1)	+/*** +/***			+/*** +/***	+ +/***				

Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's illustration. Upper left cells: SOEP results; lower right cells SUF VVL 2004 results. Significance levels: * p< 0.05, ** p < 0.01, *** p < 0.001. + indicates positive regression coefficient; - indicates negative regression coefficient. Red cells: contradiction in regression coefficients; green cells: consistency in regression coefficients.

3.8 Testing the Feasibility of a Statistical Matching

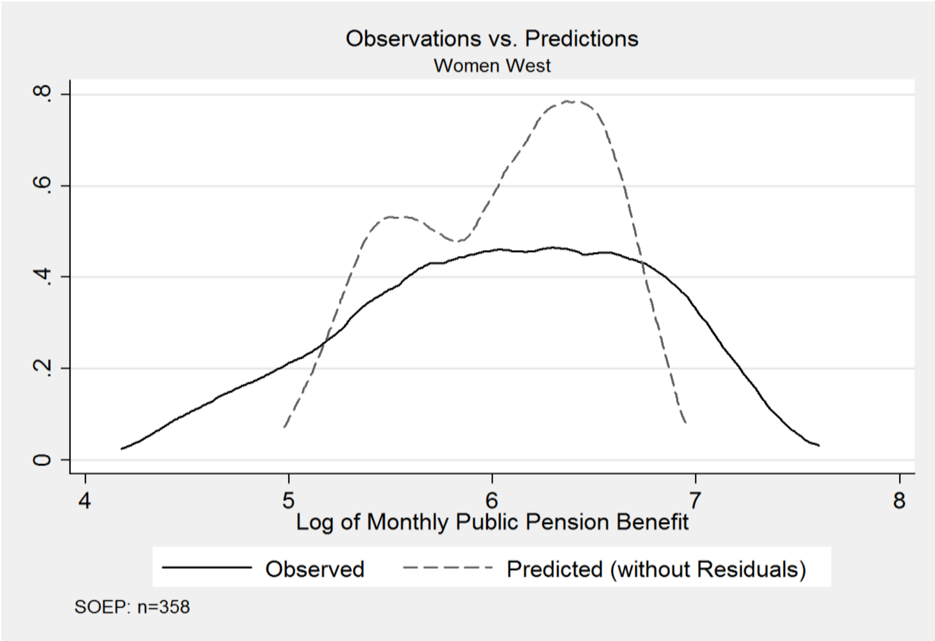
After successfully aligning the sample populations and identifying matching variables, this last step tests the feasibility of a statistical matching of population representative survey data from the SOEP and administrative pension records, namely the SUF VVL 2004. The underlying idea of this feasibility study is to estimate regression models based on SOEP data with the monthly public pension benefit being the dependent variable. We then apply the estimated SOEP coefficients to SUF VVL 2004 data in order to make out-of sample predictions. If it is possible to replicate the SOEP results with out-of-sample predictions based on SUF VVL 2004 data, we consider the preparatory steps for a statistical matching of the two datasets to be successful.

3.8.1 In-Sample Predictions

Using the matching variables identified in Section 3.7, we want to test how well our regression models predict the logged monthly pension benefit. Ideally, the predicted values $\hat{y}_i$ come as close as possible to our observed values y_i. The quality of the predictions would be highest if all the variance in the dependent variable was explained by our model. *Figure 10* compares the observed and predicted benefits for West German women based on SOEP data (*in-sample prediction*).[130] Obviously, the predicted values deviate quite clearly from the observed pension benefits.

[130] We performed the following analysis for the total sample population and eight subsamples (cp. to *Table 27*). We illustrate our procedure for West German women only. Results for the other samples are available *Figures A2-A12* in the Appendix.

Figure 10 Example for In-Sample Predictions, Women West

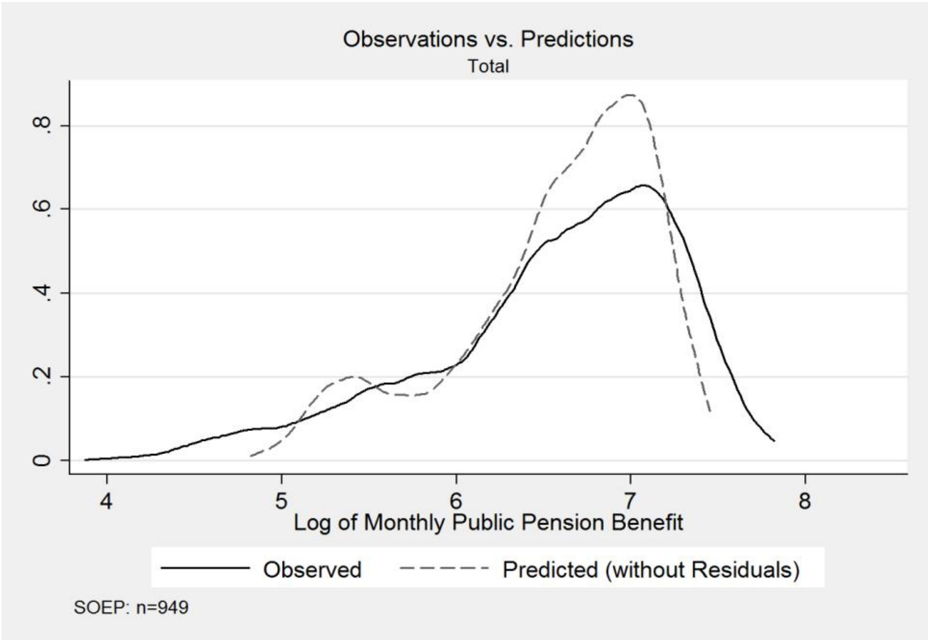


Source: SOEP 2005; Author's calculations

Figure 10 shows that the distribution of the observed logged public pension benefit is centered on a mean value of 6 (roughly 400 Euro). The distribution of the predicted value shows two peaks: one at the log value 5.3 (roughly 200 Euro) and the second at 6.3 (roughly 550 Euro). The model appears to be not well-suited for the prediction of the smallest and largest values in the distribution. This lack of precision is a result of the *regression to the mean* effect (Copas 1997). The effect implies that predicted values tend to move closer to the observed sample mean than one might anticipate from the distribution of observed values. As a result, the distribution of the predicted values is shrinking, displaying a far smaller variance than the distribution of observed values. For the sample of West German women, the variance of observed benefits is 0.534 compared to a variance of 0.222 for the fitted values. According to Copas, the *regression to the mean* effect is linked closely to the goodness-of-fit statistics. The better the model's fit, the smaller the shrinkage effect. The problem is more pronounced in the case of small sample sizes and/or a high number of covariates.

In fact, the *shrinkage-* and *regression to the mean-* effect is not as pronounced in the larger subsamples (cp. to *Figure 11*). For the total population the correspondence of observed and predicted values is far more exact than for the smaller subsamples and the shrinkage effect is less distinct.[131] But also in the distribution of the total population, the smallest and largest values are less well predicted.

Figure 11 Example for In-Sample Prediction SOEP, Women West



Source: SOEP 2005, own calculations

3.8.2 Assigning Random Residuals

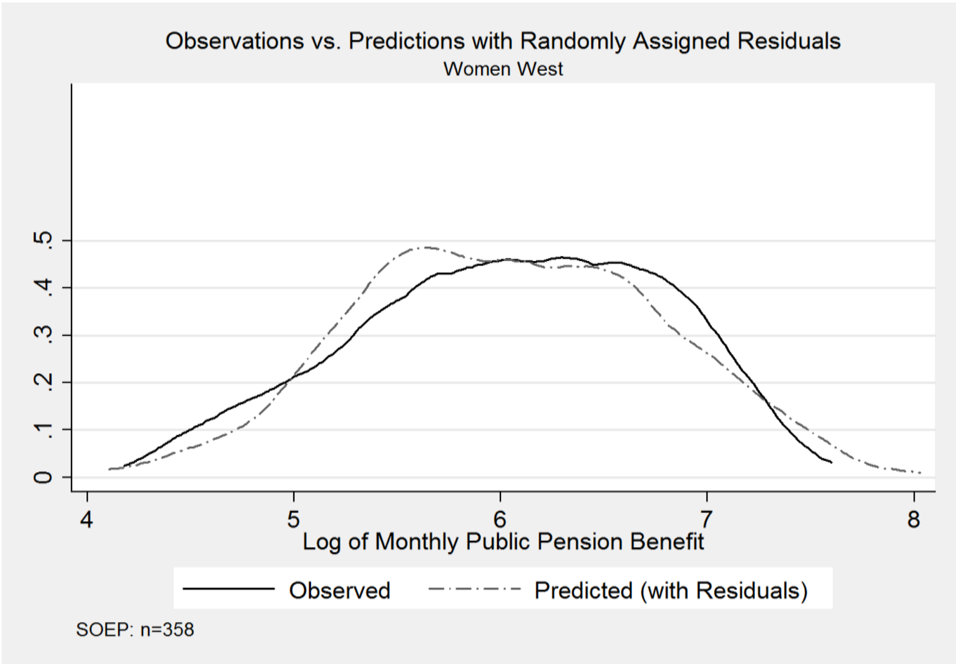
Ultimately, the quality of the statistical matching depends on the predictive power of the matching variables. If predictions systematically deviate from observations and if the variance is significantly smaller for the predicted than for the observed values, the statistical matching is likely to be biased. Assigning random residuals on top of the predictions recaptures the variance and prevents a bias in the statistical matching. In this paper,

---

[131] The total sample population has a sample size of n=949, the Women West sample size equals n=358. *Table 27* provides the sample sizes for all subsamples.

we apply the following procedure: For each pair of observations and predictions, we calculate the respective residual. If we add these residuals on top of the prediction, we obtain identical values for observations and predictions, which is not in our interest. In order to introduce some randomness, we assign each residual randomly to a new observation and add this residual on top of the prediction. Plotting the observations against the newly predicted values (including the randomly assigned residuals) illustrates whether the procedure was successful (cp. to *Figure 12*).

Figure 12 In-Sample Predictions with Randomly Assigned Residuals SOEP, Women West

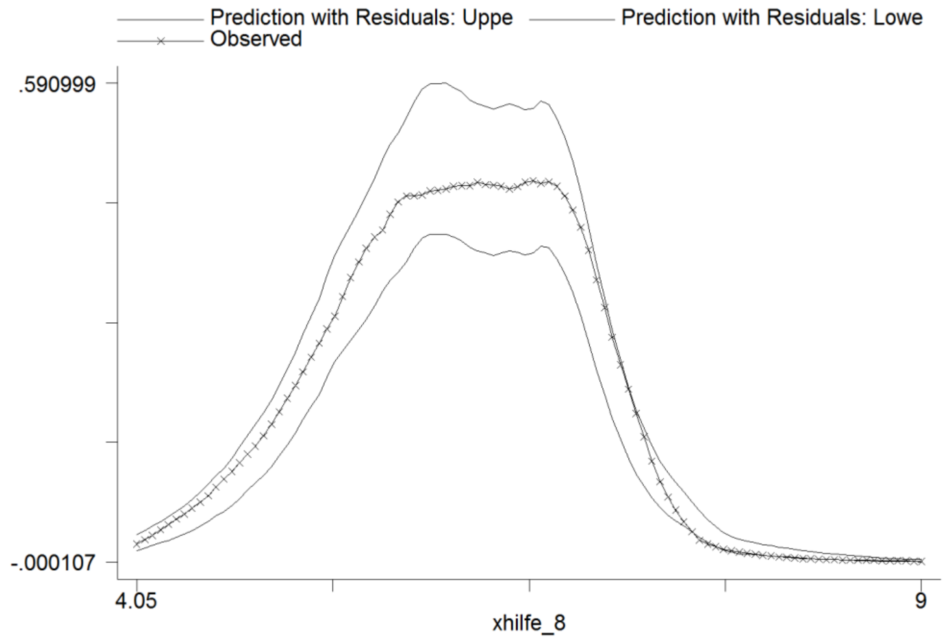


Source: SOEP 2005; Author's calculations

The comparison of *Figure 10* and *Figure 12* shows that the randomized assignment of residuals was successful. The procedure largely improves the accordance between observed and predicted values. The distribution of predicted values (dashed line) no longer has two peaks and shifted downwards to the distribution of observed values. The procedure also recaptures the variance in the distribution. The variance of the new predictions is 0.521, compared to 0.534 for the observations.

Finally, it was useful to control whether the observations and the predictions with randomly assigned residuals are significantly different from each other. Therefore, we plotted the distribution of observed values against the 95% confidence bands of the new predicted values and. If the two variables were significantly different from each other, the observed values would lie outside the 95% confidence bands. *Figure 13* depicts the results.

Figure 13 95% Confidence Bands for Predictions with Randomly Assigned Residuals, Women West



Source: SOEP 2005; Author's calculations

The solid lines represent the upper and lower limit of the 95% confidence band, whereas the dashed line represents the observed values for West German women. The observations lie perfectly within the confidence bands. Hence, observations and modified predictions are not significantly different from each other.

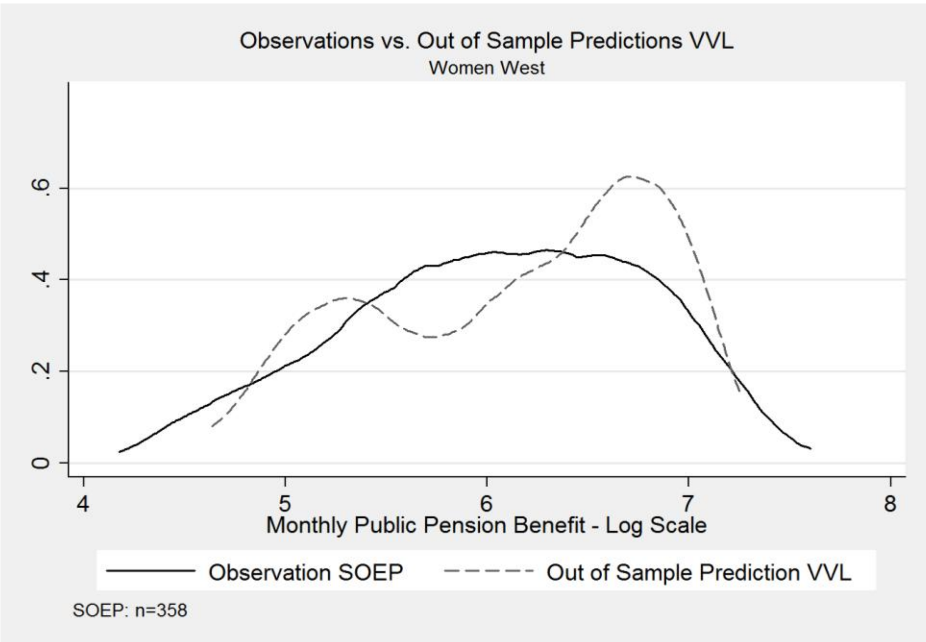
3.8.3 Out-Of-Sample Predictions

This last steps tests, whether it is possible to replicate the results based the SUF VVL 2004 with SOEP data. This test involves an out-of-sample prediction. The out-of-sample prediction takes coefficients estimated based on the SUF VVL 2004 and applies these coefficients to SOEP data. We choose to perform the replication exercise going from the administrative records to survey data, because the SUF VVL 2004 estimates are much more robust than those estimated on basis of the SOEP. The following equation shows how to calculate the out-of-sample predictions, where α is the estimated constant and β_1 to β_{12} are the estimated coefficients of the regression model based on SUF VVL 2004 data. These coefficients were then multiplied by the respective individual values (x_1 through x_{12}) in the SOEP.

$$\hat{y}_{SOEP\ Out-of-Sample} = \alpha_{VVL} + \beta_{1_{VVL}} * Years\ in\ School\ (x_1)_{SOEP} + \beta_{2_{VVL}} * Years\ in\ Training\ (x_2)_{SOEP} + \ldots + \beta_{12_{VVL}} * Education\ High\ (x_{12})_{SOEP}$$

Figure 14 illustrates the out-of-sample predictions for West German women comparing the distribution of SOEP observations with out-of-sample predictions based on coefficients estimated in the SUF VVL 2004.

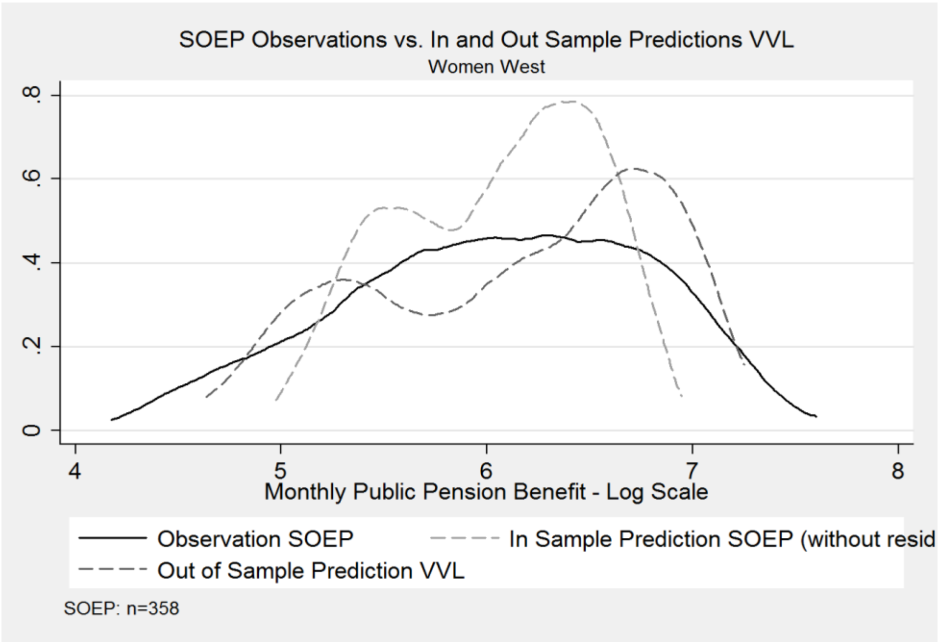
Figure 14 Example for Out-of-Sample Prediction, Women West



Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

Compared to the first in-sample-prediction in *Figure 10*, the out-of-sample prediction for West German women deviates quite clearly from the distribution of observations. The distribution of out-of-sample prediction also has two peaks: one at a log value of 5.3 (approximately 200 Euro) and the other at 6.8 (approximately 900 Euro). Compared to *Figure 10* the second peak shifts slightly to the right. The *regression to the mean* effect is less pronounced in the out-of-sample prediction. Hence, it performs better in the prediction of the smallest and largest values in the distribution. *Figure 15* confronts the distribution of the observations with the in- and out-of-sample predictions.[132]

Figure 15 Comparison of Observations and In and Out of Sample Prediction, Women West



Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

The graph shows that the shrinkage effect is less distinct in the out-of-sample prediction relative to the in-sample prediction. The variance of the out-of-sample prediction is 0.494, compared to a variance of 0.534 for the observations. As a reminder, the variance of the original in-sample prediction is 0.222. One explanation for the less distinct shrinkage effect might be the more robust SUF VVL 2004 coefficients, which we applied to the

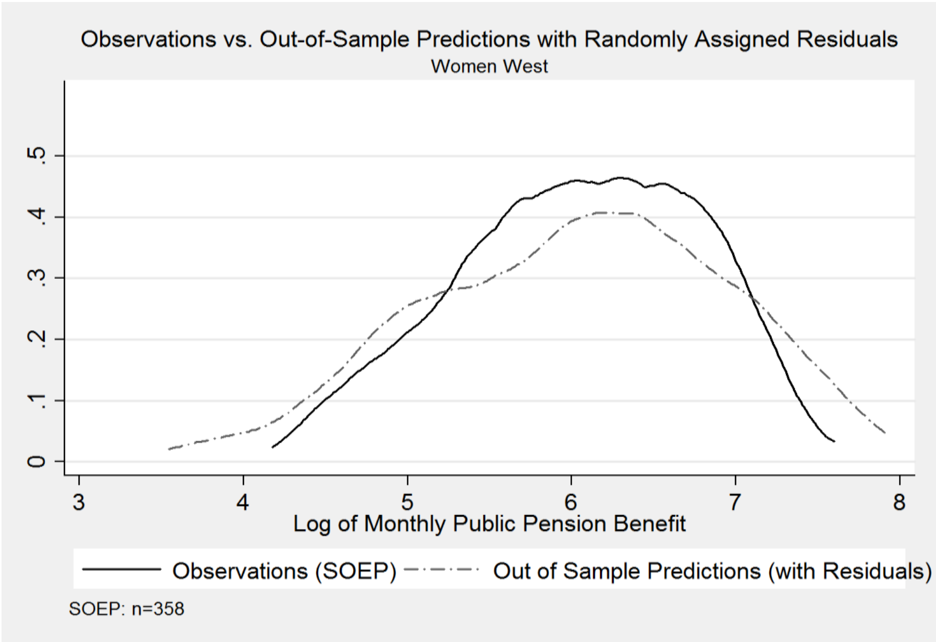
---

132 The out-of-sample predictions for the other demographic groups are illustrated in *Figure A2* to *Figure A12*.

SOEP.

We apply the same procedure as before and assign the residuals randomly to our out-of-sample predictions. In this instance, the residual is the difference between the observation and the respective out-of-sample prediction. *Figure 16* shows how the results improve after assigning random residuals to the out-of-sample prediction.

Figure 16 Out-of-Sample Predictions with Randomly Assigned Residuals SOEP, Women West

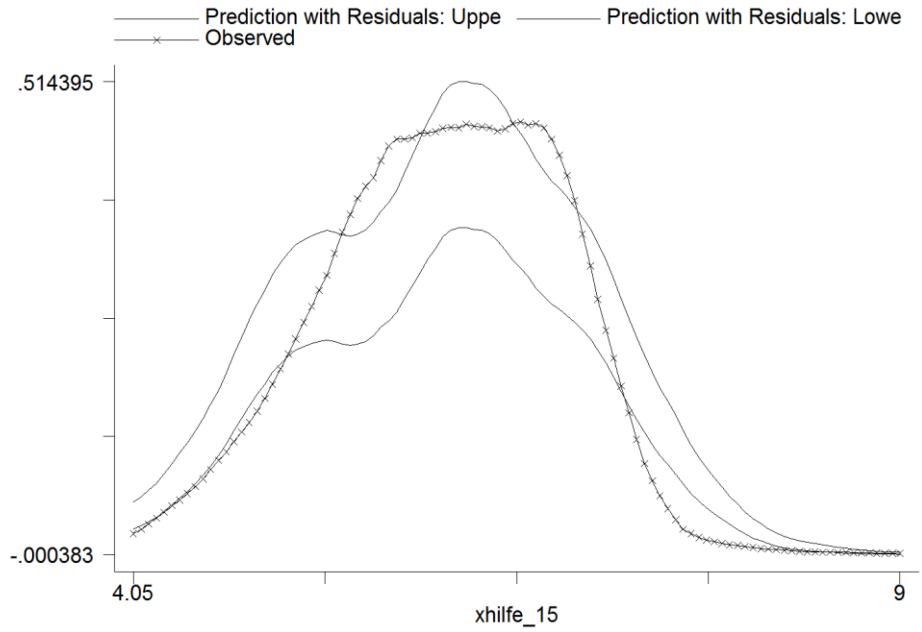


Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

Again, applying the proposed procedure largely improves the accordance between the distribution of observed and predicted values. The distribution of out-of-sample predictions with randomly assigned residuals (dashed grey line) no longer has two peaks and shifted below the distribution of observed values. It has a much larger variance than the observations, because we obtain more extreme predictions for smaller values as well as for larger values.

Plotting a 95% confidence interval controls whether the out-of-sample predictions are significantly different from the observations (cp. to Figure 17).

Figure 17 95% Confidence Bands for Out-of-Sample Predictions with Random Residuals



Source: FDZ-RV - SUFVVL2004 & SOEP 2005; Author's calculations

The results for the out-of-sample predictions wit randomly assigned residuals are not as satisfying as for the in-sample predictions (cp. to *Figure 13*). In some segments of the distribution, observations lie outside of the confidence band, which implies that observations differ significantly from out-of-sample predictions. However, for the most part, the observations lie within the confidence bands of the out-of-sample predictions. The results of the out-of-sample predictions underline that it is possible to replicate SOEP results with the SUF VVL 2004 out-of-sample predictions.

3.9 Conclusion and Outlook

This paper presents a feasibility study for a statistical matching of administrative pension records, namely the dataset *Completed Insurance Biographies 2004* maintained by the statutory pension insurance and population representative survey data from the *Socio-Economic Panel*. These preparatory steps for a statistical matching involve the correct specification of the analysis population in both datasets and the identification of potential matching variables. These first steps are crucial, because data are not representative for the same total population and presumably similar variables do not necessarily measure the same. For this reason, we thoroughly compare the distributions of potential matching variables and align them if necessary. With the distribution of variables being sufficiently similar, we run separate multivariate regression models for each dataset and for different demographic groups with the monthly public pension benefit as the dependent variable. Ideally, the estimated coefficients of the independent variables (e.g. the matching variables) point in the same direction. If they do, the variables work as matching variables and are considered in the ultimate step of the feasibility study. This last step checks, whether it is possible to replicate SOEP results with out-of-sample predictions based on SUF VVL 2004 data.

This feasibility study gave us not only a thorough understanding of the compatibility of both datasets, but also encouraging results that motivate the implementation of the actual statistical matching of administrative pension records and population representative survey data. For the implementation, we turn to the Sample of Active Pension Accounts *(Versicherungskontenstichprobe* - henceforth, SAPA) another dataset maintained by the statutory pension insurance that does not cover a sample of recent retirees, but represents a sample of individuals holding a pension account. This sample allows us to analyze the social security wealth of future retirees, especially those cohorts who will fully experience the effects of pension reforms and the repercussions of changing employment trajectories.

From this study we know, how to specify the sample populations and understand which variables work as matching variables and which don't. Our knowledge is limited when it comes to finding the best matching technique for the data at hand. In the next paper, we will compare the performance of different matching and imputation approaches.

4 Statistical Matching of Administrative and Survey Data – An Application to Wealth Inequality Analysis[133]

4.1 Introduction

Since the early 1980s, the popularity of statistical matching has been steadily increasing not only among sociologists and economists (Heckman et al. 1997; Rosenbaum and Rubin 1983), but also in other disciplines such as medicine (Baiocchi et al. 2010). In the social sciences, studies based on statistical matching typically evaluate the efficacy of government policies and programs (Dehejia and Wahba 1999). In these applications, researchers face the fundamental problem of not knowing how a treated individual (e.g. a person who participated in a certain program) would have fared if he or she had not received the treatment (Arceneaux et al. 2010; Morgan and Harding 2006). Statistical matching helps to overcome this constraint by comparing the difference in outcomes between individuals who receive a certain treatment and those who don't, but share a set of common background characteristics (Caliendo and Kopeinig 2008).

Statistical matching also comes into play when required data are not available in one, but several datasets that cannot be linked over a unique identifier (Kadane [1978] 2001; Moriarity and Scheuren 2001; Rodgers 1984).[134]

---

133 This paper is joint work with Markus M. Grabka, and Joachim Frick. Anika Rasner is the first author.

134 So-called *record linkage* requires a unique identifier and the informed consent of survey respondents, because it links identical persons in two datasets. A 2009 pilot study tested the willingness of SHARE respondents in Germany to allow for record linkage using their Social Security Number (SSN). While 77 percent gave their consent, only 64 percent of those respondents provided their SSN. It is not yet verified whether the SSN provided is always correct. This outcome implies that less than 50 percent of respondents participating in the pilot study agreed to record linkage. There is good reason to believe that there are systematic differences between consenters and non-consenters that would add bias to the results. For literature on consent patterns see Jenkins, Cappellari, Lynn, Jäckle, and Sala (2006).

This problem occurs if mounting a new survey is too costly or if survey respondents are unable to provide reliable information. In these situations, statistical matching concatenates records from one dataset with needed information from the second source (Rässler 2002; Rubin 1986). The linked information is not from the same, but from observations with identical (or nearly identical) background attributes in both datasets (Elliott and Davis 2005). However, statistical matching is not always the best strategy to complement one dataset with information from another. Alternatively, one can turn to imputation techniques to add relevant data (Little 1988; Rässler et al. 2008).

This paper compares four statistical matching and imputation techniques to complement survey data on wealth from the population representative German Socio-Economic Panel (SOEP) with information on social security wealth (SSW) from the Sample of Active Pension Accounts (SAPA), an administrative dataset maintained by the German statutory pension insurance.[135] The unique properties of the matched data provide an effective control for the quality of matches under each matching and imputation strategy that allows the authors to assess which technique fares best for the data at hand. An additional link to divorce statistics that administer the pension rights splitting between divorcees controls for otherwise unconsidered effects that arise from a marital split.

The newly assembled data opens up countless research directions. We exploit the data and perform a wealth inequality analysis that includes SSW. So far, SSW was omitted in wealth analyses, because of the lack of adequate micro data. While the SOEP collects extensive wealth data, information on SSW is difficult to obtain because respondents typically lack

[135] Record linkage is infeasible for data confidentiality reasons. Moreover, no unique identifier is available in both datasets.

knowledge of their pension entitlements.[136] In contrast, the statutory pension insurance keeps data on SSW, but lacks information on standard wealth categories and covariates that go beyond the purpose of the agency. Especially in comparative analyses, the omission of SSW from wealth estimates has so far raised issues of comparability (Frick and Headey 2009). Including SSW helps us to draw a more precise picture of the distribution of wealth in Germany.

The remainder of this paper is organized as follows: Section 2 sets out some background information on the system of old age provision in Germany and the concept of SSW. Section 3 describes the data and section 4 presents the overall matching strategy. Section 5 contrasts four statistical matching and imputation techniques and compares their performance for the groups of retirees. The technique that serves our purpose best is then applied to the total population providing the basis for the extended wealth measure that includes SSW. Section 6 presents the results of the wealth inequality analyses and finally section 7 closes with some concluding remarks.

4.2 The System of Old Age Provision in Germany

Distinctive institutional features account for differences in old age provision across occupational groups in Germany. The statutory pension insurance covers the majority of the German resident population but systematically excludes certain groups, such as civil servants or the self-employed. To appreciate differences in coverage and the accumulation of SSW across occupational groups, this section briefly sketches the system of old age provision in Germany and highlights the consequences for the

[136] In 2006, the SOEP group performed a pre-test asking for the person's SSW. The question generated more than 92% missing values, the reliability of information given by the remainder of respondents are questionable.

statistical matching and imputation exercise.

The statutory pension insurance is compulsory and by far the most important pillar in the provision of retirement benefits in Germany. Throughout their adult working lives, more than ninety percent of the population gets in touch with the public pension scheme for at least once. Today, the scheme covers more than 35 million actively insured individuals and pays benefits to almost 25 million retirees (Deutsche Rentenversicherung 2010).[137] Benefits from the public pension scheme are still the predominant source of retirement income for this group (Kortmann and Halbherr 2008), with occupational and private pensions clearly playing a secondary role.[138] For employees the accumulation of SSW usually starts with the first job that is subject to social insurance contributions and ends with the transition into retirement. In these type of jobs, employees pay contributions into the social security system, a certain fixed share of their earnings up to some maximum amount. The employer matches these payments.[139] By paying contributions into the system, employees accumulate SSW. Individuals can also accrue pension rights for certain periods of non-employment, such as spells of education, unemployment or sickness. And finally, divorce affects the individual's SSW as well. The statutory pension insurance carries out a splitting of pension rights accrued by husband and wife during their marriage (cp. to Section 4.4.4 for further details).[140] Taking all these factors into account,

---

[137] Actively insured persons have at least one period of paid contributions (payment of compulsory or voluntary social insurance contributions, marginal employment) or creditable periods.

[138] In occupational pension schemes individuals can also accumulate SSW. These schemes are typically not compulsory and vary with respect to the replacement rate.

[139] For 2010, the contribution level is 19.9 percent paid in equal parts by employee and employer.

[140] The partner who earned higher pension rights transfers half of the difference in entitlements (*deduction*) to his/her former spouse (*premium*). In practice, women are the principal beneficiaries of pension splitting, because of their comparatively weaker labor market participation. For the majority of divorced couples the splitting takes

SSW corresponds roughly to the individual's earnings history and retirement benefits are a proxy for a person's life cycle labor market attachment.[141]

A separate non-contributory civil servants pension scheme covers the 1.8 million active civil servants and provides benefits to 680,000 former public sector employees and to 300,000 survivors in Germany (Bundesministerium für Arbeit und Soziales 2008). By definition, civil servants become part of the civil servants scheme from which they cannot and have no reason to opt out. Their SSW does not accumulate over the life cycle, but solely depends on the final salary before retirement and the years of service. With its generosity the German civil service pension scheme stands out in comparative perspective, mitigating any need for additional retirement income.[142] Typically, we don't observe civil servants in SAPA data. Exceptions are those individuals who were not civil servants from the beginning of their career. In this case, they have worked for a couple of years and accumulated SSW within the statutory pension insurance.

For the 4.5 million self-employed in Germany, the system of old-age provision is most heterogeneous. In fact, about 25 percent are permanently insured by compulsory schemes, such as farmers or self-employed in liberal professions, but to very unequal conditions in terms of coverage and the provision of benefits (Loose and Frommert 2009). While benefit levels for farmers are comparatively low, self-employed in the liberal professions

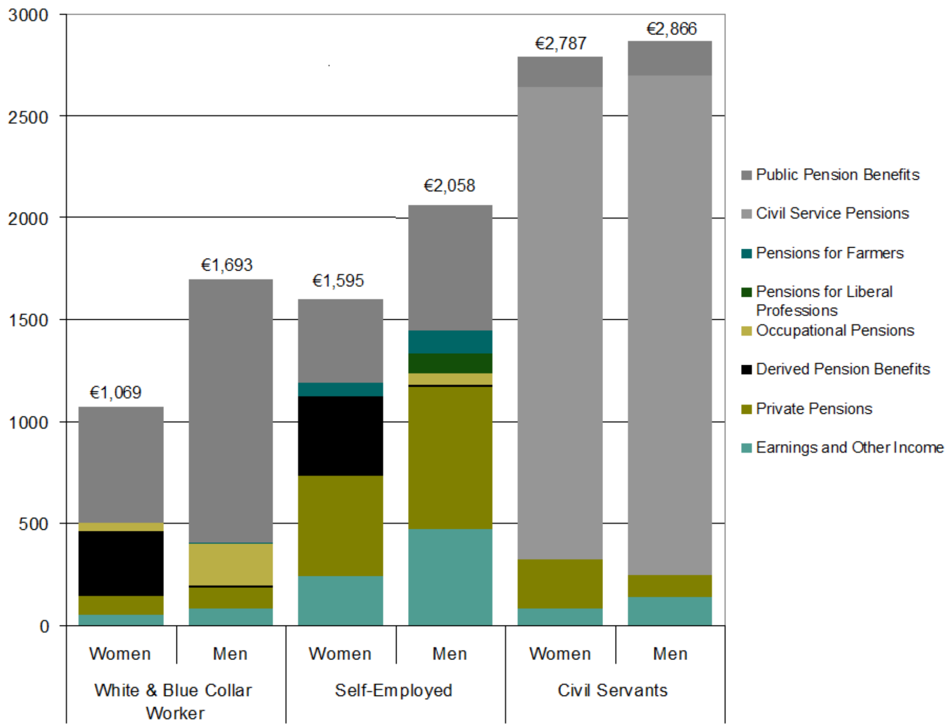
place right upon divorce, premium or deduction remain unaffected by remarriage of either ex-partner.

141 The principle of equivalence, one of the guiding principles of the public pension system, implies that benefits are roughly equivalent to contributions paid into the system.

142 The replacement rate of civil servants who retire after forty years of full-time employment amounts to 72 percent of their last gross earnings. In contrast, the replacement level for the standard retiree (worked 45 years with average earnings) reaches approximately 48 percent of previous earnings.

such as lawyers or medical doctors enjoy replacement rates comparable to those of civil servants. The rest of the self-employed lacks formal coverage: some rely exclusively on voluntary private pension investments, whereas others accumulate entitlements in several different schemes. The heterogeneity also has repercussions for the accumulation of SSW. For self-employed in compulsory schemes, SSW accumulates over the entire life cycle. The growing number of self-employed without employees typically has alternating spells of self- and dependent employment, which therefore also accumulate SSW in the public pension scheme.

Figure 18 Composition and Level of Monthly Old-Age Income Before Taxes for Men and Women Aged 65 and Older



Source: ASID 2007; Author's calculations

Figure 18 captures old-age provision across occupational groups. It illustrates differences in the level and composition of monthly old-age income for men and women aged 65 and older by occupational status.[143] For men, between-group differences in total income are very pronounced: on average, civil servants enjoy 1.7 times higher incomes than blue- and white-collar employees. Self-employed are somewhere in between. On average, they receive 22 percent higher income than workers, but 29 percent lower income than civil servants. Even more striking are differences in old-age income among women aged 65 and older. With an old-age income barely above € 1,000 per month female white- and blue-collar workers fare much worse than other women. On average, self-employed women have 1.4 time and civil servants enjoy 2.5 times higher old-age incomes than white- and blue collar workers. Self-employed women clearly lag behind civil servants in terms of their monthly old-age income as well. Also noticeable are differences in the composition of old-age income for men and for women. Civil servants receive more than 90 percent of their total income out of the civil service pension scheme. Workers draw a large share of their total income out of the public pension scheme complemented by income from occupational and private pensions. Female workers rely heavily on survivor's benefit. This income component accounts for a third of their total income. The composition of retirement income is most heterogeneous among formerly self-employed. They not only draw benefits from several different old-age schemes, but also receive a significant share of their income from earnings indicating that many continue to work above age 65.[144] Formerly self-employed women also rely on survivor's benefits amounting to 25 percent of their total income.

This brief glance at the system of social security in Germany illustrates the quantitative relevance of the public pension scheme for the majority

---

143 The categorization is based on information about the person's last occupational status prior to retirement.

144 For the group of self-employed, working above age 65 is not necessarily an economic exigency, but an intrinsic motivation.

of the active and retired population. Over the course of their working life, they pay a significant portion of their earnings into the system that cannot be invested in alternative forms of old age provision. But this sketch also shows pronounced differences across occupational lines. These differences require special diligence in the matching and imputation exercise.

4.3 Data

In this paper, we want to complement standard wealth categories with information on SSW as of 2007. Along with data capturing net worth[145] and SSW, the authors present an extended wealth measure for a more comprehensive analysis of wealth and inequality in Germany. For this purpose, the authors present a double-match involving three datasets. The first match employs 26 waves of panel data from the population representative German Socio-Economic Panel Study (SOEP) and links them with the Sample of Active Pension Accounts (SAPA). The SOEP is a broad interdisciplinary household panel study that started in 1984 (Wagner et al. 2007). It covers a representative sample of the total population living in private households in Germany. The most recent accessible data was collected in 2009 with about 11,000 households and 20,000 individuals being interviewed. The micro-data provide detailed information on individuals, households, and families, and enable researchers to monitor stability and change in living conditions over time. The standard components are surveyed year by year, whereas certain special topic modules are asked every few years. In 2007, a wealth module collected detailed wealth data at the individual level (for further details see Frick et al. 2007), except for information on SSW.

---

145 Net worth is the sum of owner-occupied and other real estate holdings, financial assets, assets from life insurance policies and private pension schemes, building loan contracts, business assets, valuables, net of any outstanding mortgage and consumer debt.

The SAPA is a one percent random sample of pension accounts, containing records for approximately 570,000 individuals, both actively insured and recently retired. These records are representative of all individuals holding a pension account.[146] SAPA contains demographic and detailed benefit information, including the individual's aggregated SSW as of 2007. The longitudinal files provide information on monthly earnings, unemployment spells, periods of child care and long-term care, etc. Unlike survey data, SAPA provides individual, but no household information, whatsoever.

The second match links data from the Divorce Statistics administering the Pension Rights Splitting between Divorcees (henceforth, Divorce Statistics) maintained by the statutory pension insurance to the matched data using record linkage.[147] The Divorce Statistics cover all divorce settlements – a total of 5.5 million cases – that involved a splitting of pension rights between ex-spouses since its introduction in 1977. Further, these statistics contain information about marriages and divorces that go beyond those available in SAPA.

4.4 First Match: Linking SOEP and SAPA

4.4.1 Notation and Conditional Independence

Whether we opt for statistical matching or imputation to link SOEP and SAPA, we face the following initial situation. There are two sample files A (SOEP) and B (SAPA) that share a set of common variables and some variables unique to each dataset. The background attributes observed in both datasets are referred to as X variables, $X = (X_1, ..., X_p)$. Variables

[146] A personal pension account is conditional on having at least one event over the life-course that constitutes rights in the statutory pension insurance.

[147] SAPA data and Divorce Statistics provide the identifiers for record linkage.

unique to the SOEP will be referred to as Y variables, $Y=(Y_1,\ldots,Y_Q)$, whereas variables unique to SAPA will be referred to as Z variables, $Z=(Z_1,\ldots,Z_R)$. By means of statistical matching or imputation, we complement Y-variables on net worth with the Z-variable on SSW using the common X-variables. The population representativeness of SOEP data dictates the matching direction: SOEP is the recipient and SAPA the donor file.[148]

We face an identification problem, because we don't observe net worth and SSW jointly and their covariance is unknown. For the results to be meaningful, the conditional independence assumption (CIA) has to hold, meaning that Y- and Z-variables are conditionally independent given the matching variables X (Caliendo and Kopeinig 2008).

4.4.2 Statistical Matching or Imputation: Four Alternatives

Simple *hot deck imputation* completes records with missing data points with values from statistically similar, but complete records (Andridge and Little 2010). Typically, both records – complete and incomplete – are part of the same dataset. In our application, SSW is missing for all SOEP, but available for all SAPA records, which corresponds to the setup for *cold deck* routines. In this paper, we pretend that sample file A and B originate from the same data. Therefore, we apply hot deck imputation to replace missing values of SSW in SOEP data with observed values in SAPA. The im-

---

[148] For reasons of sample selectivity, SAPA is not the adequate recipient file. SAPA is selective for several reasons: First, certain groups are systematically excluded from having a pension account in the statutory pension insurance. Second, account validation adds selectivity. The statutory pension insurance asks every insured person to confirm the information stored in their account, but they are not obliged to do so. However, some persons are more likely to validate than others. The authors restrict the analysis to validated pension accounts, which significantly reduces the sample size (336,069 instead of 568,586 observations). We accept the reduction in sample size, because validated accounts provide the most reliable information. For more information on account validation and selectivity see Rasner, Frick, and Grabka (2011).

putation is carried out within predefined matching strata using a number of common categorical *X*-variables.[149] Within these groups, missing values are imputed by random assignment.

Based on SAPA data, *regression-based imputation* estimates multivariate OLS regression models within different imputation classes. In these models, the individual's SSW is a linear function of the *X*-variables, the set of background attributes available in both datasets. Based on the estimates, the authors perform out-of-sample predictions of SSW, imputing the respective value for all SOEP observations. To mitigate the *regression to the mean effect* inherent in predictions, residuals are randomly assigned to the respective predictions to preserve the variance of the distribution (Copas 1997).

Strictly speaking, *univariate imputation sampling* (UVIS) also belongs to the group of hot deck imputation routines. *UVIS* differs from the above hot deck approach in that it combines parametric and non-parametric techniques to impute a single variable with missing values (SSW) using predictive mean matching. In a first step, *UVIS* makes use of a parametric model (OLS regression) that describes the individual's SSW as a function of all matching variables. In a second step, *UVIS* selects from all fully observed units the nearest neighbor donor that has the smallest distance with respect to the prediction for each incomplete observation (Chen and Shao 2000). Unlike regression imputation, *UVIS* imputes values observed in the distribution of SAPA but no predicted values (including random residuals).

---

149 Categorized continuous variables can also enter the imputation proceeding as a stratification variable. Any combination of stratification variables builds an imputation class or matching stratum. Both terms are used interchangeably. Cross-classification of a number of categorical variables can lead to many imputation classes, which possibly collides with an insufficient number of observations.

Statistical matching aims at finding statistically similar observations in both datasets. In this paper, we make use of the *Mahalanobis distance* (Mahalanobis 1936), a procedure frequently used in cluster analysis. The procedure calculates a Mahalanobis distance d_{ij} comparing each observation i_A in the SOEP to each observation j_B in SAPA based on the vector of common X-variables. The statistical donor minimizes the distance d_{ij} between the SOEP respondent and the SAPA observation. Unlike the Euclidean distance, the Mahalanobis distance score incorporates both correlations between matching variables and differences in variances. First, this implies that highly correlated matching variables do not enter the computation of the Mahalanobis distance with the same weight. We will show later why this property is useful in this application.

4.4.3 Matching Variables

The set of common X-variables distinguishes categorical *slice* and *continuous* matching variables.[150] Slice variables partition the data to only match or impute individuals within certain predefined strata. The partitioning avoids matches of individuals that are sufficiently dissimilar, especially if these groups are believed to differ in how they accumulate SSW. Matching variables determine the best imputation or matching partners in SOEP and SAPA. The statistical matching and imputation occurs within six imputation classes. Both datasets are stratified by gender, region and immigrant status, resulting in the following classes: West German men, West German women, East German men, East German women, male migrants, and female migrants.[151]

---

150 There are numerous names for this type of variables (also *cohort* or *stratification* variables). In the remainder of this paper, the authors use the term *slice* variables the groups are named matching strata or imputation classes.

151 Migrants are not divided into East and West Germany to assure a sufficiently high number of observations.

As for the continuous matching variables, annual income measures dominate the matching and imputation exercise. We use an aggregate income measure that summarizes all types of income that qualify for the accumulation of pension rights (earnings, unemployment benefits, sickness allowances, etc.). These variables are highly relevant, because they are the best predictor for the individual's SSW. The income measure enters the equation as a three-year moving average to smooth individual income histories (average annual income for the years 1983-1985, 1984-1986, 1985-1987 ... 2004-2006). For all 2007 SOEP respondents with incomplete income profiles, we impute missing information starting in 2006 and going backwards to 1983 for West Germany and 1991 for East Germany. The imputation makes maximum use of all available longitudinal information since the respondent's first participation in the SOEP.[152] For reasons of comparability, earnings are cut at the effective maximum contribution ceiling for each year.[153] Differential treatment of earnings in the statutory pension insurance depending on whether a person is a regular employee, civil servant or self-employed requires consideration in the imputation.

A woman's fertility history is an additional piece of information which enters the matching/imputation as it determines the number of child care credits a woman receives. In line with pension legislation, we assign women one year of pension credits for all children born before 1992 and three

[152] This reverse completion of income information was necessary, because otherwise cases with missing values are excluded from the matching process. Furthermore, it improves the efficiency of the matching exercise assuming sufficient quality and representativeness of the imputed income data.

[153] In the survey data at hand, respondents report their monthly earnings, whereas in social security data earnings are cut at the maximum contribution ceiling, e.g. the amount above which no additional social insurance contribution have to be paid and no additional entitlements are accrued. For 2007, the maximum contribution ceiling was fixed at €5,250 in monthly gross earnings for West Germany and €4,550 for East Germany (Deutsche Rentenversicherung 2010). For previous years, the income thresholds need to be adjusted accordingly with the respective year- and region-specific values.

years for all children born thereafter.[154] We include the total number of child care credits a women has in 2007. In addition, various duration variables enter the computation. These measures reflect the number of years spent in different pension-relevant activities such as employment, unemployment, education, compulsory military or community service (only for men), as well as long-term care giving. Finally, the matching and imputation exercise includes the age of the respondent in 2007.[155]

The four techniques tested in this paper make different use of the available matching information: Simple hot deck imputation ignores the continuous matching variables and imputes randomly within the six strata, in contrast to the other three approaches that also take the continuous variables into account.[156]

4.4.4 Second Match: Record Linkage of SAPA and Divorce Statistics

The special role divorce plays in the accumulation of SSW cannot be adequately considered in the first match of SOEP and SAPA. This inadequacy is due to the fact that at the time SOEP respondents report their monthly pension benefit, it is impossible for them to tell entitlements from employment (or other individual pension relevant circumstances) and those resulting from the pension splitting because of a divorce apart. Hence, the authors expect that the premium or deduction resulting out of the pension splitting between former husband and wife leads to a system-

[154] Each credit is worth one earnings point, equivalent to the average earnings of all contributors in the respective year.

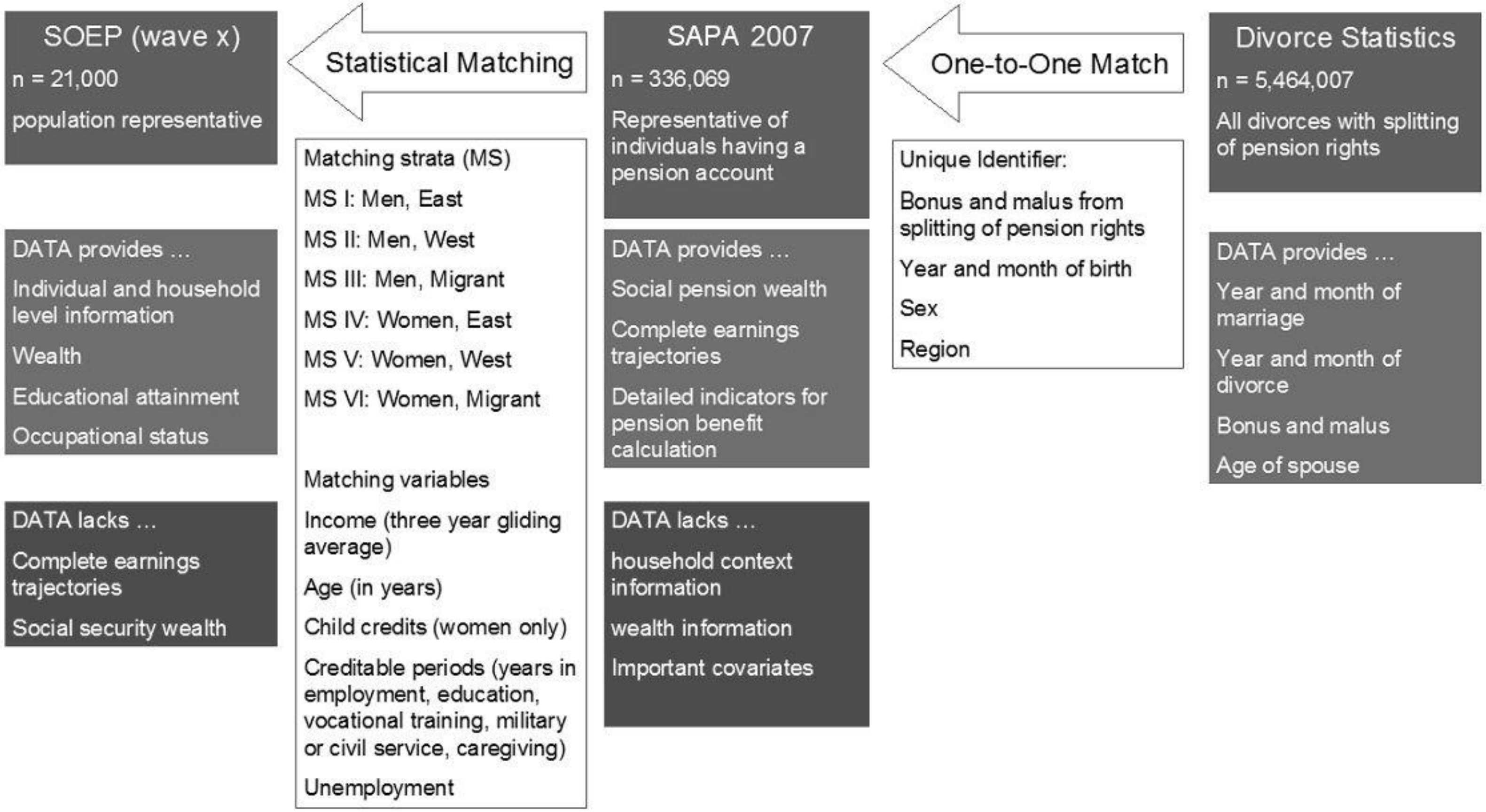
[155] The authors ran a total of nine matching algorithms with different combinations of matching variables and sample specifications. *Table A36* in the Appendix provides a summary of these tests.

[156] The authors are aware that hot deck imputation could be further refined through the inclusion of grouped income as an additional slice variable. However, the small sample size of our test population does not allow for this refinement.

atic bias in the linkage of SOEP and SAPA.

These potentially biased matching results motivate a second match of SAPA and Divorce Statistics using record linkage. The unique identifier is a combination of the exact amount of the premium or deduction from pension splitting, gender, region, and age. Information available in the Divorce Statistics allow for the correction of the divorce bias. With estimates from OLS regression models for each of the six matching classes, we correct for this bias in SOEP data. *Figure 19* summarizes the double-matching approach.

Figure 19 Statistical Matching Process at a Glance- SOEP, VSKT and Divorce statistics



Source: Author's Illustration

4.5 Assessing the Quality of Matches: Which Technique Performs Best?

4.5.1 Sample Specification and Evaluation Criteria

We now test the performance of each matching and imputation technique based on the population of retirees. For this group, the data provides an effective benchmark to assess the quality of the matches. SOEP respondents provide the presumably *true* monthly pension benefit (henceforth, *observed benefit*) that allows for the comparison with the *simulated* benefit (henceforth, *matched benefit*) from each imputation and matching technique, respectively.[157]

We perform this test for recently retired individuals aged 60 to 67. To guarantee the consistency of both sample populations, the samples do not include disability pensioners, because of significant differences in eligibility rules and pension benefit calculation as compared to old-age pensioners. The analysis excludes civil servants because they lack pension-relevant income for most parts of their working life. If they accrued any entitlements in the public pension scheme, they are typically credited against their civil servants pension as they retire. For self-employed individuals in the SOEP, income information was set to zero for the years of self-employment because it is typically not pension-relevant. The samples consist of 659 SOEP, and 34,353 SAPA observations.

We assess the validity of each approach using three criteria: 1) the correlation coefficient of the *observed* and *matched* public pension benefit; 2) the av-

---

[157] In the remainder of this paper, the terms *observed* and *reported* benefits are used interchangeably. Both describe the public pension benefit information provided by SOEP respondents. As with all survey information, data are prone to measurement error (Kreuter et al. 2010).

erage differences between *observed* and *matched* benefit; and 3) for the overall fit a graphical representation (*kernel density plots*) of the individual differences between *observed* and *matched* benefit. Each criterion is evaluated for the total population and within each of the six matching strata.[158]

4.5.2 Results

For the total population of retirees, pairwise correlations between *observed* and *matched* public pension benefits are best for Mahalanobis matching. The correlation coefficient r_{MAHA} of almost 0.7 is slightly higher than for UVIS and the regression-based approach with 0.67 and 0.68, respectively. Hot deck imputation clearly lags behind (r_{HOT}=0.22). The lack of association between observed and matched benefit for hot deck imputation is also true for the within-group correlations that range from -0.53 for female migrants to 0.17 for East German women. This result is due to the random assignment of matching partners without taking further continuous information into account.

For the other three techniques, within-group correlations always fall below the correlation coefficient of the total population except for the rather small group of female migrants (r_{MAHA} =0.82, r_{REG} =0.79, and r_{UVIS} =0.76). Concerning the other matching strata, UVIS performs best for East German men (r_{UVIS} =0.45) and male migrants (r_{UVIS} =0.63), the regression-based approach for East and West German women with 0.55 and 0.65, respectively. Mahalanobis fits best for female migrants (r_{MAHA} =0.82) and West German men (r_{MAHA} =0.43). Since the results are rather inconclusive with none of the techniques standing out, the correlation coefficient alone is no sufficient criterion to decide which technique to apply.

---

[158] We ran additional robustness tests to assess the stability of results by drawing five random samples with replacement and five disjoint random samples without replacement showing no notable variability. *Table A37* and *Table 38* in the Appendix show the results for the correlation coefficients of the total population.

The second evaluation criterion is the mean difference between observed and matched benefit:

$$\bar{d}_{i,j} = \frac{SOEP_{SSW_{retired}} - SAPA_{SSW_{retired}}}{n_{retired}}$$

The difference describes how far off the matched benefit is from the observed public pension benefit. A small average distance and standard deviation are indicators for a good match. Considering the distance criterion for the total population, hot deck imputed values fare best when it comes to the average distance (d_{HOT}= 16.4), but poorly with respect to the standard deviation of 542.8. This standard deviation is significantly higher than for all other techniques. Mahalanobis is second best in terms of distance and best with respect to the standard deviation (d_{MAHA}=-74.8; std. dev. = 320.6). The regression-based approach and UVIS are quite similar in their performance, but clearly lag behind Mahalanobis matching.

Table 33 Average Distance between Observed and Matched Benefit across Imputation and Matching Techniques

	Men East (n=126)	Men West (n=138)	Men Migrant (n=47)	Women East (n=141)	Women West (n=154)	Women Migrant (n=28)	Avg. Rank
Hot deck	-124.24 4	112.24 4	202.97 4	-53.56 1	39.08 1	90.88 1	2.5
Regression	-114.73 3	-42.81 2	-163.02 3	-125.80 4	-111.41 3	-142.05 3	3
UVIS	-105.79 1	-70.02 3	-131.47 2	-90.56 3	-119.46 4	-145.24 4	2.8
Mahalanobis	-106.86 2	-34.77 1	-90.69 1	-70.47 2	-76.91 2	-110.36 2	1.6

Source: SAPA 2007 and SOEP 2007; Author's calculations

In *Table 33*, we rank the performance of techniques (second row in each cell) with respect to the within-group average distance between observed and matched benefit (first row in each cell).[159] Mahalanobis matching

[159] For a complete overview, confronting mean, median and standard deviation for each

works best for West German men (d_{MAHA}=-34.7) and male migrants (d_{MAHA}=-90.7). Hot deck imputation performs best for East and West German women (d_{HOT}=-53.6; d_{HOT}=39.1) and female migrants (d_{HOT}=-110.4), but provides clearly the worst results for men. UVIS yields the best results for East German men (d_{UVIS}=-105.8). Across all groups, Mahalanobis renders the best outcome with respect to the average distance criterion.

Despite comparatively small average distances for groups 4 to 6 under hot deck, the standard deviation is by far the highest. With respect to the standard deviation, UVIS performs best for East German men as well as East and West German women, whereas Mahalanobis is better for West German men and male migrants. The distribution of matched values from the regression-based imputation has the lowest standard deviation for the group of female migrants.[160]

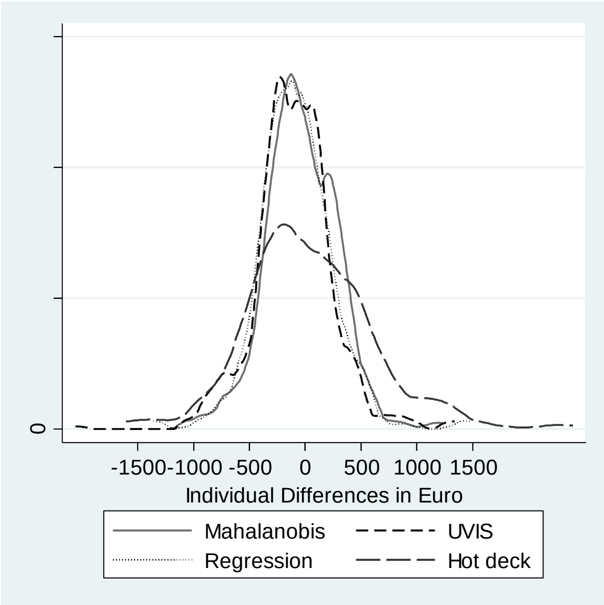
Kernel density plots depict the distribution of differences between observed and matched benefit information for all four approaches. Ideally, these plots are symmetric, unimodal and clustered around zero with a small standard deviation. *Figure 20* presents the kernel density plots for the total population.

---

technique and across all groups, see *Table A39* and *Table A40* in the Appendix.

160 Tables *A40* to *A41* present the results for the total population under the alternative matching algorithms.

Figure 20 Kernel Density Plots for Individual Differences between Observed and Matched Benefit Information - Total Population



Source: SAPA 2007 and SOEP 2007; Author's Illustration

The graphic representation underlines that hot deck is not the appropriate imputation technique. The distribution of differences has a substantial standard deviation with rather thick tails to both sides. Despite its better performance relative to hot deck, the UVIS distribution has no unambiguous peak. The distribution (dashed black curve) appears to be much wider at the top with several smaller peaks. The kernel density plots for the regression approach (dotted black curve) and Mahalanobis (solid grey curve) come closest to the ideal. The distribution for Mahalanobis is centered on zero but shows a small bump at +250 Euro. The kernel density curve for the regression-based technique has no such bump, but the peak of the distribution is more spread out.[161]

The divorce correction by and large improved the quality of matches. *Ta-*

[161] *Figure A13* in the Appendix displays the kernel density plots for the six matching strata.

ble 34 presents the absolute average distances (using *Mahalanobis distance*) with and without the divorce correction for the divorced population only.[162]

Table 34 Effect of the Divorce Correction – Only Divorcees

Difference between true and simulated pension benefit (Euro)		With divorce correction	Without divorce correction	Absolute effect
Total (n=137)	Median	-63.2	-46.7	16.5
	Mean	-36.8	-38.3	-1.5
	Standard Deviation	347.6	379.3	31.7
Men East (n=26)	Median	-47.1	-92.0	-44.9
	Mean	-50.9	-124.0	-73.1
	Standard Deviation	296.3	338.3	42.0
Men West (n=33)	Median	77.9	-127.3	-205.2
	Mean	3.4	-190.9	-194.3
	Standard Deviation	382.3	425.9	43.6
Women East (n=25)	Median	-123.2	-63.9	59.3
	Mean	-60.7	-8.5	52.2
	Standard Deviation	212.1	219.2	7.1
Women West (n=42)	Median	-109.4	29.0	138.4
	Mean	-58.4	117.5	175.9
	Standard Deviation	388.0	363.9	-24.1

Note: Migrants are omitted because of the small number of divorced migrants in the SOEP population. Source: SAPA 2007 and SOEP 2007; Author's calculations

Without the divorce correction matched benefits were too large for divorced men and too small for divorced women.[163] The divorce correction shifts results in the expected direction. Consequently, the absolute average distance changed from $d_{W/O\ Correction}$ =-124.0 to $d_{With\ Correction}$ =--50.9 for East German men and from $d_{W/O\ Correction}$ = -190.9 to $d_{With\ Correction}$ = 3.4 for West

162 Because of the small number of divorced migrants in our sample, those results are not further discussed.

163 Adding divorce as a slice variable is infeasible due to the small number of divorcees in the test population.

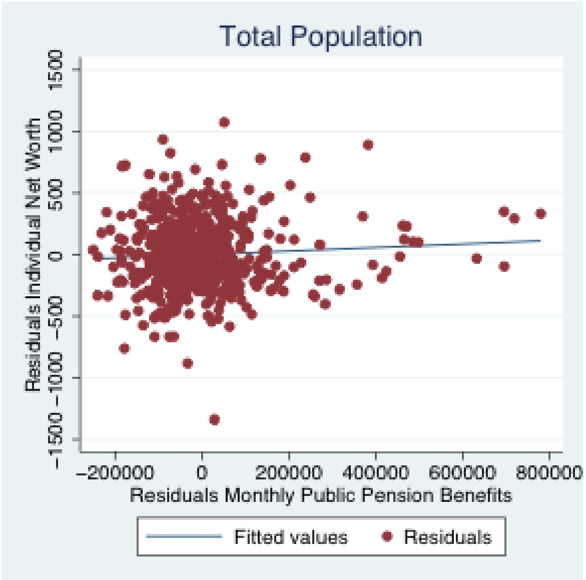
German men, respectively.[164] In turn, differences for women shifted in the other direction. The absolute average distance for East German women is $d_{W/O\ Correction}$ = -8.5 without and $d_{With\ Correction}$ = - 60.7 with the correction. For West German women, the difference shifts from $d_{W/O\ Correction}$ = 117.5 to $d_{With\ Correction}$ = - 58.4. For almost all groups, the divorce correction reduces the standard deviation of differences between observed and matched benefit. Hence, the quality of matches for divorcees converges to the quality of non-divorced individuals.

We also test for conditional independence. First, we run two OLS models: Model I regresses the matching variables on the individual's SSW; Model II regresses the matching variables on total net worth. For the conditional independence assumption (CIA) to hold, the residuals derived from both predictions have to be uncorrelated. For the total population of retirees, we find a correlation r=0.057 with a t-statistic of t=1.44. Therefore, we fail to reject the hypothesis that the correlation coefficient is significantly different from 0. This finding indicates that the CIA holds for the population of retirees and we assume this proposition to be true for the total population. *Figure 21* plots the residuals from Model I against the residuals from Model II.

---

164 The shift is more significant for West than for East German divorcees, because pension splitting was only introduced in 1991 and confined to entitlements earned and marriages divorced thereafter.

Figure 21 Residual Plot Testing for Conditional Independence



Source: SAPA 2007 and SOEP 2007; Author's calculations

4.5.3 Discussion

Mahalanobis distance matching performs best if we factor in all three criteria. Nonetheless, certain patterns require further explanation: First, a systematic negative bias in the average difference between observed and matched benefit. Second, significant between-group differences indicating better matches for some groups than for others.

The average distances are negative for the total population as well as for most of the individual groups, which indicates that the matched information is systematically higher than the reported benefit. The payment of insurance contributions for health and long-term care is one possible explanation for this bias. SAPA data provides the gross public pension benefit.[165] In turn, SOEP respondents likely report their public pension benefit

165 From this gross benefit, the statutory pension insurance pays health and long-term care premiums and then transfers the net benefit to the retiree.

net of health and long-term care premiums. Another explanation for the systematic bias is that the matching/imputation exercise fails to take actuarial adjustments for early retirement into account on the part of SAPA data, which possibly leads to an overestimation of approximated pension benefits.

We find substantial differences in the quality of matches: on average, the quality is better for men than for women, better for West than for East Germans and better for Germans than for migrants. Results of additional robustness tests indicate that the quality of matches is closely linked to the number of years a person has been observed in the SOEP as well as the number of years with income information larger zero. Second, the quality of the match depends on how good a predictor the observed information is for the individual's final public pension benefit given that we don't observe a significant portion of the SOEP respondents' lives.[166]

Matches are particularly good for West German men. For this group of SOEP respondents, we observe the annual income for more than half of their working life, which is a good predictor for their final public pension benefit.[167] In contrast, matches are less good for East German men. SOEP data collection in East Germany started only after the fall of the wall. Therefore, the matching/imputation exercise dismisses almost two thirds of elderly East German men's working life that are however relevant for their public pension benefit: First, this group has claims in special and additional pension schemes for GDR elites that increase the final benefit

---

[166] SAPA data provides earnings information for the entire working life, in contrast to the SOEP that started data collection in 1983 and 1991, respectively. For the matching, we only use earnings information for the years available in both datasets.

[167] We also tested a matching algorithm exclusively restricted to income information. This variant rendered exceptionally good results for West German men, underlining the predictive power of income for their final public pension benefit. *Table A41* and *Table A42* in the Appendix display the matching results for the *only income* variant (Model #8).

significantly but cannot be controlled for in the data. Second, the labor markets East German men worked in and their earnings before reunification had nothing in common with the situation after the German reunification. In the centrally planned economy, everybody had a job and unemployment was no issue. This job security and the continuous working careers were favorable for the accumulation of SSW of East German men. However, post-reunification labor market experiences differed greatly with age-earnings profiles being flat and returns to tenure and experience significantly lower when compared to West German men (Orlowski and Riphahn 2009).[168] Unemployment, for a significant share of this group even long-term unemployment largely limited the ability to accumulate SSW.[169] These reasons explain why the observation of only the most recent years might not be the best predictor for the final pension benefit of East German workers.

The quality of matches is better for East German women than for East German men. East German women benefit to a lesser extent from the transfer of entitlements from special and additional pension schemes than men. The biasing effects these benefits have on the matching quality are therefore less strong for women.[170] Further, women in our sample were disadvantaged with respect to earnings and occupations before, but also lacked proper employment opportunities after reunification. The ob-

---

168 Orlowski and Riphahn suggest that for many East German men job-specific human capital was out dated and did not match the requirements of the job market in unified Germany (2009).

169 Given that the *production* of SAPA data is directly linked to administrative processes the available information is by nature more accurate, in particular when measuring short spells of unemployment. In the SOEP, respondents might not perfectly recall these shorter spells. Due to higher unemployment rates in East Germany following reunification these differences in measurement might contribute to a greater difference between observed and matched pension benefit when compared to West Germany.

170 With a share of 92 percent men were highly overrepresented in the special pension schemes and to a lesser extent so in the additional pension schemes with 54 percent (Seitz 2003).

served years of East German women's lives in the SOEP are to a greater extent representative for the unobserved years, which improves the matching quality.

The matching results for West German women appear to be less good than those for East German women. Despite a longer period of observation, data mostly covers the period of economic inactivity of these birth cohorts following the years of childbearing and -rearing. This inactivity is particularly prevalent among older birth cohorts of women as a consequence of the strong *female caretaker/male breadwinner* notion promoted in the West German postwar welfare state promoted.[171] Due to these rather uniform working patterns, the years observed in the survey are not necessarily representative for the unobserved period of life. Therefore, the second half of West German women's working lives is not such good a predictor for their final public pension benefit. Following this line of argument it comes as no surprise that the matching quality is poorer for migrants. On average, years observed for both male and female migrants fall short of those observed for natives. It is likely that these years are not representative for the total employment biography. Pensions based on bilateral social insurance treaties with other countries also account for large differences between observed and matched benefit information among migrants.[172] For the analysis, it is infeasible to separate benefits earned in Germany from benefits earned in other countries.

Based on the results, it is safe to apply the double-match to the working age population to obtain the best estimate of SSW. The authors expect the

[171] The weak labor market attachment of West German women is reflected in the matching variables: For each year observed, more than half of this group has a pension-relevant income equal to zero.

[172] Persons who worked and accrued pension rights in Germany and another country receive a so-called *Vertragsrente* (Himmelreicher 2005). Individuals qualify for the payment of such a pension if the two countries have a bilateral social security agreement (also *totalization agreement*).

matching quality to be even better for this segment of the population: First, the number of observations is significantly higher (14,247 SOEP with 288,655 SAPA observations), hence SOEP observations have more potential matching partners to choose from. And second, the SOEP covers a greater share of peoples' working life, which feeds more reliable information in the statistical matching and therefore reduces uncertainty.

4.6 Wealth Inequality

4.6.1 Determining the Present Value of Pension Entitlements

The inclusion of SSW in the wealth inequality analysis requires the calculation of the present value of pension entitlements (for a detailed description see Rasner et al. (2011)). The present value considers pension rights from the statutory, company and private pensions.[173] For retirees, we substitute the matched SSW with the true amount stated in the survey, because this information seems to be more reliable than any simulated benefit could possibly be. For the working age employees, we keep the SSW as assigned by the Mahalanobis distance matching. The entitlements for active civil servants are approximated in the following way: As a final-salary scheme, gross earnings during the last three years of service and the number of service years are the basis for the calculation of benefits.

The actual calculation of the present value of recurring pension payments requires information on life expectancy, retirement age and the taxation of retirement income. As for the life expectancy, the authors rely on the 2005/2007 life tables of the German Federal Statistical Office. The life ta-

[173] In contrast to the SSW from the statutory pension insurance, there is no comparable information such as the SAPA available for company pensions and entitlements for liberal professions. However for the population of retirees these pensions are directly surveyed in the SOEP. This procedure leads to an underestimation of SSW in the population of currently active members in the labour force.

bles provide information on remaining life expectancy by sex and region (East and West Germany). Finally, differential taxation of retirement income by occupational group calls for inclusion.[174] For the calculation of the present value of such entitlements, different 2007 tax rates apply depending on the occupational group. We assume the future indexation of pension payments to be in line with inflation, so that the real value of entitlements stays constant over time. For discounting purposes, we assume an interest rate of two percent for most of the analyses, but illustrate how alternative specifications with varying discount rates affect the present value of social security wealth in the next paragraphs.

4.6.2 Social Security Wealth Assuming Different Discount Rates

In the following analyses, we assess the present value of social security wealth under the assumption of differing discount rates. Entitlements grow with increasing age up to the time around retirement, with the slope in the second phase of working life being somewhat steeper (cp. to *Figure 22*). The shape of the present value of social security wealth is similar to age-earning profiles, because paid contributions are a fixed share of earnings up to the taxable maximum (*Beitragsbemessungsgrenze*). At retirement, the individual exchanges the accumulation of entitlements for pension payments for the rest of his or her (statistical) life, therefore gradually decreasing the present value of pension entitlements.[175]

Figure 22 also illustrates that the level of the discount rate matters: With a discount rate of two percent the net present value of all pension entitle-

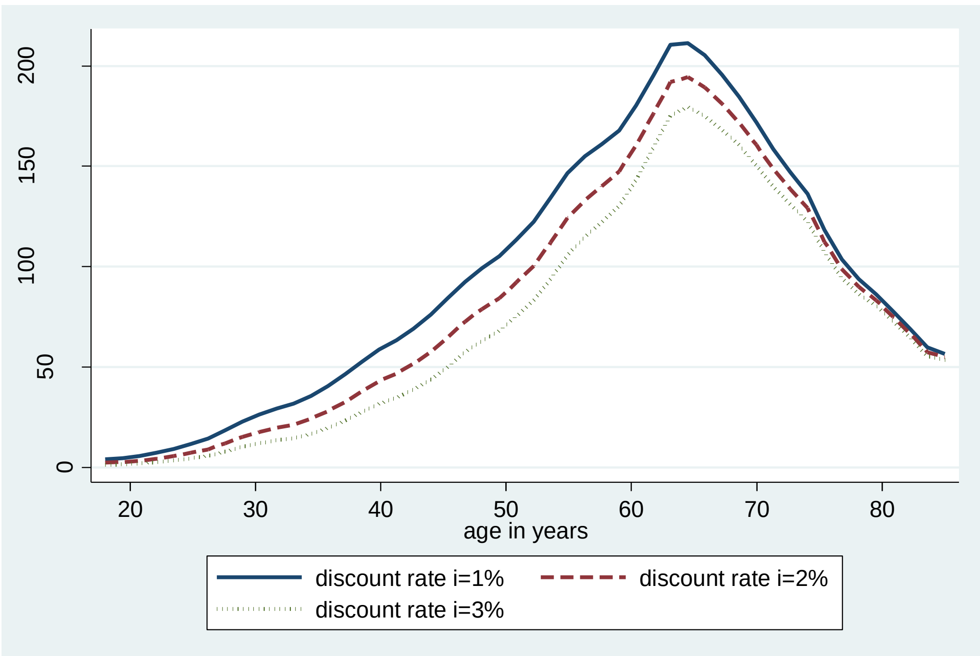
[174] So far, we only have the *gross* SSW for employees and civil servants. Their retirement income is subject to a differential tax treatment. The annuities of civil servants are already fully taxed. In contrast, life annuities, benefits from the public pension scheme, agricultural old age funds, or pension schemes organized by professional associations are taxed only to a certain degree (see §22 of the German Income Tax Act).

[175] With the *statistical death* of an individual, the present value of entitlements equals zero.

ments equals about 5.6 trillion Euros in 2007 which corresponds to an average value of 78,500 Euros for every adult in Germany or a median of roughly 47,000 Euro. While choosing a discount rate of two percent appears somewhat normative, this value reflects the long-term real interest rate for federal bonds in Germany. Alternatively, an interest rate of 1% and 3% yields an aggregated net value of pension wealth of 6.5 and 4.9 trillion Euros, respectively. The corresponding means amount to about 91,000 and 68,000 Euros.

More important for the sake of our analysis: The choice of the discount rate impacts the level of the present value of pension wealth, but it does not change the shape of the distribution over the life cycle. For all three curves in *Figure 22*, the present value peaks at retirement age. The maximum value for a discount rate of 1% is 210,000 Euro, for a rate of 2% more than 190,000 Euros and finally for a rate of 3% the maximum equals 170,000 Euros.

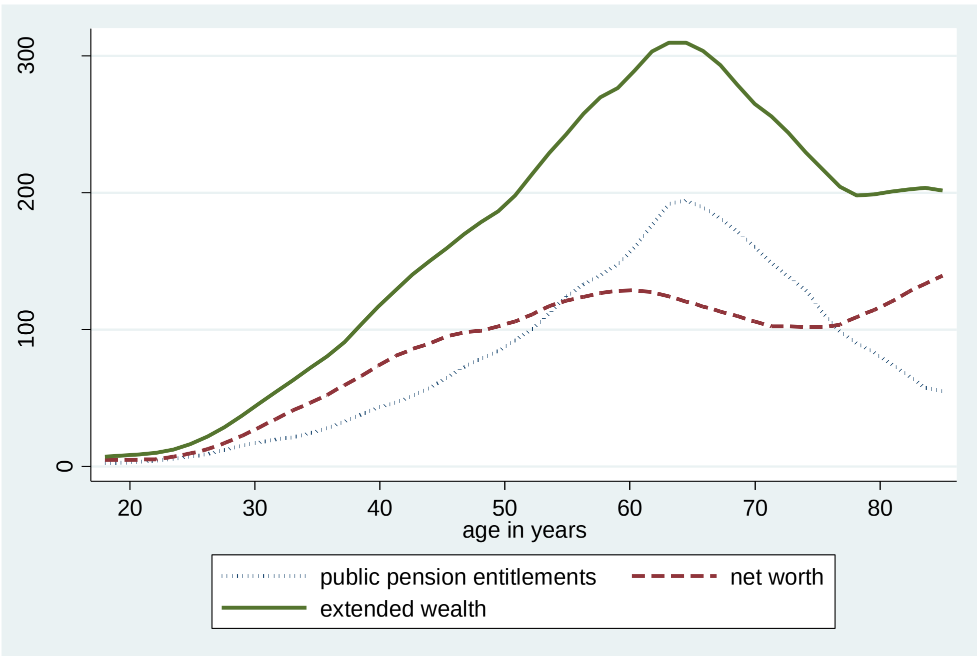
Figure 22 Present value of pension wealth entitlements by age for different discount rates, Germany 2007



Source: SOEP 2007; Author's calculations

What happens to the distribution of total net worth if we incorporate the present value of pension entitlements? *Figure 23* compares the age profiles for (public) pension entitlements, total net worth, and the extended total net worth measure including pension entitlements. Pension entitlements peak at around retirement (65 years) with more than 190,000 Euros, whereas total net worth peaks somewhat earlier at around age 60 with a value of about 140,000 Euros. Net worth decreases after age 60, but at a slower rate than pension entitlements. This decrease in net worth is possibly due to early inheritances or gifts to children and grandchildren. Interestingly, net worth increases again for the oldest-old (ages 80 and over), which possibly reflects a concentration of wealth holdings among widow(er)s following inheritances, but also demographic processes such as selective mortality in favor of wealthy elderly (survival of the fittest, here: wealthiest). Finally, for the extended wealth measure we observe a maximum amount of more than 300,000 Euros around retirement age.

Figure 23 *Net worth, present value of pension wealth entitlements and extended wealth by age, Germany 2007*



Note: Public pension wealth using a discount rate of 2%. Source: SOEP 2007; Author's calculations.

4.6.3 The Distribution of Total Individual Net Worth and SSW

The wealth inequality analysis compares a standard measure of net worth to an extended measure of net worth that includes SSW (henceforth, *extended wealth*) obtained from the statistical matching.

In a first step, we separate total net worth and SSW and take a look at their respective distributions. The first column of *Table 35* indicates that the aggregate net worth for individuals in private households in Germany amounts to about 5.9 trillion Euro in 2007. If this amount was evenly split, each adult person would have about 83,000 Euro at his or her disposal. Comparing mean and median of the distribution of total net worth gives a first indication for the degree of overall inequality. In fact, median wealth equals 15,000 Euro. Thus, the mean exceeds the median by factor 5.5. With about 78,500 Euro, the average SSW comes close to average net worth (see column 2 in *Table 35*).[176] But, the distribution of SSW is less skewed than that of total net worth, because mean and median are closer. The mean exceeds the median by factor 1.7.

[176] The above calculations apply a discount rate of two percent. While this choice might appear normative, this value reflects the long-term real interest rate for federal bonds in Germany. Alternatively, an interest rate of 1% and 3% yields an aggregated net value of pension wealth of 6.5 and 4.9 trillion Euro, respectively. The corresponding mean values amount to about 91,000 and 68,000 Euro.

Table 35 Net Worth and Social Security Wealth in Germany[1], 2007

	(1) Net Worth (in €)	(2) Social Security Wealth (in €)[2]	(3) Extended Wealth (in €)	(4) Change (%) [(1) / (3)]
Sum in trillion Euro	5.908	5.581	11.489	94.5
Basic Statistics				
Mean	83,077	78,479	161,556	94.5
Median	14,751	46,680	94,675	541.8
Wealth shares (in %)				
lowest Quintile	-1.5	0.9	0.4	126.7
2nd Quintile	0.4	5.2	4.5	1025.0
3rd Quintile	3.9	12.0	11.8	202.6
4th Quintile	17.3	24.1	22.4	29.5
highest Quintile	79.9	57.7	60.9	-23.8
Population with zero or negative wealth (in %)	28.1	4.5	3.3	-88.3
Indicators of Inequality				
Gini	0.80	0.56	0.60	-24.6
HSCV	6.51	0.73	2.02	-68.9
P90:P50[3]	14.15	4.11	3.82	-72.9

Note: [1] Population: persons in private households aged 17 or older (N=69,321,834). [2] With a discount rate of two percent, without provision for dependents. [3] Lowest value of the top ten percent in the wealth distribution in relation to the median (50 percent). Source: SAPA 2007 and SOEP 2007; Author's calculations.

The evidence on the relationship between mean and median suggests significant differences in the distributions of net worth and SSW. The wealth shares provided in *Table 35* further support this evidence. The top 20 percent of the adult population hold almost 80 percent of total net worth, whereas the three bottom quintiles own less than 3 percent of total net worth. About 28 percent of the adult population has no or even negative net worth, indicating that liabilities exceed gross wealth. In contrast, less than five percent of the total population did not accumulate any SSW. Pension entitlements are by far more evenly distributed than net worth, mainly because almost everybody accumulates pension entitlements at least once over the working life. In addition, income is subject to contributions only to an upper limit in the statutory pension insurance. Nevertheless, the highest quintile still holds the bulk of SSW with almost 58 percent of total SSW. The *Gini coefficient* reflects the differences in inequality in the distributions of these two wealth concepts. For net worth the coefficient equals 0.8 indicating a high degree of inequality, whereas for SSW it amounts to 0.57.[177] The inclusion of SSW in our extended wealth measure almost doubles the average net worth (161,500 Euro). However, a much stronger increase could be observed for the median which hit nearly 95,000 Euro. Inequality is decreasing by one quarter for the Gini coefficient when moving from the standard to the extended measure of net worth. Those in the middle of the distribution profit the most with wealth shares mounting by almost eight percentage points.

4.6.4 Net worth and Extended Wealth across Occupational Groups

The individual's occupational status is fundamental determinant not only for a person's income level and his or her ability to save but also an important proxy for the level of net worth.[178] Beyond that, the individual's occupational status is also relevant for the accumulation of SSW. The occupation determines the type of pension scheme a person belongs to and the rules by which SSW accumulates (compare to *Section 4.2*). *Table 36* provides evidence on how occupational status relates to wealth holdings in Germany. It compares the three measures of interest: standard net worth, present value of SSW, and extended wealth. For a more complete picture, *Table 36* gives information on the age/sex composition of each occupational group (median age and share of females) to better take compositional differences into account.

---

177 Other indicators such as half-squared coefficient of variation (HSCV) or the P90 to P50 percentile ratio point in the same direction. For example, the results for HSCV are even more pronounced with 6.5 for net worth and 0.7 for SSW, mainly because of the top-sensitivity of this indicator. For more details see Rasner et al. (2011).

178 In the following section, a person's occupational status refers to the information provided in the SOEP individual questionnaire of 2007. It is however possible, that a person has previously worked in another profession, which may affect the level of net worth and SSW.

Table 36 Net Worth and Social Security Wealth by Selected Occupational Groups in Germany[1], 2007

	Occupational Group	Net Worth (in €)	Social Security Wealth (in €)[2]	Extended Wealth (in €)	Relative Change (in %)	Median age in years	Percent Female
Workers and Employees	Unskilled, semi-skilled, salaried employees without an apprenticeship	33,618	53,965	87,582	161	43	55
	Trained and skilled, salaried employees in low qualification positions	46,964	56,043	103,007	119	42	41
	Foremen, masters, supervisors, salaried employees in qualified positions	69,256	60,128	129,384	87	42	58
	Salaried employees with extensive management responsibilities	122,778	74,955	197,734	61	42	33
Civil Servants	Sub-clerical or clerical service class	67,019	92,135	159,154	137	40	36
	Executive or administrative class	145,775	149,484	295,259	103	47	41
Self-Employed	Without any employees[3]	169,683	56,296	225,980	33	47	39
	With one to nine employees	351,185	38,064	389,249	11	46	25
	With ten or more employees	1,138,372	35,909	1,174,281	3	45	26
Not working	Persons of working age not gainfully employed	74,553	39,620	114,173	53	44	89
	Unemployed	15,406	52,070	67,476	338	42	53
Retired	GRV-Pensioners	98,956	129,763	228,719	131	71	56
	Retired civil servants	187,510	313,436	500,946	167	69	20
	Total	83,077	78,479	161,556	94	48	50

[1] Population: persons in private households aged 17 or older (N=69,321,834). [2] With a discount rate of two percent, without provision for dependents. [3] Including family members helping out. Source: SAPA 2007 and SOEP 2007; Author's calculations.

The results reveal substantial differences in wealth holdings across occupational groups. We also find that some occupational groups benefit to a greater extent from the inclusion of the present value of SSW than others. On average, net worth is highest for the self-employed given that they must save more to private old-age pension plans and because of their business capital itself which makes a significant difference. Evidently, the more employees a self-employed person employs, the higher their total net worth. In contrast, unskilled, semi-skilled workers and salaried employees (without vocational training) hold roughly 34,000 Euro in financial and material assets. In turn, skilled workers such as foremen or masters come close to 70,000 Euro in assets, while employees with management responsibilities hold more than 120,000 Euro.

In general, civil servants own above average net worth, which is especially true for civil servants in executive or administrative positions with an average individual net worth of more than 140,000 Euro. Civil servants in the sub-clerical or clerical service accumulate substantially less (67,000 Euro), but still more than skilled workers and salaried employees.

In line with the standard life cycle model of savings (Modigliani 1988), the elderly have above average net worth. This age effect is particularly striking for retired civil servants with a measure of net worth of nearly 190,000 Euro. In comparison, pensioners in the statutory public pension scheme have net worth of less than 100,000 Euro at their command. Civil servants are at an advantage in the accumulation of wealth not only because of the on average higher educational attainment, but also because they do not have to pay contributions into their pension scheme, which allows for a higher saving rate.

The inclusion of the present value of SSW benefits civil servants most. Retired civil servants have more than 310,000 Euro SSW, while the respective figure for pensioners in the statutory public pension scheme not even reaches half of this amount (130,000 Euro). In the active population, it is also the group of civil servants that profit the most from the inclusion of SSW. For low and medium level civil servants, pension entitlements amount to 92,000 Euro. For high level civil servants (executive and administrative class) these entitlements are even higher (almost 150,000 Euro). In fact, their SSW nearly doubles their net worth. Dependent employees do not benefit to the same extent from the inclusion of the present value of SSW. For the various groups of blue and white collar employees, SSW ranges from 54,000 Euro to 75,000 Euro. Currently unemployed have an average SSW of 52,000 Euro. This finding underlines the important role the public pension scheme plays in stabilizing the individual's economic position even in case of (short term) unemployment. For the self-employed the respective figures vary on a somewhat lower level compared to dependent employees (between 35,000 Euro and 56,000 Euro). Unlike other occupational groups, it is in the individual responsibility of the self-employed to provide for old age. They typically invest in life insurance policies or property.

Following from this, the extended wealth measure clearly improves the position of civil servants relative to the self-employed.[179] Nonetheless, the self-employed by and large stay on top of the wealth distribution.

4.7 Conclusion

This paper compares four statistical matching and imputation techniques to complement data on wealth from a population representative survey with information on SSW from administrative pension records. Statistical

[179] Civil servants benefit to such a great extent from the inclusion of SSW because they typically enjoy a continuous employment career without any interruptions due to unemployment. Furthermore, the institutional design of the civil servants scheme accounts for their favorable position. In the final salary scheme, the last three years of earnings count, which are typically those years in which earnings peak. In contrast, the statutory pension insurance takes the entire wage history into account.

matching proves to be a suitable technique to link information not available in one but several datasets that cannot be linked over a unique identifier. Rigorous robustness tests for the group of retirees identify Mahalanobis distance matching to be the best performing approach for the data at hand when compared to three alternative imputation techniques.

Applying the statistical matching strategy to the total population allows for the calculation of the present value of pension entitlements. The results illustrate that SSW represents a considerable source of wealth worthwhile to consider in wealth inequality analyses. Overall, SSW roughly amounts to 5.6 trillion Euro or - on average - 78,500 Euro per adult. When combined with net worth, SSW almost doubles the measure of extended wealth with an average of more than 160,000 Euro. The extended measure of wealth reduces inequality (Gini coefficient) by one quarter compared to standard distributional analyses that only take financial and material assets into account. This marked reduction in inequality is mainly the result of the lesser spread in the distribution of SSW and due to the fact that almost every adult in Germany has at least some entitlements in the various old-age pension schemes. We also find striking differences in levels of SSW across occupational groups. With respect to their position in the wealth hierarchy, civil servants benefit most from the consideration of pension wealth in the extended measure of wealth, whereas the wealth position of self-employed becomes somewhat less favorable as they tend to typically invest in financial and material assets for old age and thus hold rather low entitlements in traditional pension schemes.

Future research in this area should aim at developing formal indicators to assess the matching quality. Ideally, these indicators should work even in the absence of an effective benchmark, such as the reported pension benefits of the group of retirees observed in the survey. In this application, Mahalanobis distance matching is the best matching technique, but it may not be in others. The comparison of several matching and imputation techniques is appealing because it helps us better understand the compatibility of both datasets and also which technique works best in which context. These robustness tests also come into play if we plan to complement one dataset with more than just one variable from another dataset. For example, statistical matching could be used to complement complete marital biographies from SOEP data with complete earnings trajectories from SAPA data, which would allow us to address a multitude of new research questions. One such application follows in Section 5 analyzing the interdependencies of marital trajectories and pension accumulation of women in Germany and the U.S.

Finally, establishing multiple statistical matching strategies might be one research direction worth following, because it could be one way to reduce the uncertainty inherent in the matching process. Multiple statistical matching follows the idea of multiple imputation that has become the standard method to deal with missing data (Rubin 1987).

5 Women's Marital Trajectories and the Accumulation of Pension Benefits in Germany and the United States

5.1 Introduction

Employment-centered pension schemes favor continuous employment careers with no work interruptions and (above) average earnings. In this type of pension scheme, every year of employment counts towards the individual's future pension benefit, because with the payment of social insurance contributions or payroll taxes from their labor income, individuals earn pension entitlements that accumulate over the life cycle and qualify for the receipt of benefits as they retire. Career interruptions stop the accumulation of pension rights immediately and may also have persistent scar effects for the years following the interruption by slowing the accumulation down (Gangl 2006).

It is needless to point out that employment-centered pension schemes are more conducive to the employment careers of men, who work more continuously, interrupt work less often and earn more than women (Arza 2008; Frericks et al. 2007). For this type of career, pension schemes guarantee benefits upon retirement that allow workers to have more or less the same standard of living as if they were working. Typically, men - both married and unmarried - benefit from this pension design, whereas women - in particular married women – with different life cycle patterns of work and family choices are structurally disadvantaged (Lillard and Waite 2000). Motherhood and child care responsibilities, but also the resulting disruption in the accumulation of human capital and job-specific skills atrophy limit women's opportunities to accumulate sufficient pension entitlements on their own (Gangl and Ziefle 2009).

In the past, policymakers saw no reason for concern in the low pension rights accrued by women and the considerable gender pension gap , because the majority of women was assumed to be best protected through the old-age pension received by their husband or in case he dies, through the payment of survivor's benefits. The equal sharing of resources in old age among married couples was expected to compensate women in life-long marriages for their (intermittent) withdrawal from the labor market that limited their chances to earn pension rights on their own (Joshi and Davies 1991). The successes most industrialized countries had in alleviating poverty amongst the elderly further supported this viewpoint in most policymakers that there was no need for action, especially since poverty rates in the elderly population dropped below the levels of younger age groups (Bundesregierung 2008; Engelhardt and Gruber 2004).

However, increased divorce rates experienced by most industrialized countries put pension systems under pressure. Divorce ends the sharing of financial resources in couples not only during working life but also in old age (Ginn and Price 2002). In terms of pension entitlements, divorce is no problem for individuals who were always fully engaged in the labor market during marriage. However, individuals with a weak labor market attachment who are financially dependent on their partner are particularly vulnerable. For them to catch up in pension building is difficult given that they jeopardized their earnings capacity during the time being married. It is a well-known fact that the economic consequences of divorce are more severe for women than they are for men (Burkhauser et al. 1994; Duncan and Hoffman 1985; Smock 1994). Hence, policymakers push for a stronger individualization of pensions promoting the accumulation of independent pension rights away from a reliance on derived spousal or survivor's benefits (Frericks and Maier 2008).

The strong ties between marital status and labor supply are well-documented in the literature, but evidence is limited as to how marital trajectories affect the accumulation of pension rights across the individual's working life (Haider et al. 2003). This paper argues that the extent to which marital trajectories and pension building are intertwined largely depends on the welfare state context both processes are embedded in. This study compares the interdependencies of women's marital trajectories and the accumulation of pension rights in two diverging welfare states: Germany and the U.S. Both countries under study have mature public pension schemes that are based on a rather similar rationale, but exhibit considerable variations in their welfare state conceptions, the generosity of benefits provided, and the way they handle marital transitions over the life-course. At the same time, both countries experienced massive changes in partnership patterns during the last decades. Germany and the U.S. both saw significant increases in divorce rates in spite of differences in timing and magnitude.[180]

Systematic analyses of cross-country differences possibly identify pervasive incentives in welfare states that perpetuate gender-specific employment patterns, economic dependencies in couples, and insufficient financial resources in later life depending on their respective marital choices. So far, these studies have been rare because of the lack of access to adequate and comparable longitudinal data that allow us to study the long-run financial consequences of marital choices for pension building by tracking them through to retirement.

Based on two new and unique datasets linking extensive longitudinal survey data with administrative pension records that cover life cycle pension-relevant earnings, this study seeks to shed light on the dynamics of pension accumulation and marital trajectories in Germany and the U.S. For Germany, this study uses data from the Socio-Economic Panel Study (SOEP) matched with the Sample of Active Pension Accounts (SAPA) from the German statutory pension insurance. The reference data for the U.S. provides the Health and Retirement Study (HRS) linked with the 2004 Permissions: Wage and Self-Employment Income (W2), an adminis-

---

[180] For a comparison of the development of the crude divorce rate in Germany and the U.S. see *Figure A14* in the Appendix.

trative dataset maintained by the Social Security Administration.[181] The U.S. Permissions data are not available each year. To permit a meaningful analysis with a sufficient number of observations for both countries, this paper uses 2004 data for the U.S. and 2007 data for Germany.

The paper is structured as follows: The next session summarizes the existing literature dealing with the interdependencies of marital choices and labor supply and discusses possible repercussions for the accumulation of pension rights. Section 5.3 provides a detailed picture of the policy background in Germany and the U.S. with respect to the institutional design of their public pension programs, the treatment of marital transitions in these programs and the underlying welfare state conceptions in both countries. Also in this Section, I state the research hypotheses that direct the empirical analysis. Section 5.4 presents the data and the analytic approach. For the population of retirees, Section 5.5 provides descriptive and multivariate evidence on the interplay of marital trajectories and retirement outcomes. In Section 5.6 I take on a dynamic perspective and show how marital trajectories and the process of pension building are intertwined for the pre-retirement cohorts. Section 5.7 concludes and ends with an outlook on future research.

5.2 Literature Review

So far, our knowledge is limited as to how marital choices affect the dynamics of pension building over the adult's working life, mainly because adequate data became only recently available.[182] More is known about

---

181 Access to these confidential data is restricted. Data are only available under terms of a formal agreement between the Health and Retirement Study and the researcher. The author thanks Prof. Don Taylor and Anne Fletcher at the Sanford School of Public Policy at Duke University for providing access as a supplemental user under the HRS DUA 2004-011 in the name of Donald H. Taylor, Jr.

182 In Germany, access to administrative data was long time restricted because of confidentiality concerns. National initiatives like the *German Data Forum* (*Rat für Sozial- und*

how marital status or marital transitions affect the short- to medium-term labor supply of men and women. This evidence is valuable because in employment-centered pension schemes, labor supply directly translates into pension entitlements.

Since the late 1970s, much academic attention has been devoted to the link between labor supply and marriage. These studies found that women cut back on their labor supply upon marriage, but also in the years prior to getting married (Heckman and Macurdy 1980; Johnson and Skinner 1986; Mincer and Solomon 1978). Aughinbaugh shows that remarriage in women does not change the odds for the decision of whether to work or not when compared to the first marriage. However, women who worked during their first marriage work more hours in their second marriage (2010). Despite the reduction in labor supply of married women, they still enjoy higher levels of economic well-being than divorced women (Smock et al. 1999).[183] Nevertheless, in terms of pension rights, a decrease in labor supply of married women means less pension rights to accumulate.

During the last decades, the transition to divorce and the respective short- and medium-term consequences have received increasing attention. These studies analyzed how marital dissolution modifies the labor supply of individuals (Haardt 2006; Haurin 1989; Johnson and Skinner 1986) or the individual's income position (Jarvis and Jenkins 1999; Jenkins 2008; Poortman 2000) both for men (Kalmijn 2005; McManus and DiPrete 2001) and for women (Smock 1993; Smock 1994). Cross-country studies

Wirtschaftsdaten) have largely improved the data infrastructure for the social, economic, and behavioral sciences by opening up access to valuable data in the branches of the social insurance system and other public institutions (Rat für Sozial- und Wirtschaftsdaten 2010).

183 The authors show that the economic benefit married women have over divorced women has been overestimated in previous studies, because divorced women would not benefit to the same extent from marriage if they were to remain married (Smock et al. 1999).

helped to understand the role institutions play in mediating the economic consequences of divorce (Andreß et al. 2006; Wagner and Weiß 2006). The trigger events[184] literature also stressed how institutions can buffer the impact of potentially disruptive events on household income mobility, distinguishing job-related and demographic events such as unemployment (Gangl 2006; McManus and DiPrete 2001)[185], childbirth (Budig and England 2001), or divorce (DiPrete and McManus 2000; McManus and DiPrete 2001).

Another strand of literature focuses on differences in retirement outcomes across marital status groups. Studies in Germany and the U.S. identified divorcees, especially divorced women, to face an increased risk of old-age poverty (Bundesregierung 2008; Munnell 2004). Divorced women typically have low social security benefits on their own and no additional income sources to rely on when living alone (TNS Infratest 2009; Vartanian and McNamara 2002).[186] Based on the available empirical literature, the economic well-being of elderly divorced individuals seems to be more of a concern in the United States (Butrica and Iams 2000; Favreault and Steuerle 2007; Haider et al. 2003) and the UK (Ginn and Price 2002; Joshi and Davies 1991) than in Germany.[187] Little is known about the population of never married individuals that will increase over the next decades. For the U.S., studies show that never married elderly Americans are more

---

184 *Trigger events* are critical life course events that are likely to have an (negative) impact on the household's income situation (Gangl 2006).

185 These studies have a comparative focus in order to analyze how varying institutional contexts mediate the effects of trigger events.

186 In general, poverty rates for elderly single women are higher in the U.S. than in Germany. In both countries, poverty rates for divorced women are more than double the overall old-age poverty rates. Smeeding and Sandstrom come to the conclusion that elderly living arrangements matter more than age (2005).

187 In the U.S., cohorts born between 1940 and 1950 were the first to experience large increases in divorce rates. The majority of these birth cohorts will quit work between 2005 and 2015. In Germany, the cohorts who experienced the rapid rise in divorce in the 1980s will retire approximately ten years later.

likely to experience economic hardship than other marital status groups (Tamborini 2007).

All these works either focus on how the current marital status or marital transitions affect the individual's labor supply or alternatively, how retirement outcomes differ by current marital status. However, little is known as to how the marital history affects retirement outcomes. An exception is the paper by Wilmoth and Koso who illustrate that marital history matters when it comes to wealth outcomes (2002). They show large differences in the wealth outcomes of preretirement adults depending on their marital history as well as substantial gender effects. Substantial changes in the marital histories of the baby boomers also affect eligibility for spousal and widows benefits in the U.S. social security system (Tamborini et al. 2009). Due to the shorter duration of marriages that end in divorce, a growing share of divorced women foregoes eligibility for spousal and widows benefits.

This paper follows the line of argument that retirement outcomes and the process of pension building does not depend on marital status or a specific marital event, but rather on a sequence of different marital events across the life-course, namely the individual's marital trajectory. Furthermore, I expect that the repercussions marital trajectories have on the accumulation of pension rights differ depending on the institutional context these trajectories are embedded in. The following research questions guide the empirical analyses.

1. Do retirement outcomes differ across marital trajectories?
2. Do marital trajectories result in different paths of pension building?
3. If marital trajectories matter, to what extent do welfare state context and the institutional design of the pension system explain differences in retirement outcomes?

5.3 The Policy Background

5.3.1 The Process of Pension Building

This paper analyzes how marital trajectories affect retirement outcomes and pension building primarily in the public pension scheme in Germany and the U.S. Focusing on the public pension pillar dismisses a great deal of the retirement income picture. It is certainly true that other forms of old age provision such as occupational and private pensions, but also home-ownership (Frick and Grabka 2003; Frick et al. 2010b) complement social security benefits. Nevertheless, as a compulsory program in both countries, the social security scheme covers the majority of the population, which is not true for any other type of old-age provision. Social security benefits also make up for the largest share of the total retirement income in Germany and the U.S.

On average, stock market has done greatly over the last years. However, individuals don't live in averages. This means that if a person needs to cash out in the year of retirement, he/she is subject to considerable different set of risks. Social security is not prone to these risks and therefore the most important source of old-age income.

Retirement outcomes are the pension benefits individuals receive as soon as they retire. Pension building describes the process of accumulating pension rights across the working life. Typically, individuals earn pension rights through gainful employment. The process of accumulation starts with the first job that is subject to social insurance contributions or payroll taxes and ends with the transition into retirement.[188] The extent to

[188] In most pension schemes, individuals who work past retirement no longer accumulate pension entitlements but are allowed to work without any limits on earnings as long as they reached the full retirement age, which equally applies to the German and U.S. pension scheme. Matters are more complicated for individuals who retired early.

which the accumulated social security wealth is a reflection of the individual's earnings history and retirement benefits are a proxy for the person's life cycle labor market attachment depends on the institutional design of the pension scheme. More specifically, it depends on whether the scheme is funded or pay-as-you-go (PAYG) and on how strict a relation exists between pension benefits and the workers previous contributions or taxes (Barr and Diamond 2006).

5.3.2 The Institutional Design of Pension Schemes

In defined-contribution (DC) schemes, also called funded individual accounts, individuals pay a fixed share of their earnings that are invested in assets (Barr 2006). The accumulating assets and their returns are typically paid out as an annuity upon retirement. In this type of pension scheme, the size of the final pension benefit is related to a person's lifetime contributions, but more so to the successful accumulation of assets, the rate of interest and life expectancy at the time of retirement. Furthermore, DC plans face a multitude of risks that are out of the individual's control that potentially weaken the relation between pension benefits and previous contributions (e.g. macroeconomic shocks, future earnings, etc.).[189]

Pay-as-you-go schemes are defined benefit (DB) systems. It's in the nature of these schemes that individuals currently in the workforce pay contributions that finance the pensions of current retirees. In PAYG schemes, the relation between benefits and previous contributions is more straightforward than in DC schemes: the retirement income is a reflection of the person's work history. Year in and year out, employees and employers

Individuals under the full retirement age are subject to a limit on earnings and if they exceed this limit, a certain amount of their pension benefit is withheld.

189 For a detailed discussion of the risks associated to fully funded pension schemes see Barr and Diamond (2006).

pay a fixed share of their earnings into the public pension scheme that entitles workers to draw benefits as they retire. The pension benefit formula determines how closely contributions and benefits are linked depending on whether all years, the final years or a fixed number of best years enters the equation. Clearly, the contribution-benefit link is closest if all years with pension-relevant earnings are considered in the benefit calculation. In best years schemes however, benefits reflect a positive selection of a person's working career, which is beneficial to individuals with a few years of employment but otherwise weak labor market attachment. Policymakers can install additional provisions that weaken the benefit/contribution link in favor of individuals with a weak labor market attachment or below average earnings. These redistributive provisions include pension entitlements for certain forms of non-employment. Another instrument is the upgrade of below average contributions, for example during the child-rearing years of mothers. Alternatively, pension schemes can redistribute between low and high earners (Barr and Diamond 2006).

The rationale behind the German and U.S. social security scheme is quite similar.[190] Both systems are mature PAYG schemes. They are compulsory and cover more than 90 percent of the population in both countries (Kruse 2007; Tamborini et al. 2009). Both systems are employment-centered in that pension building primarily depends on the individual's lifetime pattern of economic activity. The U.S. benefit formula considers the best 35 years, whereas the German considers all years of employment. Because of the stricter benefit/contribution link, the German public pension system installed several provisions to weaken this tie. For example, it gives caretaker credits to the parent who predominantly cares for children or family members in need of care. The U.S. pension system gives no such

[190] For a brief sketch of the German and U.S. public pension scheme and their respective benefit calculation formulas, see *Tables A45* and *A46* in the Appendix.

credits, but redistributes between high and low earners with its progressive benefit formula that provides higher returns to the first than to the last dollar paid in contributions (Steuerle et al. 2004). This brief synopsis illustrates that the institutional design of both pension schemes is quite similar, which is surprising given the significant differences in the way they treat marital transitions in their pension programs and their welfare state conceptions.

5.3.3 The Impact of Marital Transitions

Despite their similarities in pension design, Germany and the U.S. differ with respect to the eligibility for benefits related to the individual's marital history. The marital status in itself has no impact on the accumulation of pension entitlements, because pension rules are by definition marriage neutral.[191] This neutrality implies that a certain amount of contributions or payroll taxes does not directly render more pension rights for married than for never married individuals. However, certain marital transitions over a person's life-course bring about changes in his or her pension rights, because they might establish indirect benefit eligibility (Tamborini et al. 2009).

In Germany, married individuals are not entitled to any kind of spousal benefit, whereas the U.S. social security system provides benefits to dependents as soon as the eligible worker retires conditional on the couple being married for at least one full year prior to the application for benefits.[192] The spousal benefit amounts to half of the worker's primary insurance amount (PIA) that depends on the work record of the entitled worker. The U.S. social security system pays spousal benefits as long as

191 In contrast, marriage neutrality does not apply to the German and U.S. tax system.

192 The early retirement age of 62 is the earliest possibility for workers to draw social security benefits.

the beneficiary is alive. If a married spouse, in addition to the spousal benefit, qualifies for his or her own retired-worker pension, the person is dually entitled, but collects only one benefit, whichever of the two is larger.[193]

The German and U.S. social security system both introduced benefits to divorced spouses in the late 1970s. However, the countries differ with respect to the timing the divorce settlements takes place and the amount of benefits paid. In the U.S., divorced elderly individuals might be entitled to divorced spouse benefits as soon as the entitled worker starts to collect the retired worker benefit. The benefit for the divorced spouse equals 50 percent of the ex-spouses Primary Insurance Amount (PIA).[194] The level of benefits depends on the ex-spouse's work record (Lillard and Waite 2000). In principle, divorced spouse's benefits are equal to spousal benefits. Both, married and divorced spouses can simultaneously collect benefits from their (ex)-husbands earnings record (The Urban Institute 2009).[195] However, divorced spouses are only eligible to collect benefits from their former spouse's record if the marriage lasted at least ten years.[196] Divorced spouses lose eligibility to divorced spousal benefits if they remarry and then stay married.[197] Remarriage at age 60 and older doesn't terminate the payment

---

193 Harrington-Meyer points out the misleading nature of the term *dual entitlement*. Women who are dually entitled receive a combination of benefits consisting of their own retired workers benefit and a share of the husband's benefit. Dually entitled persons receive the same amount of benefits they would receive, if they had never participated in the labor force (1996).

194 The PIA is the monthly pension benefit a person receives upon reaching the normal retirement age. Ten years of covered earnings over the entire work life are necessary for workers to qualify for the payment of Social Security benefits (Dickert-Conlin and Meghea 2008).

195 In fact, there is no limit on the number of divorced spouses that can collect benefits from their ex-husband. If none of the former spouses remarried, they are all eligible to draw divorced spouse's benefits (The Urban Institute 2009).

196 Before the reform in 1977, divorced workers were eligible for spouse and survivor benefits if their marriage lasted at least 20 years (Steuerle and Spiero 1999).

197 In case they divorce multiple times, women are entitled to the highest benefit among all husbands.

of divorced spousal benefits from the ex-husband. Women are the main beneficiaries of this benefit type in the U.S. (Dickert-Conlin and Meghea 2008).

Germany introduced the property settlement of pension entitlements as an element of the new divorce law that came into effect in 1977 (Bundesgesetzblatt I vom 15. Juni 1976).[198] This new law considered the financial effects on both partners and introduced the splitting of pension rights earned within the statutory pension insurance during the marriage. The partner that earned higher pension rights has to transfer half of the difference in entitlements to his/her former spouse. In practice, women are the principal beneficiaries of pension splitting. Typically, women receive premiums, whereas men face a deduction in pension rights. The pension splitting is one way to compensate women for their role as primary caregivers (Mayer and Wagner 1996) at no cost for government. Ex-spouses of high earners who had no or only low earnings at their command during the marriage receive larger premiums than ex-spouses with roughly equal earnings. The German pension splitting allocates pension rights equitably between ex-spouses at no additional cost for the government. Given that the splitting takes place right upon divorce, premiums and deductions remain unaffected by remarriage and are not conditioned on the retirement or death of the ex-spouse.

Social security systems in both countries provide benefits to surviving spouses in case an insured worker or entitled retiree dies. In the U.S., the social security system pays 100 percent of the deceased workers PIA as soon as the widow(er) reaches the full retirement age. Divorced survivors

198 East and West Germany differed with respect to their divorce legislation. Following reunification, West German divorce law replaced East German law. The new law came into effect starting October 3, 1990. Prior to reunification, divorced East German women were only entitled to two years of alimony payments (Boele-Woelki et al. 2004).

may equally qualify for the surviving spouse benefit in case the ex-spouse deceased. This benefit also amounts to 100 percent of the ex-spouses PIA if the marriage lasted longer than ten years and the surviving spouse did not remarry prior to age 60. For surviving spouses the same rules for dual entitlement apply as for married spouses.

In Germany, the surviving spouse qualifies for survivor's benefits if the deceased spouse reached a minimum qualification period of at least five years of contributory and non-contributory periods or already collected a retired workers pension. If the surviving spouse is 45 years or older, he or she qualifies for the major widow(er)'s pension (*große Witwenrente*) that amounts to 60 percent of the deceased spouse's pension benefit.[199] Unlike U.S. widows, the surviving spouse in Germany receives the widow(er)'s benefit immediately and does not have to reach the full retirement age (Immergut et al. 2007). However, own income – whether a retired workers pension or labor income – counts against the survivor's pension. German and the U.S. also differ with respect to the payment of survivor's pensions in case of remarriage. Women in Germany forego their right to a survivor's benefit if they marry again, whereas women in the U.S. are still eligible for survivor's pensions. *Table 37* summarizes the U.S.-German differences in benefits eligibility and marital status.

[199] There is also a minor widow(er)'s pension for women who are younger than 45 years. The minor benefit amounts to 25 percent of the entitlement of deceased spouse. For survivor's to qualify for the benefit, the marriage had to last at least one year. Since January 1st 2002, the survivor's pension amounts to only 55 percent of the deceased husbands benefit, but the surviving spouse receives additional credits in case of children (Deutsche Rentenversicherung 2011).

Table 37 Marital Status and Eligibility for Social Security Benefits in Germany and the U.S.

Marital Status	Germany		U.S.	
	Retired-Worker Benefit	Auxiliary Benefit	Retired-Worker Benefit	Auxiliary Benefit
Never Married	Must have at least 5 years of contributory or non-contributory, but pension-relevant periods	None	Must have at least 40 quarters of covered employment	None
Married	Must have at least 5 years of contributory or non-contributory, but pension-relevant periods	None	Must have at least 40 quarters of covered employment	Up to 50 percent of spouse's PIA if spouse is still living and retired
Divorced	Must have at least 5 years of contributory or non-contributory, but pension-relevant periods	Splitting of pension rights accumulated by both partners during marriage	Must have at least 40 quarters of covered employment	If spouse living, benefits for married women apply; if spouse deceased, widow benefits apply. Only if married for at least 10 years and no remarriage, otherwise none.
Widowed	Must have at least 5 years of contributory or non-contributory, but pension-relevant periods	55 percent of deceased spouse's benefit if widow above age 45 and at least one year of marriage, otherwise 25 percent. Foregoes eligibility upon remarriage.	Must have at least 40 quarters of covered employment	Up to 100 percent of deceased spouse's PIA at full retirement age

Source: For the U.S. Tamborini et al. (2009); For Germany Author's illustration

5.3.4 Welfare States and Incentives

Following Esping-Andersen's seminal work on the varieties of welfare capitalism, Germany and the U.S. fall into two distinct welfare state clusters (Esping-Andersen 1990; Esping-Andersen 1999). While the U.S. is prototypical for the liberal welfare state, Germany represents the conservative type. Each welfare state brings about a unique pattern of social stratification with differences in socio-economic outcomes and employment patterns (DiPrete and McManus 2000; Gangl 2006). Not only do the U.S. and German regime differ with respect to the type and level of benefits provided, they also put a distinct emphasis on what role both government and the market play (Uunk 2004). Furthermore, welfare state institutions set incentives that determine the extent to which men and women engage in the market and in home production (Misra et al. 2007).

In the early 1990s, Fraser identified the crumbling gender order as one of the driving forces behind the crisis of the welfare state (Fraser 1994). Welfare state institutions being based on the classic male breadwinner – female caregiver notion were out of tune with people's real lives leading to inadequacies in social protection. The clear-cut division of labor within the family was abandoned and replaced by more modern welfare state strategies. Today, women are expected to be both - earners and carers, while the role of men remained largely unchanged.

Following Fraser, Misra et al. distinguish four welfare state strategies that deviate more or less clearly from the male breadwinner – female caretaker model (Misra et al. 2007). The *carer strategy* comes closest to the classic model of gender division, expecting women to be caregivers in the first place and wage earners in the second. Under the *earner strategy*, priorities are reversed in that employment comes first and care giving in the second place. The provision of care services by the state is inadequate under both regimes. The *universal breadwinner* and the *caregiver parity* model follow a new gender order. The first strives for gender equity in the labor market by promoting women's opportunities and leaving the provision of care to government. The latter supports a gender egalitarian division of caretaking and accordingly working. The gendered assumptions embedded in each welfare state strategy are reinforced by a corresponding set of policies and practices.

Germany follows the *carer strategy* with a mix of tax and family policies that contribute to a weak labor market attachment of married women, in particular married women with children.[200] Joint income taxation lacks labor supply incentives for married women, but favors single-earner families with one stay-at-home parent (Steiner and Wrohlich 2004).[201] Family policies set strong incentives for mothers to withdraw from the labor market for a significant amount of time with generous parental leave policies that grant mothers long-lasting job guarantees (Misra et al. 2007). With the payment of parental leave benefits (*Elterngeld*)[202] and child allowances (*Kindergeld*) an immediate return into the job is no economic necessity.[203] Even the public pension scheme compensates mothers for their labor market withdrawal through child care credits and subsidies for those working on low pay or for few hours while raising small children (Rasner 2006).[204] At

200 The *carer strategy* is laid out in Article 6 of the Basic Law of the Federal Republic of Germany, according to which "marriage and the family enjoy the special protection of the state", and further "the care and upbringing of the children is the natural right of parents a duty primarily incumbent on them" (Federal Republic of Germany 1949).

201 It is not necessarily the presence of children in the household that reduces the labor supply of married women in Germany. Drobnic et al. show that marriage has a negative impact on women's labor market attachment even if it is not linked to motherhood (1999).

202 The parental leave benefit introduced in 2007 is a wage-dependent benefit paid for a maximum of 14 months, if the father stays at home for at least two months, otherwise eligibility ends after one year. The parental leave benefit replaced the means-tested child rearing allowance (*Erziehungsgeld*) that was paid for a maximum of 24 months (Spieß and Wrohlich 2008).

203 Stier et al. argue that the interruption of employment doesn't appear to be costly to women, since they are typically working in secondary jobs (2001).

204 The subsidies for child care credits are for mothers that earn less than 75 percent of the average wage earnings. This benefit is clearly directed towards part-time working

the same time, little action is taken to enable mothers to reconcile work and care responsibilities through adequate provision of public or private care services (Drobnic et al. 1999), especially for women with children below age three (Köppen 2010; Kreyenfeld and Geisler 2006) This explains why employment patterns of married women in Germany are tied much closer to their family life cycle than in most other countries (Blossfeld and Rohwer 1997) leading to a strong economic dependency on the husband's (family) wage (Stier et al. 2001). It is important to point out differences in the welfare state conceptions between East and West Germany. The former East German regime promoted dual earner couples and provided sufficient government-run child care facilities, which resulted in a much stronger labor market attachment of East German mothers that still prevails today (Hanel and Riphahn 2011).

In contrast, the U.S. falls in the cluster of welfare states that follow an earner strategy. Even though the income tax system is marriage-centered as well, U.S. practices and policies aim at treating women as economic equals relative to men (Sainsbury 1999). While the government's role is limited to providing equal opportunities for both sexes in the workplace, market criteria determine whether women operate in the market place or engage in home production (Stier et al. 2001). According to the U.S. welfare state strategy, the primacy of the market renders family-oriented policies and practices that induce married women to work superfluous. This orientation explains the lack of adequate state provided child care services (Misra et al. 2007) or the absence of other forms of compensation for limited labor supply while raising small children. In accordance with the strong market principle, the lack of public child care is compensated for by the broad availability of private sector child care arrangements (Stier et

mothers, whereas these subsidies barely pay-off for full-time working mothers.

al. 2001).[205]

In theory, both countries' welfare state regulations are gender neutral in that they don't explicitly assign the carer role to the wife and the earner role to the husband. However, in practice gender inequalities exist in Germany and the U.S., with women working and earning less, staying at home more often and interrupting employment because of care responsibilities.

5.3.5 Research Hypothesis

Based on the roles assigned to women in each respective welfare state, we expect significant differences in retirement outcomes (average public pension benefits and total retirement income) as well as in the process of pension building. Married women in Germany are caught in the *carer regime*, primarily meeting their obligation as caregivers and facing significant disincentives to take up employment, whereas for women in the U.S. *earner regime*, market work takes priority over caregiving. In both social security systems, retirement benefits reflect the individual's earnings history and pension entitlements are a proxy for a person's life cycle labor market attachment. Because of the *carer strategy* embedded in the German welfare state, married women's labor market attachment is almost inevitably weak. The proportion of part-time work and intermittent employment is significantly higher among married women in Germany when compared with the U.S. (Drobnic et al. 1999; Stier et al. 2001). The marginal labor market involvement and accordingly low pension entitlements of married women in Germany, don't pose a policy challenge if the couple stays married till death do them part. The earner orientation of the U.S. welfare

[205] Even if private sector care is much more prominent in the United States, it is not necessarily affordable for every household (Gornick and Meyers 2003). In contrast, private sector child care is rather uncommon in Germany (Evers et al. 2005).

state results in a stronger labor market involvement of married women. Part-time work or longer periods of labor market withdrawal are rather uncommon, even for women with smaller children. Differences in female labor supply according to marital status are therefore far less pronounced in the U.S. than in Germany, and employment patterns only weakly attached to the family life cycle.[206]

Hypothesis 1: Marital trajectories matter in both countries when it comes to retirement outcomes and the process of pension building. The longer women are married, the lower the pension entitlements they accumulate. The negative effect of being married will be stronger in Germany than in the U.S.

Not only do these differences in welfare state strategies affect pension building in the public pension pillar, but also the access to private and occupational pension schemes (Ginn 2003; Ginn and Price 2002).[207] With the shifting emphasis from pension building being an insurance against income loss to a method for developing assets, the coverage with private and occupational pensions gains importance (Shuey and O'Rand 2006). Given that participation in private and occupational pension schemes highly correlates with labor market involvement, we expect:

Hypothesis 2: Access to private and occupational pension schemes varies by marital history. The longer women in Germany are married, the lower the extent to which they benefit from private and occupational schemes

---

[206] In Germany, the marital union enjoys the special protection of the state, while U.S. public policies are neutral towards family formation and dissolution (DiPrete and McManus 2000). The German welfare state favors marriage over other family forms by granting additional social rights to non-working spouses and their children, such as health insurance or survivor's benefits (Berghahn 2003).

[207] For the UK, Ginn and Price show that private pension coverage is highest for full-time employees in their thirties, which coincides with the time of family formation for the majority of young couples and hence, a weaker labor market attachment of women (Ginn and Price 2002).

when compared to their U.S. counterparts.

How do women in Germany and the U.S. fare in terms of retirement outcomes and the process of pension building when they experience a divorce? During their marriage, wives benefit - albeit to a different extent in both countries - from a household income mainly provided by their husband. The adverse financial effects of divorce are therefore stronger for women (Burkhauser and Duncan 1989) and it takes them much longer than men to return to the level of pre-divorce material well-being (Burkhauser and Duncan 1988). In spite of alimony and child support payments[208] from their ex-partners, women are required to work in order to make a living and to accrue their own pension rights unless they opt to remarry.

The distinct welfare state strategies followed in Germany and the U.S. can provide an indication of how women differ in their coping strategies in the aftermath of a divorce. For German women, divorce not only sets an end to the strong reliance on the husband's economic resources. It also forces them to give up their role as primary caregivers and to step up in their role as earners. At the same time, care duties remain the same with adequate state-provided care support lacking (Gornick and Meyers 2003). This shift in responsibilities implies that divorced women have to become earners under adverse conditions.[209] In contrast, U.S. women are expected to be primarily earners whether they are married or divorced. The earner orientation implies a weaker economic dependency on their husband and gives them a stronger financial autonomy. Hence, divorce doesn't trigger a

208 Burkhauser et al. stress that the U.S. is less successful in the enforcement of alimony and child support payments when compared to Germany (Burkhauser et al. 1991).

209 The German welfare state strongly preserves status differentials. Two tiers of welfare state provision draw a clear line between those inside and outside the labor force with a different set of benefits (means-tested vs. social insurance benefits) being associated to each respective status (DiPrete and McManus 2000).

change in the roles of U.S. women. They remain earners in the earner regime.

The length of marriage effect persists even after the marital break-up. Following human capital theory, periodic separations from the labor market as well as part-time work, both of which are more prevalent in Germany, lead to a depreciation of general and specific human capital and consequently to a decrease in the person's earning capacity (Mincer and Ofek 1982; Polachek 1975). The longer the break, the larger the decline in wages at reentry and the longer the restoration phase to get back to the level of exit wages (Mincer and Ofek 1982).[210] Part-time work restricts women's economic opportunities as well and results in a part-time pay penalty letting them fall behind the wages of women working full-time.[211]

Hypothesis 3: The stronger depreciation makes it more difficult for German women to catch up in pension building following a divorce and takes them longer to make up for their reduced labor supply during marriage. Also, a larger share of divorced women is not able to catch up at all.

5.4 Data & Analytic Approach

This study employs two unique datasets well designed to study the accumulation of pension rights and retirement outcomes as determined by the work and family choices of women aged 50 to 80 in Germany and the U.S. Both datasets are unique in that they link survey data with administrative pension records. These linked data have never been used in a com-

---

[210] Mincer and Ofek stress that the process of depreciation and restoration is restricted to general human capital of intermittent workers. The loss of job-specific capital is considered to be a *once-for-all* phenomenon due to the separation from the job (Mincer and Ofek 1982).

[211] Among other things, the gap is associated with differences in the type of jobs held by part-time and full-time working women (Manning and Petrongolo 2008) providing only restricted access to job training for part-timers.

parative study before.[212]

German data come from a link of the German Socio-Economic Panel Study (SOEP) with the Sample of Active Pension Accounts (SAPA) maintained by the German Social Security Administration. The SOEP is a broad interdisciplinary household panel study that started in 1984. Today, 26 waves of data are available that cover a representative sample of the total population living in private households in Germany. Most importantly, the data provide detailed information on retirement income, but also extensive information on individual's education and work histories (for more details about the data see Wagner et al. 2007). SAPA data covers a one percent sample of all active pension accounts. These administrative records provide unusually strong pension-relevant earnings histories that stretch back to age 15 and provide monthly information on individual earnings. For Germany, we link 2007 SOEP and SAPA data, with the SOEP being the recipient and SAPA being the donor file. The total 2007 SOEP survey population covers 21,232 individuals. Restricting the population to women aged 50 to 80, leaves us with a sample population of 4,777 individuals.

For the U.S., we use data from the Health and Retirement Study (HRS) merged with administrative records from the Social Security Administration. In 1992, the core survey of the HRS started with a sample of 12,656 individuals that was interviewed every other year. The steady-state design of the study requires that a new cohort of respondents populates the sur-

[212] German and U.S. data differ in how survey and administrative information were linked. For Germany, statistical matching was used to match survey and administrative data. This technique doesn't aim at finding the exact same person, but statistically similar individuals in both datasets. For the U.S., record linkage was used to match survey information and administrative records, which links information for identical persons in both datasets. In Germany, record linkage is infeasible for confidentiality reasons. Moreover, no common identifiers are available (for further details see Rasner et al. 2011).

vey every six years to have the sample including all age groups above age 50 (Leacock 2006). Today, the HRS covers more than 26,000 Americans above age 50. Information is collected on their financial situation, retirement, employment, health, and family, etc. For the purpose of this paper, the 2004 HRS core data (wave 7) is merged with restricted data from the U.S. Social Security Administration. The 2004 Permissions: Wage and Self-Employment Income (W2) data covers the HRS respondent's earnings between 1937 and 2003. These earnings are taken from the individual's annual W2 form.[213] From the 16,859 HRS respondents in 2004[214], a total of 7,685 respondents gave permission for their data to be merged with their earnings information. Restricting the population to women aged 50 to 80 leaves us with a sample population of 3,823.[215]

First, the empirical analysis requires a reconstruction of marital trajectories for women in Germany and the U.S. using data from previous waves of data collection in order to have consistent marital histories for each individual starting at age 15. For the older cohorts, these marital histories go from age 15 to 65. Of no interest for the research questions of this paper are information on marital status and marital transitions after age 65, because they have no direct effect on the individual's retirement income and the accumulation of pension income, because the majority of individuals retires at age 65.[216] Marital trajectories of the younger age cohorts go from age 15 to the actual margin of the data, namely 2004 for the U.S., and 2007 for Germany.

---

213 For information to be available, earnings have to be recorded in the Master Earnings file of the Social Security Administration.

214 All respondents from the original AHEAD study that covers individuals born in the year 1923 and earlier were excluded from the sample population, because we are not focusing on the oldest-old.

215 *Table A47* in the Appendix illustrates that respondents who gave permission are not systematically different from respondents who refused the permission for data to be merged.

216 An exception is the transition into widowhood that changes retirement income through the payment of survivor's benefits.

For the analysis of the relationship of marital trajectories and the accumulation of pension entitlements, we split the sample populations in a retirement and a preretirement sample. The retirement sample includes women aged 65 years and higher, whereas the preretirement sample covers females aged 50 to 64 years. For the retirement sample, the level of pension benefits can be directly observed, because the majority of women already receive retirement income. Therefore, the focus of the analysis is on how marital trajectories impact retirement outcomes: First, retirement income from the public pension scheme, namely the statutory pension benefits in Germany and benefits from the Old-Age, Survivors, and Disability Insurance (OASDI) in the U.S.; and second, total retirement income from public, private and employer pensions. For the younger cohorts, we look specifically at how marital choices affect the process of pension building. This part of the study analyzes annual individual pension-relevant earnings and marital trajectories simultaneously.

In order to compare retirement outcomes across similar marital trajectories in the U.S. and Germany, I first use sequence analysis and optimal matching (OM) techniques (Abbott 1995; Abbott and Tsay 2000).[217] Sequence analysis serves the description of marital trajectories in terms of the number of marital states, the order of marital sequences and their respective length. The application of optimal matching serves the purpose of comparing these marital sequences across large numbers of observations. OM is the most suitable technique to detect similarities between marital sequences. In order to compare sequences, optimal matching makes use of the so-called *Levenshtein distance* (Levenshtein 1966). The distance reflects the costs of transforming any given sequence in the data into another se-

[217] For the sequence analysis and optimal matching procedure I make use of a special program *sq.ado* written for the statistical software package Stata. For more details on this program see Brzinsky-Fay et al. (2006). For a more detailed description of the method see Simonson et al. (2011)and Brzinsky-Fay (2007).

quence. The Levenshtein distance calculates the costs based on the *Needleman-Wunsch algorithm* (Needleman and Wunsch 1970). In the calculation, the Levenshtein distance allows for three operations in order to transform one marital sequence into another: substitution, deletion and insertion. In order to calculate the distances between each marital sequence and to identify the minimum costs involved in transforming one sequence into another requires the assignment of costs to each operation. In this application, I choose the default setting of the program routine, a cost of one for insertion and deletion and a cost of 2 for substitution. Based on these assumptions, optimal matching calculates a distance matrix that compares each sequence to every other sequence.

The distance matrix of marital sequences, however, is not meaningful for the analysis. Therefore, a cluster analysis follows that groups marital sequences with similar distances into distinctive clusters. I merge U.S. and German data on marital trajectories to perform the optimal matching and cluster analysis jointly, however keeping the separation of the retirement and preretirement samples. The results allow for an analysis of differences in retirement outcomes and pension building across different clusters of marital trajectories in Germany and the U.S. *Table 38* summarizes the sample year, sample specification, and sample sizes for younger and older birth cohorts.

Table 38 Sample Specification in SOEP-SAPA and HRS-SSA

	Germany	United States
Data source	SOEP-SAPA	HRS-SSA
Year of Data Collection	2007	2004
$\text{Sample}_{\text{Preretirement}}$	Birth Cohorts: 1943 -1957 Ages: 50 to 64 years Sample Size: 2,679	Birth Cohorts: 1940-1954 Ages: 50 to 64 years N: 1,989
$\text{Sample}_{\text{Retirement}}$	Birth Cohorts: 1927-1942 Ages: 65 to 80 years Sample Size: 2,098	Birth Cohorts: 1924-1939 Ages: 65 to 80 years N: 1,834

Source: Author's illustration

5.5 Marital Trajectories and Retirement Outcomes: Comparing Retired Women in Germany and the U.S.

5.5.1 Descriptive Results of Marriage Patterns

Changing partnership patterns might increase the vulnerability of certain demographic groups in terms of insufficient financial resources in old age. Both, Germany and the U.S. experienced massive changes in the patterns of family formation and union dissolution. *Table 39* provides summary statistics for the retirement sample population derived from the reconstructed marital trajectories that reflect differences and similarities in marriage patterns in Germany and the U.S. It reports relevant measures for marital status at age 65, the prevalence of selected marital transitions for first and higher-order marriages as well as information on the duration of marital sequences between ages 15 and 65.

Table 39 Marital Status and Marital Transitions of Women in Germany & the U.S., Retiree Population

Variables	HRS-SSA	SOEP-SAPA
Marital Status at Age 65 (in %)		
Never Married	4.0	6.4
First Marriage	51.7	57.4
Second+ Marriage	16.9	6.6
First Divorce	8.2	8.3
Second+ Divorce	5.0	1.1
First Widowhood	10.4	19.8
Second+ Widowhood	3.9	0.5
Marital Dynamics (in %)		
First Marriage	96.0	93.6
→ Stays Married	53.8	61.3
→ Transition into Divorce	30.4	16.7
→ Transition into Widowhood	15.7	22.0
Second Marriage	58.1	25.7
→ Stays Married	52.4	74.9
→ Transition into Divorce	31.5	19.5
→ Transition into Widowhood	16.1	5.6
Third Marriage	45.4	37.5
→ Stays Married	49.4	88.3
→ Transition into Divorce	41.7	11.6
→ Transition into Widowhood	8.9	0.0
Time Spent in Marital States and Length of Marriage (average number of years)		
Never Married	8.4	11.5
Married	36.7	34.2
Divorced	3.9	2.7
Widowed	1.9	2.5
Length of 1st Marriage (of all 1st Marriage)	32.3	34.9
1st Marriages → Divorce	14.8	16.4
Length of 2nd Marriage (of all 2nd Marriages)	19.0	15.7
2nd Marriages → Divorce	10.9	6.9
Length of 3rd Marriage (of all 3rd Marriages)	11.6	9.1
3rd Marriages → Divorce	6.5	5.4
n	1,834	2,098

Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

The descriptive results show that the proportion of never married women is higher in Germany (6.4 percent) than in the U.S. (4.0 percent). In both countries, the majority of elderly women are married at age 65 either in their first or a higher-order marriage (U.S.: 68.6 percent; Germany: 64.0

percent). However, the share of individuals still married in their first marriage is significantly higher in Germany (57.4 percent) than in the U.S. (51.7 percent).[218] In both countries, the share of first time divorcees amounts to more than 8 percent, whereas the share of women with two or more divorces is higher in the U.S. when compared to Germany (5.0 vs. 1.1 percent). With 19.8 percent the share of first time widows is twice as high in Germany as it is in the U.S. (10.4 percent). This finding might indicate that widows in Germany do not remarry, whereas widows in the U.S. are more inclined to remarry after experiencing the death of their spouse.

The middle panel of *Table 39* provides information on marital dynamics. While almost all women enter a first marriage in Germany and the U.S. (96 vs. 93.6 percent), the countries differ with respect to the share of individuals that stay married, get divorced or widowed. In the U.S., only 53.8 percent of individuals stay married in their first marriage, compared to more than 61.3 percent in Germany. More than thirty percent of first marriages in the U.S. end in divorce, but only 16.4 percent in Germany. In turn, the share of individuals whose first marriage ends in widowhood is lower among U.S. women with 15.7 percent compared to more than 22 percent in Germany. U.S. women do not only stand out because of the higher prevalence of first marriages that end in divorce, but also because of higher remarriage rates. Of all women whose first marriage ends in widowhood or divorce, more than 58 percent in the U.S., but only 26 percent in Germany get remarried. Also in second and third marriages, individuals in Germany are more likely to stay married and less likely to get divorced when compared to their U.S. counterparts.[219]

---

218 *Figure A15* in the Appendix shows the distribution of marital status between ages 15 to 65.

219 Note that the number of observations for women that enter a third marriage in Germany is very small, hence results have to be interpreted with caution.

Despite the greater fluctuations between marital statuses in the U.S., the average time spent in each status reveals only minor differences when compared to Germany. The time individuals spent being never married is 3.1 years shorter in the U.S. than in Germany, which is due to women in the U.S. marrying at a younger age. Regardless of the higher divorce propensity, women in the U.S. spent more time being married than their German counterparts (36.7 years in the U.S. compared to 34.2 years in Germany).[220] In the U.S., the time being divorced is on average more than one year longer than in Germany (3.9 vs. 2.7 years, respectively).[221] Widowhood plays only a minor role, which is due to the fact that widowhood starts to become more prevalent at higher ages.

The higher prevalence of divorce in the U.S. does not go along with a shorter duration of marriages. On average, first marriages last about 32 years in the U.S., and 35 years in Germany. Clearly, first marriages that end in divorce are significantly shorter (14.8 years in the U.S. and 16.4 years in Germany). The average duration of marriages decreases for higher parity marriages in both countries. Differences in the average length of marriages between the U.S. and Germany are negligible.[222]

220 An explanation for this somewhat surprising finding is that women in the U.S. have higher divorce, but also higher remarriage rates. In this case, divorce and widowhood are an interruption of two marriage sequences. In Germany, divorce and widowhood are more often absorbing states, which implies that after a divorce or widowhood experience, women are not as likely to remarry, but remain in this state (Sackmann and Wingens 2003).

221 Note that not all individuals in the sample population experienced a divorce. Even though the prevalence of divorce is higher in the U.S., the average time spent being divorced for those who got divorced is significantly higher in Germany than in the U.S. (17.6 vs. 12.9 years, respectively).

222 On average, individuals marry at a younger age in the U.S. than in Germany (23 years compared to 25 years). The difference in the average age at marriage increases for higher-order marriages. Individuals that enter a second marriage in the U.S. are more than five years younger than in Germany (40 years vs. 45.6 years). For the third marriage this differences increases to more than 7 years (46.8 in the U.S. vs. 53.9 years in Germany).

5.5.2 Clusters of Marital Trajectories

After examining the differences in marital dynamics of the retirement sample, this Section focuses on similarities in marital patterns among women in Germany and the U.S. using cluster analysis to group similar marital sequences based on the timing, order and length of marital episodes. This analysis clusters the marital trajectories of women in Germany and the U.S. jointly. This proceeding allows for the comparison of retirement outcomes across the same marital clusters that are embedded in two different welfare state contexts. The cluster analysis identifies a total of six clusters that are displayed in *Table 40*.[223]

The reference cluster consists of individuals with the lowest attachment to marriage, namely women who were not married during the entire observation period (*never married*).[224] Characterizing for the second cluster (*late spouses*) are two long sequences: never been married and married. Individuals that fall in this cluster spend an average time of 35.7 years in the status never married, and 12.1 years in the status married. Divorce and widowhood sequences are negligible in the second cluster. Cluster 3 covers *early-life divorcees* that spent most of the time between ages 15 and 65 being divorced (29.1 years). Women in this cluster get married around age 22 and spend less than five years being married. The fourth cluster (*mid-/late-life widows*) consists of individuals that experience widowhood between ages 15 and 65. Women in cluster 4 spend an average time of 25 years being married and almost 18 years being widowed. A significant share of women in this cluster experience widowhood at a relatively young age. The fifth cluster covers *late-life divorcees*. These women are married for more than 30 years and spend almost 14 years being divorced. The most dominant cluster contains *continuously married* women (cluster 5). Individuals in this cluster

---

[223] For a graphical display of the six clusters see *Figure A16* in the Appendix.

[224] In the remainder of the paper, I refer to the cluster name given in parenthesis.

enter marriage at a relatively young age and stay continuously married. On average, this cluster is married for 42.1 years. If they interrupt marriage because of divorce or widowhood, they tend to remarry quickly.

Table 40 Clusters of Marital Trajectories and their Prevalence in Germany and the U.S., Retiree Population

	Cluster 1 Never Married	Cluster 2 Late Spouses	Cluster 3 Early-Life Divorcees	Cluster 4 Mid-/Late-life Widows	Cluster 5 Late-Life Divorcees	Cluster 6 Continuously Married
Average Duration ...						
Never Married	51.0	35.7	7.2	7.7	6.6	7.4
Married	0	12.1	4.6	25.1	30.7	42.1
Divorced	0	2.5	29.1	0.5	13.6	0.8
Widowed	0	0.7	0.1	17.7	0.1	0.6
Average Number of Marital Episodes	1	2.4	3.7	3.1	3.7	2.6
Prevalence of Cluster ...						
Germany Share in Percent/(n)	6.4 (85)	1.3 (27)	6.3 (105)	11.9 (209)	3.0 (52)	71.1 (1,620)
U.S. Share in Percent/(n)	4.0 (76)	1.4 (32)	8.0 (156)	8.3 (161)	5.3 (94)	73.1 (1,315)

Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

Table 40 also provides information on the distribution of clusters among women in Germany and the U.S. including the number of observations in parentheses. The continuously married cluster is the most dominant across all groups, but more so among women in the U.S. than in Germany. The higher remarriage rates in the U.S. might be one explanation for this finding. In case of a marital split, women in the U.S. tend to remarry instead of staying divorced. Never married women are more common in Germany than in the U.S., whereas late spouses are equally rare in both countries. Because of the higher divorce propensity in the U.S., it is not surprising that women in the U.S. are more likely to fall in one of the two divorcee clusters. It applies to both countries that the share of early-life divorcees is higher than that of late-life divorce. The pattern of widowhood in mid- or late-life without remarriage is more common among women in Germany than in the U.S. The average number of episodes in each cluster ranges from 1 in the never married cluster to almost 4 different marital episodes in the both divorcee clusters, which underlines that some marital trajectories are more dynamic than others. Overall, the number of observations in each cluster is sufficiently high to analyze marital trajectories and retirement outcomes of women in Germany and the U.S. The cluster covering late spouses is the only exception; hence results have to be interpreted with caution.

5.5.3 Retirement Outcomes and Marital Trajectories

5.5.3.1 Descriptive Findings

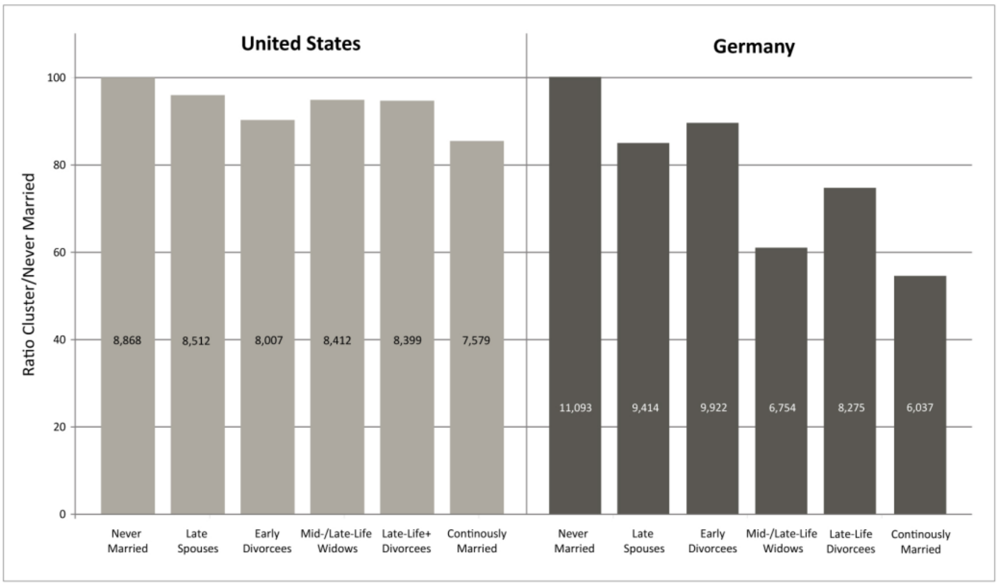
Can we observe differences in retirement outcomes across marital trajectory clusters? This section compares two indicators for retirement outcomes, namely 1) the average social security benefit (*Figure 24*) and 2) average total retirement income[225] (*Figure 25*) across the six marital clusters.[226] The never married cluster is the reference cluster to which the other clusters relate to (see ratio columns). In line with the research hypotheses set out in *Section 5.5.3*, marital cluster differ greatly with respect to their retirement outcomes, with differences being more pronounced in Germany than in the U.S. Overall, women in the never married cluster fare best when compared to women in the other clusters. This finding applies to Germany and the U.S. and holds for social security as well as total retirement income. In Germany, there is a clear tendency that clusters with longer episodes of marriage have lower levels of social security benefits, which does not apply to women in the U.S.

---

[225] The total retirement income includes benefits from the public pension scheme (own retired-workers and survivor's benefits), employers' pensions as well as annuities from private pension funds.

[226] The tables provide annual benefit and income information as of 2004 for the U.S. and 2007 for Germany, respectively. Differences in years of data collection and currencies are of no concern, because the paper's focus is on within- instead of between-country differences.

Figure 24 Average Public Pension Benefit across Marital Clusters in Germany and the U.S.



Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

In both countries, the social security benefit gap is widest between the continuously married and never married cluster. However, with 46 percent the gap is far more pronounced in Germany than in U.S. (15 percent). The strikingly low social security benefits of continuously married women in Germany indicate a weak labor market attachment and go along with a strong financial dependence on their husband. In contrast, continuously married women in the U.S. with low retired worker pensions on their own, benefit from the payment of spousal benefits, whereas women in Germany don't. Unfortunately, the U.S. data does not allow for a separation of benefits from own contributions and benefits from the (deceased) spouse or ex-husband of dually entitled women. However, statistics from the Social Security Administration provide valuable insights.[227] These numbers indicate that about 28 percent of women receive

[227] Of all women aged 65 to 80 in 2004, more than 70 percent are entitled as workers and 30 percent as wives only (14 percent as wives and 16 percent as widows). Of the women who are entitled as workers, 61 percent are entitled as workers only, 22 percent receive a combined retired-worker benefit and secondary spousal benefit and 17

spousal benefits, either as a combined or exclusive benefit. The benefit amounts of these women are below average with the retired-worker benefit being much higher than the secondary spousal benefits. Hence, spousal benefits are not the only explanation for the better financial position of continuously married women in the U.S., but also their stronger labor market attachment.[228] The on average significantly higher age at first marriage of women in the late spouses cluster in Germany makes them more similar to never married women. In the U.S., late spouses fare even better because they are potentially eligible for spousal benefits on top of their own social security benefit.[229] For the U.S., women in the mid-/late-life widows cluster benefit from the payment of survivor's benefits, which explains why average social security benefits of this cluster come close to the benefits of never married women in the U.S. Dually entitled widows receive almost $1,200 in monthly social security benefits (Social Security Administration 2006). The average benefit of the German cluster of mid-/late-life widows is only slightly higher than that of continuously married women. They receive 60 percent of never married women. If survivor's benefits were included, the average social security benefit of mid-/late life widows in Germany would increase from 6,755 to 12,410 Euro, topping the average benefit of the never married cluster by 12 percent.

Figure 24 reveals interesting differences in the two divorcee clusters. When compared to the cluster of never married women, early divorcees fare better than late divorcees in Germany, whereas it is the other way around in

---

percent receive a combined retired-worker benefit and secondary widow's benefit. Women eligible for worker benefits only, receive on average $800, whereas dually entitled wives receive an average benefit of $570 with 66 percent of this benefit coming from own contributions and 34 percent from the spousal benefit. Women who receive spousal benefits exclusively receive an average payment of $480 (Social Security Administration 2006).

228 The analysis on pension building in the preretirement sample will shed more light on this question.

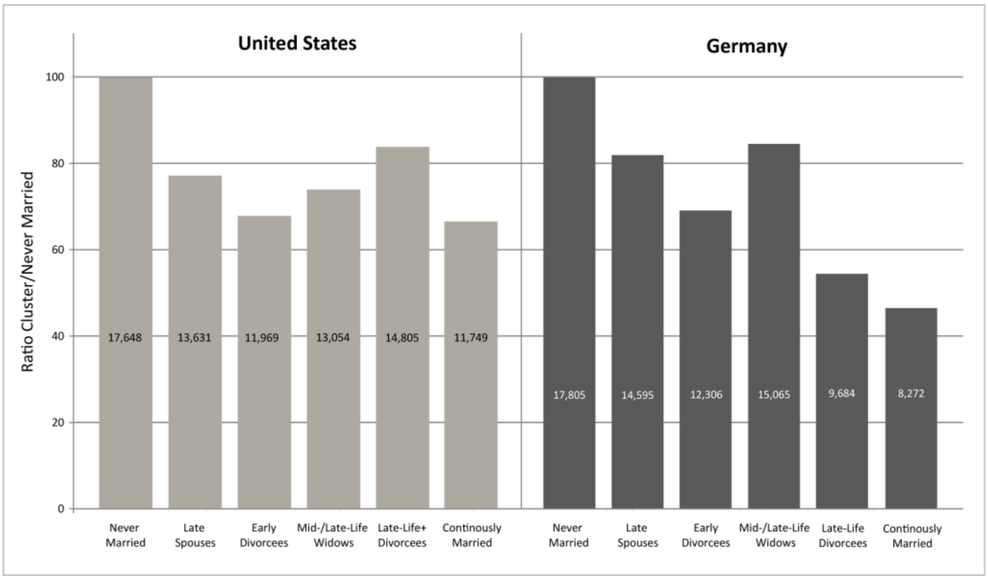
229 However, if late spouses have high retired-workers benefits they might not be eligible for additional spousal benefits.

the U.S., even though differences are rather negligible. For Germany, findings for the two divorcee clusters highlight again that benefit levels decrease with an increasing duration of marriage. Women who experience divorce early and then stay divorced have to provide for themselves because they can no longer rely on the sharing of financial resources with their husband. Hence, an increase in labor supply equally boosts their social security entitlements. In contrast, women who experience divorce later in life, often have little time to catch up in making provisions for retirement if they (partially) withdrew from the labor market during their marriage. The splitting of pension rights between ex-spouses upon divorce results in higher pension rights with divorced women being the primary beneficiaries. However, not all women in this cluster benefit from the splitting because the policy was not introduced until 1977. Late divorcees in the U.S. fare better than their counterparts in Germany. Because the majority of these women are eligible for the receipt of divorced spousal benefits, whereas early divorcees possibly forego eligibility because their marriage lasted less than 10 years and they cannot collect benefits from their ex-husbands record. Late-life divorcees might also benefit from the even higher widow's benefits if their ex-spouse deceased.

Figure 25 compares results for the total retirement income of the six clusters. The overall rank order within each country stays more or less the same after the inclusion of retirement income from employer's pensions and annuities. Across the board, U.S. women benefit to a greater extent from this inclusion than German women, who obviously have limited access to additional forms of old-age provision. For the U.S., differences between the clusters become more pronounced when looking at total retirement income instead of social security benefits alone. Never married women are still on top and continuously married women at the bottom of the distribution with the gap between the two clusters being much larger (33 percent). Taking the total retirement income into account almost doubles the benefits (+99 percent) of never married women indicating good coverage with employer's pensions and annuities. In contrast, mid-/

late-life widows lose ground when compared to the other clusters. Income from employers' pensions and annuities increases the total retirement income by 55 percent. The same applies to the cluster of continuously married women. Both clusters – mid-/ late-life widows and continuously married - don't gain much through the inclusion of occupational and private pension funds, which illustrates that the majority of women in these clusters not only have low social security benefits, but also insufficient coverage or access to occupational and private pensions.

Figure 25 Average Total Retirement Income across Marital Clusters in Germany and the U.S.



Source: HRS Permissions 2004 and SOEP 2007; Author's Calculations

Late-life divorcees come closest to the average total retirement income of never married women in the U.S. The relatively favorable position of this cluster is mainly due to the high average social security benefits, but on top they also have better access to additional retirement income when compared to the widows and continuously married cluster (+ 76 percent). Early divorcees have a less favorable position. Their total retirement income goes up by only 50 percent when additional sources of old-age income are included. Women who divorce at a young age are more likely to have young children to care for, which largely limits their employment capacities, especially their ability to work full-time. Hence, care responsibilities also have repercussions for pension building in the public pension scheme as well as for the access to employer's and private pensions.

The distribution of total retirement income across marital clusters in Germany reveals even larger disparities. Continuously married women have only 46 percent of the total retirement income of never married women. With the inclusion of employers and occupational pensions, continuously married women gain only 37 percent, whereas never married women gain 61 percent from these additional forms of old-age provision. Mid-/late-life widows benefit most. This improvement is not due to their good coverage with employer's pensions and annuities, but mainly because of the inclusion of survivor's benefits. Otherwise, their gain from total retirement income would be even lower than the levels of continuously married women. Unlike late-life divorcees in the U.S., the German counterparts fall way behind. They gain only 17 percent through the inclusion of occupational and private pensions. Late-life divorcees in Germany cannot rely on the payment of divorced spousal or widow's benefits as do women in the U.S. After being married for most of time between age 15 and 65, they divorce at a relatively high age. This situation does not give these women enough time to catch up in pension building. If they had not started an occupational or private pension plan while being married, they face substantial access barriers to these sources of old-age provision after divorce. Early-life divorcees face a better situation than late-life divorcees, which is mainly due to their high social security benefits and not because of their better coverage with employer's and private pensions. For example, early-life divorcees have more in social security benefits than late-life divorcees have in total retirement income. With the inclusion of additional retirement income, early-life divorcees gain another 24 percent on top of their social security benefit.

These first descriptive findings confirm the hypothesis that marital trajectories have a significant impact on retirement outcomes in both countries. The effects are much more pronounced in Germany, where the strong carer notion embedded in its welfare state sets strong disincentives for married, especially continuously married women to work and make provisions for their own retirement. In the U.S., the picture is less clear-cut. The less marked differences might be either due to the higher labor market attachment of U.S. women irrespective from their marital choices or due to the potential availability of spousal and divorced spousal benefits as well as generous survivor's benefits that extenuate differences across marital clusters. The multivariate analysis might shed more light on this question.

5.5.3.2 Multivariate Analyses

By and large, the multivariate analyses confirm the descriptive findings. This section presents results of two multivariate OLS regression models for Germany and the U.S. with annual public pension benefits being the dependent variable.[230] Unlike other income measures, the distribution of public pension benefits is not largely skewed in both sample populations. Therefore, there is no need for a logarithmic transformation of the dependent variable.

The main explanatory variables are the six marital clusters with the never married cluster being the reference cluster. Two dummy variables indicate whether a woman experienced a divorce or widowhood if they do not fall

---

[230] By definition, public pension benefits have a lower and upper limit with the lower limit being equal to zero. The maximum contribution ceiling in Germany and the taxable maximum in the U.S. not only limits the amount of contributions and payroll taxes an individual pays, but also mark the upper limit of pension payments. This limited dependent variable calls for a *tobit* instead of an *OLS regression*, because OLS is likely to produce inconsistent estimates (Wooldridge 2002). However, robustness checks did not reveal substantial differences in coefficients and standards errors from OLS or tobit regression, which justifies the presentation of OLS results in this paper. The lack of differences in standard errors might be due to the fact that only a small share of women reaches the upper limit of pension benefits, because they typically earn less and have less continuous employment careers. In contrast, the public pension benefits of men will have positive probability mass at one or more points of the distribution.

in the widow and the two divorce clusters.[231] The model also controls for post-retirement changes in marital status, namely the transition to widowhood. This control variable should have a strong positive effect in the U.S., but not in the German model.[232] The variable indicating whether the spouse receives a social security benefit also takes into account that it is impossible to separate own retired-workers benefits from spousal benefits. The variable number of children and its interpretation is straightforward: the more children a woman gave birth to, the lower her public pension benefit. The categorical variable educational attainment takes differences in the educational system in Germany and the U.S. into account. The German variable for educational attainment distinguishes six categories: lower secondary education (*Hauptschule*), medium secondary education (*Realschule*), A-levels (*Abitur*), college degree, no degree, degree unknown. The U.S. variable for educational attainment distinguishes five categories: less than high school, General Education Development (GED), high school, some college, college degree.[233] The model also controls for vocational training, which is a particularly relevant for Germany where the institutionalized vocational education system serves as a bridge between school and labor market (Brzinsky-Fay 2007). Because social security systems in both countries are employment-centered, the models control for the years of employment, but also whether a person is still working.[234] The U.S. model controls for race and the German model for migration history and

---

231 This situation pertains to marital trajectories where divorce or widowhood constitutes an interruption of two marriages sequences, but no permanent state. Both variables are set to zero for women who belong to the *early* or *late-life divorcee* as well as the *mid-/late-life widows* cluster.

232 Remember that the optimal matching only considers marital status information between ages 15 to 65. A transition to widowhood after age 65 makes women in both countries eligible for survivor's benefits. However, German data allows for a separation of own and survivor's benefits, whereas U.S. data doesn't. Hence, the coefficient of the variable *post-retirement widowhood* gives an indication to what extent women benefit from survivor's benefits on top of their own benefit in the U.S.

233 Please note these differences in the educational attainment categories in *Table 41*.

234 Restrictions in HRS data don't allow for more detailed measures of the women's employment biographies.

whether a woman lived in East or West Germany at the time of German reunification, because I expect systematic differences between these groups (Hanel and Riphahn 2011). *Table 41* presents coefficients and standard errors that come from two separate estimations, but are arranged in one table to provide a better overview.

The results confirm that marital trajectories matter more in Germany than in the U.S. When compared to the never married, continuously married women in Germany accumulate significantly less in public pension benefits. The same is true for mid- and late-life widows who fare only slightly better than continuously married women. Both clusters are highly financially dependent on their husband's retirement income or in case he dies on the resulting survivor's benefits. Pension benefits of women in the two divorcee clusters and the late spouses cluster are not significantly different from the never married. In the U.S., only the cluster of continuously married women fares significantly worse than never married women. The models include further controls concerning the women's marital history. In the descriptive analyses, it was impossible to separate own retired-workers benefits from spousal or survivor's pensions in the U.S. Controlling for widowhood after age 65 provides a rough estimate of what women gain from receiving survivor's benefits on top of their own pension. The positive significant effect of widowhood after age 65 is therefore in line with the expectation given that women are the primary beneficiaries of survivor's benefits. Women in Germany benefit to the same or even higher extent from survivor's pensions. However, German data allows for a separation of both benefit types, which explains why the coefficient is not significant. In contrast, experiencing divorce has a strong positive effect on pension benefits in Germany.[235] One possible explanation is that women benefit from the pension splitting in case of a divorce. However,

[235] Note that the variable is set to zero for women in the two divorcee clusters.

it is also possible that women who once experience a marital split will no longer rely on their husband's financial resources but provide for themselves.

The strong link between marital trajectories and the level of public pension benefits in Germany holds even if the model controls for education- and employment-related variables, the number of children and additional demographic information. The completion of vocational training has a strong positive effect on the level of pension benefits in Germany but not in the U.S. This finding underscores the relevance of this institutionalized system for the school-to-work transition and women's later labor market attachment in Germany. Women in the birth cohorts under investigation were more likely to complete a vocational training than a college degree. For this reason, there is a positive and significant effect of having intermediate secondary education or completed A-levels when compared to women with lower secondary education, whereas a college degree does not bring about significantly higher pension benefit. In the U.S., higher educational attainment results in higher average pension benefits. Hence, women with some college or a completed college degree have significantly higher pension benefits than women who have less than high school. The number of years worked has a strong positive effect on the level of benefits, which is in line with the expectation given that both countries have an employment-centered public pension scheme. The progressive benefit formula in the U.S. might explain the somewhat weaker effect of years worked in the U.S. Working past age 65 has a negative significant effect in both countries. The coefficient might capture two phenomena: First, women who have to work past age 65 because they have insufficient funds for old-age. Or second, women who want to work past age 65 and take partial retirement. Both types of women either did not start to collect pension benefits yet or only draw partial benefits.

Table 41 Determinants of Monthly Pension Benefits of Retired Women in Germany and the U.S.

	Germany		U.S.	
	b	se	b	se
Marital Clusters (Ref.: Never Married)				
Late Spouses	-93.54	(70.35)	-45.54	(69.09)
Early Divorcees	-2.25	(49.23)	32.44	(47.35)
Mid/Late-Life Widows	-167.76***	(44.56)	92.17	(47.24)
Late Divorcees	1.6	(58.86)	1.21	(51.57)
Continuously Married	-243.67***	(39.71)	-90.87*	(46.24)
Experienced Widowhood	35.33	(29.57)	-29.87	(26.63)
Experienced Divorce	147.84***	(27.29)	36.64	(20.32)
Widowed after Age 65	-18.71	(23.68)	289.46***	(29.57)
Number of Children (Ref.: No Children)				
One Child	-56.78*	(27.29)	-59.54	(35.86)
Two Children	-72.22**	(26.36)	-69.21*	(31.83)
Three + Children	-45.54	(27.12)	-82.28**	(29.69)
Educational Attainment[1]				
No Degree	44.51	(42.66)	n/a	
Degree Unknown	5.17	(49.87)	n/a	
Intermediate Secondary School/GED	97.96***	(18.74)	3.05	(37.86)
A-Levels/High school	131.66***	(37.16)	39.72	(21.15)
Some College	n/a		99.72***	(23.81)
College Degree	54.05	(27.77)	93.55***	(26.69)
Received Vocational Training	61.61***	(16.46)	-22.45	(20.23)
Number of Years Worked	10.96***	(0.56)	4.29***	(0.52)
Working Past Age 65	-201.73***	(34.93)	-55.89*	(23.09)
Lived in East Germany	174.77***	(17.02)	n/a	
Migrant	-67.69	(40.68)	21.45	(28.10)
Race (Ref.: Non-Hispanic Black)				
Non-Hispanic White	n/a		61.36**	(22.71)
Hispanic	n/a		-24.87	(35.41)
Other	n/a		-21.98	(54.03)
Constant	425.648***	(42.72)	565.941***	(45.43)
R-Squared	0.377		0.197	
N	2098		1834	

Note: [1]The reference category for educational attainment is lower secondary school (*Hauptschule*) for Germany and less than high school in the U.S. Significance level: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Selected parameter estimates only. Abbreviations: n/a = not applicable. Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

To have children results in significantly lower pension benefits, because women are the primary caregivers in both countries and hence, cut back on their labor supply. In the U.S., results show that the higher the number of children, the lower the monthly benefit. Having three or more children in Germany does not have a significant negative effect on the level of pension benefits. Separate estimations for East and West Germany can explain this somewhat surprising finding.[236] The coefficients for West Germany are negative and significant, whereas those for East Germany are positive and only partially significant. Hence, in the joint estimation of East and West the opposing effects cancel each other out. This outcome provides evidence that the welfare state context matters. West Germany promoted the division of labor and home production with women being the primary caretaker and if working than mostly as co-earners. Institutionalized child care facilities were available for children aged three years and older, but mainly part-time. The former GDR subsidized families with children, but also promoted full-time employment of both husband and wife enabled by the broad availability of all day child care facilities starting for newborn babies. Consequently, neither the marital trajectory nor the presence of children had negative repercussions for the accumulation of pension rights. This background information explains the strong positive coefficient for living in East Germany at the time of reunification. Migration history has no effect in Germany and the U.S. Effects along racial lines are rather moderate in the U.S. Only non-Hispanic white women have significantly higher monthly pension benefits than Non-Hispanic black women. Overall, the selected variables explain more variation in the monthly public pension benefits of women in Germany than in the U.S.

[236] *Table A48* provides the results of separate estimations for East and West German women.

5.6 Marital Trajectories & Pension Building in the Pre-Retirement Cohorts

5.6.1 Marriage Patterns and Marital Clusters

Changes in partnership patterns and increasing divorce rates are even more pronounced in the preretirement cohorts. *Table 42* provides summary statistics.[237] Concerning the current marital status, more than two thirds of women in both countries are married either in their first or a higher order marriage. Women in the U.S. are far less likely to be in their first marriage than women in Germany (43.3 vs. 57.4 percent). The share of women who is divorced for the first time is quite similar, whereas the share of women in a second or higher order divorce is higher in the U.S. (6.8 percent) when compared to Germany (3.6 percent). Obviously, widows are not very common because women in the pre-retirement population are still young.

The empirical evidence concerning the marital dynamics reveals that the prevalence of divorce and remarriage is even higher in the pre-retirement than in the retiree population. In the U.S., more than 47 percent of first marriages end in divorce compared to almost 30 percent in Germany. Relative to the retiree cohorts, the divorce risk increases by 55 percent in the U.S. and 76 percent in Germany. First marriages that end in widowhood are far less common in the preretirement cohorts. Not only divorce, but also remarriage rates are higher among pre-retirement women. In the U.S., remarriage rates were already high for the retiree population (58.1 percent), but in the pre-retirement population the rate amounts to 65 percent. However, in Germany only 25 percent of women in the retiree population remarried after their first marriage ended in a divorce com-

---

[237] For reference information on the retiree population, compare to *Table 39*.

pared to 47 percent in the pre-retirement sample. Hence, remarriage becomes more common in Germany as well. The share of women who stay married in their second marriage is higher in Germany than in the U.S. (70.6 vs. 54.6 percent). Consequently, women in the U.S. are more likely to get a divorce or experience widowhood in their second marriage than women in Germany.

Table 42 Marital Status and Marital Transitions of Women in Germany & the U.S., Preretirement Population

Variables	HRS-SSA	SOEP-SAPA
Marital Status at Age 65 (in %)		
Never Married	5.8	6.3
First Marriage	43.3	57.4
Second+ Marriage	23.4	11.6
First Divorce	13.5	12.6
Second+ Divorce	6.8	3.6
First Widowhood	4.5	7.8
Second+ Widowhood	2.8	0.6
Marital Dynamics (in %)		
First Marriage	94.2	93.7
→ Stays Married	45.9	61.4
→ Transition into Divorce	47.2	29.4
→ Transition into Widowhood	6.9	9.1
Second Marriage	64.8	47.0
→ Stays Married	54.6	70.6
→ Transition into Divorce	37.4	26.8
→ Transition into Widowhood	8.0	2.7
Third Marriage	48.8	22.6
→ Stays Married	58.5	79.7
→ Transition into Divorce	36.2	15.2
→ Transition into Widowhood	5.4	5.1
Time Spent in Marital States and Length of Marriage (Average number of years)		
Never Married	9.0	10.1
Married	27.8	27.5
Divorced	5.0	3.9
Widowed	0.9	1.0
Length of 1st Marriage (all 1st Marriages)	23.2	26.7
1st Marriage → Divorce	12.0	12.0
Length of 2nd Marriage (all 2nd Marriages)	15.2	13.8
2nd Marriage → Divorce	7.8	8.9
Length of 3rd Marriage (all 3rd Marriages)	10.2	9.2
3rd Marriage → Divorce	5.6	7.2
n	1,989	2,679

Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

Differences are negligible concerning the time women spend in different marital states. On average, women spend 10 years being never married and almost 28 years being married. The number of years a women is divorced is on the rise, when compared to the retiree sample.[238] Two factors contribute to the shorter average length of marriages. First, the sample population is younger and the period of observation only goes from age 15 to 50, hence, marriages are shorter. Second, more marriages end in divorce. On average, first marriages that end in divorce last 12 years in the U.S. and in Germany. In the retiree population, the average duration was 14.8 and 16.4 years, respectively. Overall, the results indicate that the propensity to divorce and to remarry is higher in the pre-retirement than in the retiree cohorts (for reference cp. to *Table 39*). These marital patterns are still more prevalent in the U.S. than in Germany, however trends slowly start to converge.

The cluster analysis reflects the changes in marital patterns in the preretirement cohort and reveals interesting differences with respect to the number of clusters, the distribution across marital cluster and the prevalence of marital clusters in Germany and the U.S. First, the sequence analysis and optimal matching results in seven distinct marital clusters, adding one more cluster the so-called *remarriage cluster*. The first marriage of women in the remarriage cluster ends at an early age. After some years being divorced, they remarry and stay married. Because of the higher prevalence of divorce and remarriage in the U.S., this new cluster is more common in the U.S. than in Germany (6.5 vs. 2.8 percent). Second, the analysis brings about significant changes in the distribution of women's marital trajectories across marital clusters. The optimal matching for the preretirement cohorts considers marital status information only for ages 15 to 50, because information on marital status and pension-relevant in-

---

[238] These numbers are not comparable to those of the retiree population because the sample is right-censored.

come is right-censored at age 50 for the youngest birth cohort.[239] *Table 43* illustrates the distribution of marital clusters in Germany and the U.S.

Table 43 Distribution of Cluster of Women's Marital Trajectories in Preretirement Population

Cluster of Marital Trajectories	Germany Share in %	n	U.S. Share in %	n
1_Never Married	4.3	(114)	5.9	(117)
2_Late Spouses	34.6	(927)	22.4	(446)
3_Early Divorcees	2.8	(74)	4.0	(80)
4_Widows	1.8	(48)	2.2	(44)
5_Late Divorcees	7.7	(207)	10.6	(210)
6_Continuously Married	46.1	(1,235)	48.4	(963)
7_Remarriage	2.8	(74)	6.5	(129)

Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

In the retiree cohorts, the continuously married cluster was the dominant marital pattern for more than 70 percent of women in both countries. This finding is also true for the pre-retirement cohorts, but to a lesser extent. In the younger cohorts, more individuals fall in the late spouses cluster reflecting a general trend of women postponing their first marriage to a higher age. The graphical display (compare to *Figure A17* in the Appendix) of the cluster illustrates that late spouses in the pre-retirement sample are not marrying as late as late spouses in the retirement sample, however later than continuously married women.[240] The average age at first marriage of late spouses is 28 years compared to 20 years of continuously married women. More women in the German sample fall in the late spouses cluster, because they tend to marry at a higher age than women in the U.S. (34.6 vs. 22.4 percent, respectively). Permanent widowhood is rare in

---

239 The decision to right-censor age-specific marital status information at age 50 to perform the sequence analysis avoids that the missing years influence the distance between sequences of different length and hence, the cluster solution (Brzinsky-Fay et al. 2006). Changes in the distribution across marital cluster are - at least - partly driven by the shorter period of observation.

240 The increase in the share of women who fall in the *late spouses* cluster is mainly due to the shorter period of observation, which drives the distance measure between the marital sequences.

the pre-retirement cohorts in both countries and much less common than in the retiree cohorts. Late-life divorcees become more common in the pre-retirement cohorts, whereas early-life divorcees are less common when compared to the retiree population. Note that the shorter period of observation also contributes to the changes in the distribution across the two divorcee clusters with the average age at which divorce occurs being significantly lower in the pre-retirement cohort. The share of never married women increases in the U.S., whereas it decreases in Germany. In addition to the changes in the distribution across marital clusters, *Table 43* also illustrates that clusters are less evenly distributed across the two countries of study.

5.6.2 Paths of Pension Building

The analysis now turns to the process of pension building. In both countries, women accumulate pension entitlements through the payment of payroll taxes or insurance contributions from pension-relevant earnings. Taking on this perspective allows us to compare the paths of pension building over the adult's working life across different marital trajectories.[241] The advantage of this perspective is that pension building only reflects pension rights women accrue from own employment, therefore factoring out any kind of divorced (spousal) or survivor's benefits or redistributive elements, such as child care credits or redistribution between high and low earners. It allows for a straight view on how marital trajectories and women's employment interact.

Figure 26 compares the paths of pension building between ages 25 and 50 across marital clusters in Germany and the U.S. Instead of using age-

[241] In both datasets, year-specific pension-relevant earnings are available. For the analysis, these earnings are indexed to 2004 dollars and 2007 Euros, respectively and then transformed into age-specific earnings.

specific nominal earnings, we calculate the women's relative income position for each year, which relates the women's pension-relevant income in year x to the average pension-relevant income in this year.[242] Like in previous analyses, the cluster of never married women is the reference group constantly set to 100 to which we compare all other marital clusters. Plotting the upper and lower bound of the 95% confidence interval of never married women, shows whether differences to the other clusters are statistically significant. Never married women fare best in the pre-retirement sample. Overall, their pension-relevant earnings are higher than for any other marital cluster. This finding applies to Germany and the U.S. In Germany, the path of pension building of never married women differs significantly from the paths of all other marital clusters, except for early divorcees. They come close to the path of never married women. Early divorcees have a good chance to eventually to outperform never married women until retirement. In the U.S., only continuously married women and widows have significantly different paths of pension building when compared to the never married. The other clusters end up within the upper and lower bound of the 95% confidence interval of the never married.

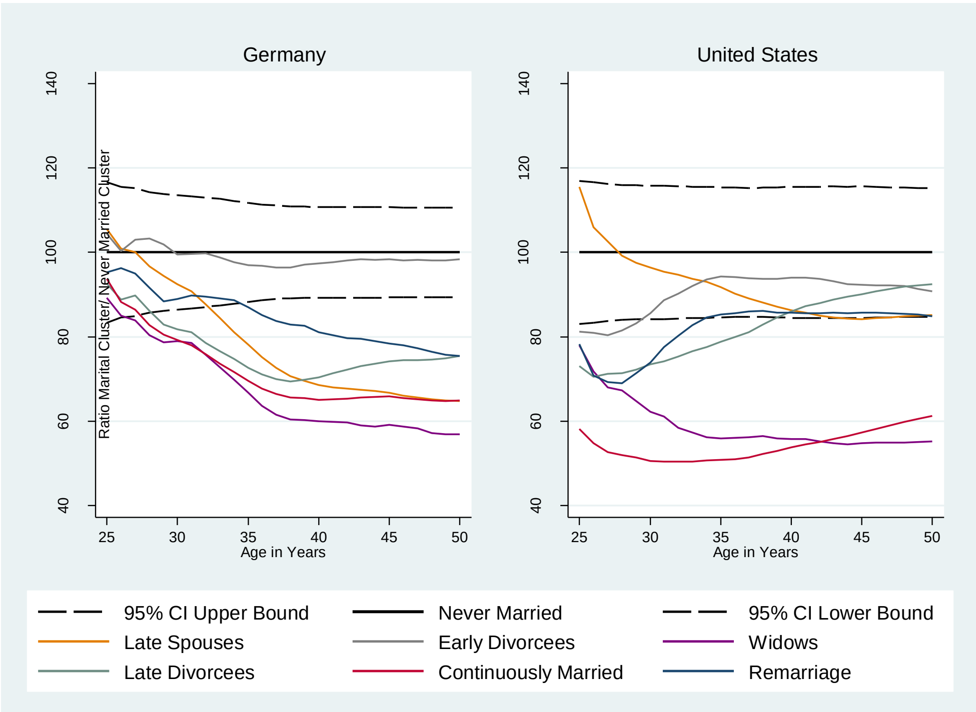
Also in the pre-retirement sample, differences across marital clusters are more pronounced in Germany than in the U.S. However, at age 50, the gap between the top and bottom cluster is the same in both countries with more than 40 percent between widows and never married women. The gap between continuously married and never married women is also quite substantial in both countries: at age 50, they have accumulated slightly more than 60 percent of pension-relevant earnings of never married women. However, there is a clearer upward trend in pension building of con-

---

[242] The year-specific average pension-relevant income comes from the National Average Wage Index in the U.S. and the *Sozialversicherung-Rechengrößenverordnung* (SGB VI Anlage 1) in Germany (Deutsche Rentenversicherung 2010; Social Security Administration 2011).

tinuously married women in the U.S. The path of pension building of late spouses in Germany is similar to that of continuously married women; however, with a five year time shift which is a consequence of the higher age at first marriage. At age 50, late spouses end up at the same level as continuously married women. In contrast, pension-relevant earnings of late spouses in the U.S. exceed those of the never married until the late 20s and then fall below. At age 50, the gap amounts to 17 percent between late spouses and never married women. Their accumulated pension-relevant earnings are within the 95% confidence interval.

Figure 26 Pension-Relevant Earnings between Ages 25 and 50 across Marital Clusters in Germany and the U.S.



Source: HRS Permissions 2004 and SOEP 2007; Author's Calculations

The results for the two divorce clusters illustrate that the timing of divorce matters in both countries. For early divorcees in the U.S., the gap relative to the never married women is largest around the time of divorce.

The same applies to late divorcees in Germany with the gap being significantly larger.[243] Obviously, the smaller gap in the U.S. allows divorced women to catch up more easily in pension building despite the clear upward trend following a divorce in both countries. At age 50, the gap between never married and early divorcees amounts to 9 percent in the U.S. and only 3 percent in Germany.[244] For late divorcees, the gap equals only 7 percent in the U.S., but 31 percent in Germany, which indicates that late divorcees in Germany have trouble catching up in pension building. The later the divorce, the more vulnerable women are and the higher their risk of insufficient provisions for old-age.

Marital trajectories also matter in the U.S. when it comes to pension building. Continuously married women and widows lag considerably behind never married women in terms of pension building. It's the contrast with the retirement cohorts that makes evident to what extent, in particular the groups benefit from the payment of spousal and survivor's benefits. The auxiliary benefits they are eligible for almost close the existing gap between continuously married women and widows on the one side and never married women on the other side. It is also important to remember that women in the U.S. benefit from the best 35 years rule. Only their best 35 years count in the pension benefit calculation. This rule compensates for their weak earnings position between the mid-20s and mid-30s. German women, in turn, benefit from caregiver credits.

5.7 Conclusion

This paper analyzed the impact marital trajectories have on retirement outcomes and the process of pension building of women in Germany and

---

243 According to the summary statistics in *Table A49* in the Appendix, early divorcees experience the marital split in their late 20s and late divorcees in their late 30s.

244 Early divorcees in East drive this result. They have a stronger labor market attachment to begin with. The small number of observations does not allow a

the U.S. The distinction between retirement outcomes and the dynamic process of pension building for retirement and pre-retirement cohorts provided valuable insights into the interplay of individual-level and institutional factors in Germany and the U.S. Differences in retirement outcomes and pension building across marital trajectories exist in Germany and the U.S. with differences being more pronounced in Germany.

For the U.S., the comparison of retirement and pre-retirement cohorts reveals greater differences in pension building among the younger cohorts, whereas differences level off when looking at the retirement outcomes of the older cohorts. Obviously, retired women benefit from the payment of (divorced) spousal and survivor's benefits on top of their own-retired worker benefits. These auxiliary provisions make women less vulnerable to the effects of marital transitions, because they compensate women, at least partially, for their weaker labor market attachment, their larger share in home and care work and the resulting financial dependency on their husband during married life. Hence, the U.S. social security system compensates women for potential disincentives embedded in the welfare state set-up. Women who got a divorce also benefit from these auxiliary benefits in case they were married for more than ten years, because they continue to be eligible for spousal and survivor's benefits despite of the marital split. This rule has two positive implications: First, in general, women have to rely to a lesser extent on the equal sharing of resources with their husbands. Second, these provisions allow women in the U.S. to opt out of marriage even at higher ages, because they do not lose eligibility.

For Germany, we found pronounced differences in the paths of pension building by marital trajectories. These differences also prevail for the retirement outcomes of older cohorts. This finding indicates that the German public pension program has almost no provisions to compensate women for their role as primary caretakers, except for the survivor's benefits they receive if their husband dies. Continuously married women have significantly lower retirement income when compared to women with other marital trajectories, but also relative to their continuously married U.S. counterparts. The lack of effective incentives embedded in the German welfare state that promote married women's labor supply has detrimental effects for successful pension building. The interplay of welfare state settings and individual behavior makes women in Germany particularly vulnerable to marital shocks. This vulnerability grows with every additional year of marriage. Hence, the strong economic dependency on their spouse and the lack of adequate compensating provisions in the German social security scheme might prevent some women who are continuously married from getting a divorce, even though this study does not provide any explicit empirical evidence to support this hypothesis. Possibly, women with low pension benefits on their own stay married because otherwise, they also forego eligibility for survivor's benefits. The empirical evidence illustrates the high economic risk women bear that opt out of marriage later in life. The pension splitting between ex-spouses appears to be a less effective instrument than the divorced spousal and survivor's benefits in the U.S. Late divorcees in Germany have not enough time to catch up in pension building and fail to close the gap to never married women. They still fare better than continuously married women, but typically they have few other income sources to rely on.[245] In turn, women who get a divorce early on in life, succeed to catch up.

From this analysis, it becomes obvious that the German approach to social security simply prolongs the strong financial dependency of women on their husbands into the retirement phase. Hence, continuously married women with low benefits on their own have to rely on the equal sharing of pension benefits with their husband. Even though, evidence on the

[245] Younger birth cohorts will benefit to a greater extent from the pension splitting in case of a divorce. Some women in the retirement cohort did not benefit this provision, because it was only introduced in 1977.

gender wealth gap raises doubts, whether resources in couples are truly equally shared (Sierminska et al. 2010).

The situation of married women in Germany can only improve with targeted work incentives that put them in the position to have continuous employment careers. For divorced women might change significantly over the next years in the light of the new German divorce legislation. This new legislation shifts the emphasis from the welfare of ex-wives to the welfare of children. Hence, the law considers it just and reasonable that women with young children have to work and achieve financial independence. This ruling will also have repercussions for the old-age provisions of divorced women.

Despite of the comparative advantage of women in the U.S. and the less strong impact of marital trajectories on retirement outcomes, it is important to keep in mind that spousal and survivor's benefits in both countries come from general tax revenues. Workers pay no extra payroll taxes or social insurance contributions for spousal benefits in the U.S. and survivor's benefits in both countries. From the perspective of women, the current legislation raises questions as to whether it is safe to rely on spousal and survivor's benefits. Given that these benefits are not financed over additional taxes makes it easier for policymakers to cut spousal and survivor's benefits. In the light of population aging, these benefits can become a substantial cost driver that easily overstrains national budgets. From the perspective of policymakers, the payment of spousal and survivor's benefits raises equity concerns. For example, never married individuals with continuous work histories might end up with lower social security benefits than continuously married individuals or widows with a weak lifetime labor market attachment who benefit from the high retired-workers benefit of their husband. These outcomes raise questions whether a more equitable allocation of benefits is feasible.

Burkhauser and Duncan wrote that "one method by which women can reduce the relative risk of dramatic drops in well-being is to become more like men" (Burkhauser and Duncan 1989, p. 20). This statement clearly refers to women's employment behavior, but dismisses how marital status and marital history affect their labor supply and consequently pension building.

6 Conclusion and Outlook

The overhaul of the German system of old-age provision raises concerns about the future distribution of pension rights. Therefore, policymakers and researchers seek to better understand the distributional consequences of cuts in the generosity of the public pension program (Börsch-Supan et al. 2003), changes in the public-to-private mix (Bonin 2009) as well as gradual increases in the retirement age (Brugiavini 2001; Gruber and Wise 1999) paired with lasting changes in employment biographies (Simonson et al. 2011). In this context, it becomes critical to appreciate: First, what role does social security wealth play in the individual's total wealth holdings? And second, what individual-level or institutional factors facilitate or impede successful pension building that allow individuals to retire with a high enough pension on their own?

These analyses require the use of complex micro data that is not readily available in Germany. While the data infrastructure largely improved over the last years, with the number of data sources on aging steadily increasing (Jürges 2010), none of the existing sources provides a satisfying basis for the research questions of interest (Hauser 2011). While survey data, in particular the population representative *Socio-Economic Panel* (SOEP), stands out for the large sample size, the longitudinal design, and extensive contextual information at the individual and household level (Wagner et al. 2008), it lacks complete life cycle earnings and reliable information on the individual's social security wealth (Frick et al. 2010a). Administrative pension records, in turn, provide these data but fall short of other standard income and wealth categories and household level information (Rasner 2007). Ideally, survey data and administrative pension records could be linked over a unique identifier, so-called record linkage (Winkler 2006). However, record linkage is infeasible because of the absence of a common identifier, but also because it requires the written consent of respondents (Jenkins et al. 2006). A viable alternative to record linkage is statistical matching

(Rässler 2002; Rubin 1986).

In this context, the purpose of this dissertation has been twofold: The major methodological contribution was to augment survey data with administrative pension records by means of statistical matching. Chapter 2 and Chapter 3 tested and prepared the feasibility of the statistical matching and Chapter 4 dealt with its implementation. The second contribution was to use these new and unique matched data to trace the consequences of life cycle work and family choices through to outcomes in old age. Chapter 4 presents the findings of an extended wealth inequality analysis that includes social security wealth information obtained from the statistical matching. Chapter 5 uses complete information on life cycle earnings from pension data to analyze the interdependencies of marital trajectories and pension building for women in Germany and the U.S.

In order to prepare the statistical matching, Chapter 2 gave a systematic account of the strengths and weaknesses of administrative data for the analysis of the material well-being of (future) elderly. This methodological groundwork has been so far neglected, even though it is one of the central recommendations of the German Data Forum concerning the improvement of the research infrastructure in Germany (Rat für Sozial- und Wirtschaftsdaten 2010). Based on evidence from the *Sample of Active Pension Accounts* (SAPA), a dataset maintained by the German statutory pension insurance (Stegmann 2008), Chapter 2 suggested that analyses based on administrative data alone are limited in scope. The key findings are: First, data are representative for all individuals holding a pension account, but not for the total population living in Germany, which limits and impedes statistical inference. Second, special working routines in the administration of the public pension program, especially the process of pension account validation, are a source for sample selection bias. Third, specifics in the data collection process as well as legislative changes directly feed into modifications in the measurement of variables, thereby compromising their validity and reliability. Fourth and most importantly, the lack of relevant covariates in administrative data that go beyond the principal purpose of the agency largely limits the explanatory power of statistical analyses. Enhancing the power of administrative pension records by means of statistical matching with population representative survey data appears therefore like a promising path worth following.

Chapter 3 illustrated the preparatory steps to test the feasibility of a statistical matching of the dataset *Completed Insurance Biographies 2004* (SUF VVL 2004) maintained by the statutory pension insurance and population representative survey data from the SOEP. The main aim of this joint work with Joachim R. Frick, Markus M. Grabka, and Ralf K. Himmelreicher is to elaborate a generic blueprint for the preparation of all sorts of statistical matching projects. This feasibility study gave us a thorough understanding of the compatibility of both datasets in terms of potential matching variables and the correct specification of matching populations. Extensive comparisons of the distributions of matching variables across various subsamples, and if necessary, a further alignment of these variables in order for them to measure functional equivalents have proven that both datasets share a sufficient number of common variables. For a certain subset of these variables, the multivariate regression analysis confirmed comparable underlying relationships in survey and administrative data. Finally, the paper illustrated that it is possible to replicate SOEP results with out-of-sample predictions based on SUF VVL 2004 data for the total population and for smaller subsamples. Overall, these results suggested that the data meet all requirements for the actual implementation of the statistical matching of administrative pension records and population representative survey data.

Chapter 4 links the two strands of research by implementing the statistical matching and using the matched data in a wealth inequality analysis. For the actual implementation of the statistical matching, we turned to the Sample of Active Pension Accounts. The methodological aim of this third paper (joint work with Markus M. Grabka and Joachim R. Frick) was to find the best statistical matching or imputation technique to augment SOEP data with life cycle earnings and information on social security wealth from SAPA data. For this purpose, we compared four statistical matching and imputation techniques for a total of nine matching algorithms. The unique properties of the linked data allowed for a straight control of the quality of matches and the performance of each technique for the group of retirees in both datasets. We identified *Mahalanobis distance matching* (Mahalanobis 1936) to be the best performing technique and applied it to the total population. We also proved that the conditional independence assumption holds (Caliendo and Kopeinig 2008), which allowed for the inclusion of social security wealth in a wealth inequality analysis. The key findings of this application are: First, the inclusion of social security wealth almost doubles the total net worth individuals hold in Germany. Second, social security wealth decreases inequality in net worth by almost 25 percent, which is a direct consequence of the maximum contribution ceiling in the statutory pension insurance that sets an upper limit to social security wealth. Third, the analysis revealed substantial differences across occupational lines. Obviously, all occupational groups benefit from the inclusion of social security wealth, but civil servants benefit most with their wealth holdings increasing by a factor of two. The group of self-employed, especially those with employees, stay on top of the wealth distribution, but they benefit to a greater extent from standard net worth, in particular property or life-insurance policies. Overall, the evidence underscored the added value statistical matching has in order to provide a more complete picture of the wealth distribution in Germany.

Chapter 5 analyzed the relationship between marital trajectories and retirement outcomes as well as the process of pension building of women aged 50 to 80 years in Germany and the U.S. For Germany, this study used the matched SOEP-SAPA data. The reference data for the U.S. came from the *Health and Retirement Study* (HRS) linked with the *2004 Permissions: Wage and Self-Employment Income (W2)*, an administrative dataset maintained by the U.S. Social Security Administration. A split of the analysis population into retirement and preretirement cohorts provided valuable insights into the interplay of individual-level and institutional factors in Germany and the U.S. A joint cluster analysis for women in both countries identified similar marital trajectories that allowed for the comparison of retirement outcomes and the process of pension building across marital clusters. The key findings of this second application are: First, differences in retirement outcomes and pension building across marital trajectories exist in Germany and the U.S. with differences being more pronounced in Germany. Second, women in Germany who get a divorce late in life bear a high economic risk because they forego a sharing of resources with their husband and lose eligibility for survivor's benefits. Their weak labor market attachment during married life makes it difficult for late-divorcees to catch up in pension building. Third, special provisions such as spousal and survivor's benefits level off differences in pension building across marital trajectories. These provisions make married *and* divorced women in the U.S. less vulnerable to marital shocks than their German counterparts. This second application further emphasizes the added value of statistical matching that allows for comparative analyses that were previously infeasible because of the lack of adequate and comparable longitudinal data.

Overall, this dissertation illustrated that statistical matching is of great use when it comes to augmenting survey data with information from other data sources that are difficult to obtain in an interview situation. The statistical matching approach put forward in this dissertation could serve as a generic blueprint for other statistical matching projects that could further improve the quality and extend the applicability and scope of data from the Socio-Economic Panel Study. These extensions could include the statistical matching with health and benefit records in order to track program outcomes and inform the social policy debate. The feasibility of the statistical matching depends on the availability of suitable matching variables.

Clearly, the two applications in this dissertation are far from fully exploiting the research potential of the matched data. The combination of real life cycle earnings trajectories with comprehensive context information at the individual and household level goes far beyond data available to policymakers and provides a good basis for the evaluation of already passed reforms or the simulation of new reform plans. The data are also of great use to test the feasibility of proposals for a redistribution of work (Vaupel and Loichinger 2006). The matched data is well-suited to identify Germany's unused labor market potentials and work out the specifics on how new life-course policies could help to spread work more evenly across the individual's life cycle without risking significant cutbacks in their standard of living.

Future work should direct increased efforts to obtain an even more complete picture of the material well-being of future retirees and the distribution of pension wealth in Germany. Since not all segments of the population are covered under the umbrella of the statutory pension insurance, future research could extend the statistical matching to occupations with separate schemes of old-age provision, such as civil servants, certain chambered professions (e.g. lawyers or medical doctors) or agricultural workers. Moreover, statistical matching could be used to gather reliable and population-representative data on occupational and private pension schemes. With the paradigmatic shift from a strong reliance on the public pension program to a multi-pillar system of old-age provision, these sources of old-age income will gain importance over the next years. It is therefore essential to gain a better understanding of the interplay of public, occupational, and private pensions in Germany at the individual and household level.

The research project *Life Course, Aging and Well-Being* (LAW project) already adopts the statistical matching approach put forward in this thesis using SOEP and modified SAPA data (Simonson et al. 2011) that allows for further refinements of the matching. The project compares the life-courses of the baby boomers to those of older birth cohorts and analyzes the repercussions of less stable family lives and changes in employment trajectories for the accumulation of pension rights.[246]

Apart from applied research, future work should also focus on methodological advancements of statistical matching. These advancements could involve: First, the development of formal instruments for the assessment of the matching quality even in the absence of controls. Second, comparisons of record linkage and statistical matching based on the same data to gain insights in the respective strengths and weaknesses as well as potential trade-offs of each data fusion technique. Third and finally, the development of multiple matching algorithms - similar to methods of multiple imputation (Rubin 1987) - could be an option to obtain better inference for matched data through improved statistical properties.

[246] For more information about the LAW project go to www.law-projekt.org.

References

Abbott, Andrew. 1995. "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology* 21: 93-113.

Abbott, Andrew and Angela Tsay. 2000. "Sequence Analysis and Optimal Matching Methods in Sociology." *Sociological Methods & Research* 29 (1): 3-33.

Andreß, Hans-Jürgen, Barbara Borgloh, Miriam Bröckel, Marco Giesselmann, and Dina Hummelsheim. 2006. "The Economic Consequences of Partnership Dissolution - A Comparative Analysis of Panel Studies from Belgium, Germany, Great Britain, Italy, and Sweden." *European Sociological Review* 22 (5): 533-560.

Andridge, Rebecca R. and Roderick J. A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review* 78 (1): 40-64.

Anger, Silke, Deborah A. Bowen, Michael Engelmann, Joachim R. Frick, Jan Goebel, Markus M. Grabka, Olaf Groh-Samberg, Hansjörg Haas, Bruce Headey, Elke Holst, Peter Krause, Martin Kroh, Christine Kurka, Henning Lohmann, Rainer Pischner, Uta Rahmann, Christian Schmitt, Jörg-Peter Schräpler, Jürgen Schupp, Ingo Sieber, Thomas Siedler, C. Katharina Spieß, Martin Spieß, Ingrid Tucci, and Gert G. Wagner. 2008. "25 Wellen Sozio-oekonomisches Panel." *Vierteljahreshefte zur Wirtschaftsforschung* 77 (3): 9-14.

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2010. "A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark." *Sociological Methods & Research* 39 (2): 256-282.

Arza, Camila. 2008. "Changing European Welfare: the New Distributional Principles of Pension Policy", pp. 109-131. In: *Pension Reform in Europe: Politics, Policies and Outcomes*, edited by Camila Arza and Martin Kohli. London: Routledge.

Aughinbaugh, Alison. 2010. "The Effects of Remarriage on Women's Labor Supply." *Journal of Population Economics* 23 (4): 1151-1176.

Baiocchi, Mike, Dylan S. Small, Scott Lorch, and Paul R. Rosenbaum. 2010. "Building a Stronger Instrument in an Observational Study of Perinatal Care for Premature Infants." *Journal of the American Statistical Association* 105 (492): 1285-1296.

Barr, Nicholas. 2006. "Pensions: Overview of the Issues." *Oxford Review of Economic Policy* 22 (1): 1-14.

Barr, Nicholas and Peter Diamond. 2006. "The Economics of Pensions." *Oxford Review of Economic Policy* 22 (1): 15-39.

Bender, Stefan, Ralf K. Himmelreicher, Sylvia Zühlke, and Markus Zwick. 2008. "ForschungsDatenZentren – Eine informationelle Infrastruktur mit erFreulichem DatenZugang", pp. 147-158. In: *Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag*, edited by Gabriele Rolf, Markus Zwick, and Gert G. Wagner. Baden-Baden: Nomos Verlagsgesellschaft.

Berghahn, Sabine. 2003. "Der Ehegattenunterhalt und seine Überwindung auf dem Weg zur individualisierten Existenzsicherung", pp. 105-131. In: *Wohlfahrtsstaat und Geschlechterverhältnis im Umbruch: Was kommt nach dem Ernährermodell?*, *Jahrbuch für Europa- und Nordamerika-Studien*, edited by Sigrid Leitner, Ilona Ostner, and Margit Schratzenstaller. Wiesbaden: VS Verlag für Sozialwissenschaften.

Berner, Frank. 2006. "Riester Pensions in Germany: Do They Substitute or Supplement Public Pensions? Positions in the Debate on the New Public Policy on Private Pensions." *German Policy Studies* 3 (3): 492-492-534.

Blossfeld, Hans-Peter and Götz Rohwer. 1997. "Part-Time Work in West Germany", pp. 164-190. In: *Between Equalization and Marginalization. Women Working Part-Time in Europe and the United States*, edited by H.-P. Blossfeld and C. Hakim. Oxford: Oxford University Press.

Boele-Woelki, Katharina, Frédérique Ferrand, Cristina González Beilfuss, Maarit Jänterä-Jareborg, Nigel Lowe, Dieter Martiny, and Walter Pintens. 2004. *Principles of European family law regarding divorce and maintenance between former spouses*. Antwerpen: Intersentia.

Bonin, Holger. 2009. "15 Years of Pension Reform in Germany: Old Success and New Threats." *The Geneva Papers* 34 (4): 548-560.

Börsch-Supan, Axel. 2000a. "Incentive Effects of Social Security on Labor Force Participation: Evidence in Germany and across Europe." *Journal of Public Economics* 78 (1-2): 25-49.

Börsch-Supan, Axel. 2000b. "A Model Under Siege: A Case Study of the German Retirement Insurance System." *The Economic Journal* 110: F24-F45.

Börsch-Supan, Axel , Anette Reil-Held, and Christina Benita Wilke. 2003. "Der Nachhaltigkeitsfaktor und andere Formelmodifikationen zur langfristigen Stabilisierung des Beitragssatzes zur GRV." *Sozialer Fortschritt* 52 (11-12): 275-284.

Börsch-Supan, Axel, Martina Brandt, Karsten Hank, and Mathis Schröder. 2011. "The Individual and the Welfare State: Life Histories in Europe." Heidelberg: Springer.

Börsch-Supan, Axel and Reinhold Schnabel. 1999. "Social Security and Retirement in Germany", In: *Social Security and Retirement Around the World*, edited by Jonathan Gruber and David A. Wise. Chicago: Chicago University Press.

Börsch-Supan, Axel and Christina B. Wilke. 2004. "The German Public Pension System: How it was, how it will be." NBER Working Paper 10525. National Bureau of Economic Research, Cambridge.

Brenke, Karl. 2007. "Die Bedeutung der Älteren auf dem Arbeitsmarkt nimmt deutlich zu." DIW Wochenbericht 74:337-345. DIW, Berlin.

Breyer, Friedrich and Mathias Kifmann. 2004. "The German Retirement Benefit Formula: Drawbacks and Alternatives." *FinanzArchiv* 60 (1): 63-82.

Brugiavini, Agar. 2001. "Early Retirement in Europe." *European Review* 9 (04): 501-515.

Brzinsky-Fay, Christian. 2007. "Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe." *European Sociological Review* 23 (4): 409-422.

Brzinsky-Fay, Christian, Ulrich Kohler, and Magdalena Luniak. 2006. "Sequence Analysis with Stata." *The Stata Journal* 6 (4): 435-460.

Budig, Michelle J. and Paula England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66: 204-225.

Bundesgesetzblatt I vom 10. Februar 1998. "Verordnung zur Neuregelung des Meldeverfahrens in der Sozialversicherung", Seite 343.

Bundesgesetzblatt I vom 15. Juni 1976. "Erstes Gesetz zur Reform des Ehe- und Familienrechts (1. EheRG)", Seite 1421.

Bundesgesetzblatt I vom 18. Dezember 1972. "Verordnung über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der Bundesagentur für Arbeit (Datenübermittlungs-Verordnung - DEÜV)", Seite 2482.

Bundesgesetzblatt I vom 24. November 1972. "Verordnung über die Erfassung von Daten für die Träger der Sozialversicherung und für die Bundesanstalt für Arbeit (Datenerfassungs-Verordnung - DEVO)", Seite 2159.

Bundesgesetzblatt I vom 26. Februar 1957. "Gesetz zur Neuregelung des Rechts der Rentenversicherung der Angestellten (Angestelltenrentenversicherungs-Neuregelungsgesetz - AnVNG)", Seite 88.

Bundesgesetzblatt I vom 26. Februar 1957. "Gesetz zur Neuregelung des Rechts der Rentenversicherung der Arbeiter (Arbeiterrentenversicherungs-Neuregelungsgesetz - ArVNG)", Seite 45.

Bundesgesetzblatt I vom 26. Mai 1994. "Gesetz zur Absicherung des Risikos zur Pflegebedürftigkeit (Pflegeversicherungs-Gesetz - Pflege VG)", Seite 1014.

Bundesministerium für Arbeit und Soziales. 2006. "Social Security Data Protection (Sozialdatenschutz)." BMAS, Berlin.

Bundesministerium für Arbeit und Soziales. 2008. "Ergänzender Bericht der Bundesregierung zum Rentenversicherungsbericht 2008 gemäß §154 Abs. 2 SGB VI." Pp. 239.

Bundesregierung. 2008. "Lebenslagen in Deutschland: Der dritte Armuts- und Reichtumsbericht der Bundesregierung."228. Bundesministerium für Arbeit und Soziales, Berlin.

Burkhauser, Richard V. and Greg J. Duncan. 1988. "Life Events, Public Policy, and the Economic Vulnerability of Children and the Elderly", pp. 55-88. In: *The Vulnerable*, edited by John L. Palmer, Timothy S. Smeeding, and Barbara B Torrey. Washington D.C.: The Urban Institute Press.

Burkhauser, Richard V. and Greg J. Duncan. 1989. "Economic Risks of Gender Roles: Income Loss and Life Events over the Life Course." *Social Science Quarterly* 70 (1): 3-23.

Burkhauser, Richard V., Greg J. Duncan, and Richard Hauser. 1994. "Sharing Prosperity Across the Age Distribution: A Comparison of the United States and Germany in the 1980s." *The Gerontologist* 34 (2): 150-160.

Burkhauser, Richard V., Greg J. Duncan, Richard Hauser, and Roland Berntsen. 1991. "Wife or Frau, Women Do Worse: A Comparison of Men and Women in the United States and Germany after Marital Dissolution." *Demography* 28 (3): 353-360.

Butrica, Barbara and Howard Iams. 2000. "Divorced Women at Retirement: Projections of Economic Well-Being in the Near Future." *Social Security Bulletin* 63 (3): 3-12.

Caliendo, Marco. 2006. *Lecture Notes in Economics and Mathematical Systems*. Vol. 568. Berlin: Springer.

Caliendo, Marco and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22 (1): 31-72.

Chen, Jiahua and Jun Shao. 2000. "Nearest Neighbor Imputation for Survey Data." *Journal of Official Statistics* 16 (2): 113-132.

Christensen, Kaare, Gabriele Doblhammer, Roland Rau, and James W Vaupel. 2009. "Ageing Populations: The Challenges Ahead." *The Lancet* 374: 1196-1208.

Copas, J. B. 1997. "Using Regression Models for Prediction: Shrinkage and Regression to the Mean." *Statistical Methods in Medical Research* 6 (2): 167-183.

Dannenberg, Andreas, Alena Degtjarjev, and Kalamkas Kaldybajewa. 2008. "Arbeitslosigkeit im Versichertenleben: Zunahme oder Rückgang - Statistische Fakten, Interpretationen und Bewertungen." *RVaktuell* 55 (Oktober): 303-318.

Dehejia, Rajeev H. and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448): 1053-1062.

Deutsche Rentenversicherung. 2005. *Rentenversicherung in Zeitreihen*. Vol. 18. Berlin: Deutsche Rentenversicherung.

Deutsche Rentenversicherung. 2006a. *Rentenbestand am 31.12.2005*. Berlin: Deutsche Rentenversicherung.

Deutsche Rentenversicherung. 2006b. *Rentenversicherung in Zeitreihen*. Vol. 19. Berlin: Deutsche Rentenversicherung.

Deutsche Rentenversicherung. 2007a. "Codebook: Versicherungskontenstichprobe (Stand: 11.09.2009)." Berlin.

Deutsche Rentenversicherung. 2007b. *Rentenzugang am 31.12.2006*. Berlin: Deutsche Rentenversicherung.

Deutsche Rentenversicherung. 2009. *Rentenversicherung in Zeitreihen*. Vol. 22. Berlin: Deutsche Rentenversicherung.

Deutsche Rentenversicherung. 2010. *Rentenversicherung in Zeitreihen*. Vol. Band 22. Berlin.

Deutsche Rentenversicherung. 2011. "Hinterbliebenenrente: Hilfe in schweren Zeiten." Berlin.

Dickert-Conlin, Stacey and Christian Meghea. 2008. "The Effect of Social Security on Divorce and Remarriage Behavior."45.

Diekmann, Andreas and Ben Jann. 2004. "The Erosion of Regular Work. An Analysis of Structural Changes in the Swiss and German Labor Markets." Working Paper, Department of Sociology, ETH Zürich.

DiPrete, Thomas A. and Patricia A. McManus. 2000. "Family Change, Employment Transitions, and the Welfare State: Household Income Dynamics in the United States and Germany." *American Sociological Review* 65 (3): 343-370.

Drobnic, Sonja, Hans-Peter Blossfeld, and Götz Rohwer. 1999. "Dynamics of Women's Employment Patterns over the Family Life Course: A Comparison of the United States and Germany." *Journal of Marriage and Family* 61 (1): 133-146.

Duncan, Greg J. and Saul D. Hoffman. 1985. "A Reconsideration of the Economic Consequences of Marital Dissolution." *Demography* 22: 485-497.

Eichhorst, Werner and Paul Marx. 2011. "Reforming German labour market institutions: A dual path to flexibility." *Journal of European Social Policy* 21 (1): 73-87.

Elliott, Michael R. and William W. Davis. 2005. "Obtaining Cancer Risk Factor Prevalence Estimates in Small Areas: Combining Data from Two Surveys." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 54 (3): 595-609.

Engelhardt, Gary V. and Jonathan Gruber. 2004. "Social Security and the Evolution of Elderly Poverty." NBER Working Paper Series.44. National Bureau of Economic Research, Cambridge.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge: Polity Press.

Esping-Andersen, Gøsta. 1999. *Social Foundations of Postindustrial Economies*. Oxford: Oxford University Press.

Evers, Adalbert, Jane Lewis, and Birgit Riedel. 2005. "Developing Child-Care Provision in Germany and England: Problems of Governance." *Journal of European Social Policy* 15 (3): 195-209.

Fachinger, Uwe and Ralf K. Himmelreicher. 2008. "Die Bedeutung des Scientific Use Files Vollendete Versichertenleben 2004 (SUF VVL 2004) aus der Perspektive der Ökonomik." *Deutsche Rentenversicherung* 61: 562-582.

Favreault, Melissa M. and C. Eugene Steuerle. 2007. "Social Security Spouse and Survivor Benefits for the Modern Family."29. Center for Retirement Research, Boston.

Federal Republic of Germany. 1949. "Basic Law for the Federal Republic of Germany",

Fehr, Hans and Heinrich Jess. 2007. "Who Benefits from the Reform of Pension Taxation in Germany?*." *Fiscal Studies* 28 (1): 73-101.

Ferber, Marianne A. and Bonnie G. Birnbaum. 1979. "Retrospective Earnings Data: Some Solutions for Old Problems." *The Public Opinion Quarterly* 43 (1): 112-118.

Fitzenberger, Bernd , Aderonke Osikominu, and Robert Völter. 2006. "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample." Center for Doctoral Studies in Economics and Management, Mannheim.

Fitzenberger, Bernd and Stefan Speckesser. 2005. "Employment Effects of the Provision of Specific Skills and Techniques in Germany." ZEW Discussion Paper No 05-77. ZEW, Mannheim.

Fraser, Nancy. 1994. "After the Family Wage: Gender Equity and the Welfare State." *Political Theory* 22 (4): 591-618.

Frericks, Patricia and Robert Maier. 2008. "The Gender Pension Gap: Effects of Norms and Reform Policies", pp. 175-195. In: *Pension Reform in Europe : Politics, Policies and Outcomes*, edited by Camila Arza and Martin Kohli. London: Routledge.

Frericks, Patricia, Robert Maier, and Willibrord De Graaf. 2007. "European Pension Reforms: Individualization, Privatization and Gender Pension Gaps." *Social Politics: International Studies in Gender, State & Society* 14 (2): 212-237.

Frick, Joachim R. and Markus M. Grabka. 2003. "Imputed rent and income inequality decomposition analysis for Great Britain, West Germany and the US." *Review of Income and Wealth* (4): 513-537.

Frick, Joachim R., Markus M. Grabka, and Richard Hauser. 2010a. *Die Verteilung der Vermögen in Deutschland: Empirische Analysen für Personen und Haushalte*, Edited by Hans Böckler Stiftung. Edition Sigma: Berlin.

Frick, Joachim R., Markus M. Grabka, and Jan Marcus. 2007. "Editing and Multiple Imputation of Item-Non-Response in the 2002 Wealth Module of the German Socio-Economic Panel (SOEP)." Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.

Frick, Joachim R., Markus M. Grabka, Timothy M. Smeeding, and Panos Tsakloglou. 2010b. "Distributional effects of imputed rents in five European countries." *Journal of Housing Economics* 19 (3): 167-179.

Frick, Joachim R. and Bruce Headey. 2009. "Living Standards in Retirement: Accepted International Comparisons are Misleading." *Schmollers Jahrbuch* 129 (2): 309-319.

Frick, Joachim R. and Henning Lohmann. 2010. "Biography and Life-History Data in the German Socio-Economic Panel (SOEP, v26, 1984-2009)." Data Documentation 52. DIW, Berlin.

Frommert, Dina and Thorsten Heien. 2006. "Retirement Pension Provision Schemes in Germany 1996 and 2005." *Schmollers Jahrbuch* 126 (2): 329-336.

Gangl, Markus. 2006. "Scar Effects of Unemployment: An Assessment of Institutional Complementarities." *American Sociological Review* 71 (6): 986-1013.

Gangl, Markus and Andrea Ziefle. 2009. "Motherhood, Labor Force Behavior, and Women's Careers: An Empirical Assessment of the Wage Penalty for Motherhood in Britain, Germany and the United States." *Demography* 46 (2): 341-369.

Ginn, Jay. 2003. *Gender, Pensions, and the Lifecourse: How Pensions Need to Adapt to Changing Family Forms*. Bristol: Policy Press.

Ginn, Jay and Deborah Price. 2002. "Do Divorced Women Catch Up in Pension Building?" *Child and Family Law Quarterly* 14 (2): 157-173.

Goebel, Jan and Markus M. Grabka. 2011. "Entwicklung der Altersarmut in Deutschland." *Vierteljahreshefte zur Wirtschaftsforschung* 80 (2): (Forthcoming).

Gornick, Janet and Marcia Meyers. 2003. *Families that Work: Policies for Reconciling Parenthood and Employment*. New York City: Russell Sage.

Grabka, Markus M. 2004. "Alternative Finanzierungsmodelle einer sozialen Krankenversicherung in Deutschland - Methodologische Grundlagen und exemplarische Durchführung einer Mikrosimulationsstudie." Fakultät VIII - Wirtschaft und Management, Technische Universität Berlin, Berlin.

Gruber, Jonathan and David A. Wise. 1999. "Social Security and Retirement Around the World." Chicago: Chicago University Press.

Haak, Carroll. 2006. "Von Künstlern lernen: Mehrfachbeschäftigung, Bildung und Einkommen auf den Arbeitsmärkten von Künstlern." WZB Discussion Paper SP 2006-123Wissenschaftszentrum Berlin für Sozialforschung, Berlin.

Haardt, David. 2006. "Older Couples' Labour Market Reactions to Family Disruptions."27. Institute for Social & Economic Research, Essex.

Hagen, Christine, Ralf K. Himmelreicher, and Hilmar Hoffmann. 2007. "Typologie des Mehrfachrentenbezugs." *RV Aktuell* 54 (3): 52-60.

Haider, Steven. J., Alison Jacknowitz, and Robert F. Schoeni. 2003. "The Economic Status of Elderly Divorced Women."39. Michigang Retirement Research Center, Ann Arbor.

Haines, Steven J. and Brian Greenberg. 2005. "Statistical Use of Social Security Administrative Data." Pp. 1138-1145 in *Joint Statistical Meeting 2005*. Minneapolis: American Statistical Association.

Haisken-DeNew, John P. and Joachim R. Frick. 2005. "Desktop Companion to the German Socio-Economic Panel (SOEP)." Berlin: Deutsches Institut für Wirtschaftsforschung.

Hanel, Barbara and Regina T. Riphahn. 2011. "The Employment of Mothers: Recent Developments and Their Determinants in East and West Germany." IZA Discussion Paper No. 5752 Institute for the Study of Labor (IZA), Bonn.

Hanisch, Jens U. 2005. "Rounded Responses to Income Questions." *Allgemeines Statistisches Archiv* 89 (1): 39-48.

Harrington Meyer, Madonna. 1996. "Making Claims as Workers or Wives: The Distribution of Social Security Benefits." *American Sociological Review* 61 (3): 449-465.

Haurin, Donald R. 1989. "Women's Labor Market Reactions to Family Disruptions." *Review of Economics and Statistics* 71 (1): 54-61.

Hauser, Richard. 2011. "Household Income, Poverty, and Wealth", pp. 1027-1040. In: *Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences*, vol. 2, edited by Rat für Sozial- und Wirtschaftsdaten. Opladen: Budrich UniPress.

Hauser, Richard, Gert G. Wagner, and Klaus F. Zimmermann. 2008. "Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung: Ein Memorandum.", pp. 36-46. In: *Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag.*, edited by Gabriele Rolf, Markus Zwick, and Gert G. Wagner. Baden-Baden: Nomos Verlagsgesellschaft.

Hawala, Sam, Martha Stinson, and John Abowd. 2005. "Disclosure Risk Assessment Through Record Linkage." in *Joint UNECE/Eurostat Work Session on Statistical Confidentiality*. Geneva.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153-161.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64 (4): 605-654.

Heckman, James J. and Thomas E. Macurdy. 1980. "A Life Cycle Model of Female Labour Supply." *The Review of Economic Studies* 47 (1): 47-74.

Heese, Claudia. 2004. "Aspekte des Datenschutzes im Forschungsdatenzentrum der Rentenversicherung.", In: *Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung im Aufbau. Bericht vom ersten Workshop des FDZ am 28. und 29. Juni 2004 in Würzburg.*, vol. 55 (Sonderausgabe). Bad Homburg: WDV Verlag.

Himmelreicher, Ralf K. 2005. "Analysepotenzial des Scientific Use Files Versichertenrentenzugang", pp. 38-92. In: *Forschungsrelevante Daten der Rentenversicherung: Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg*. Berlin: Deutsche Rentenversicherung Bund.

Himmelreicher, Ralf K. and Dirk Mai. 2006. "Retrospektive Querschnitte: Das Analysepotential des Scientific Use Files Versichertenrentenzugang 2004 unter besonderer Berücksichtigung der Berechnung von Entgeltpunkten." *DRV Schriften* 55: 23-51.

Himmelreicher, Ralf K. and Holger Viebrok. 2004. "Die "Riester-Rente" und einige Folgen für die Alterseinkünfte." Zentrum für Sozialpolitik, Bremen.

Hoffmann, Hilmar. 2007. "Wege in den Ruhestand." *Deutsche Rentenversicherung* 62 (4-5): 298-320.

Holst, Elke. 2000. *Die Stille Reserve am Arbeitsmarkt: Größe – Zusammensetzung – Verhalten*. Berlin: Edition Sigma.

Holst, Elke. 2010. "Wochenarbeitzeiten: Wunsch und Wirklichkeit nach der deutschen Wiedervereinigung bis 2008", pp. 313-327. In: *Leben in Ost- und Westdeutschland. Eine sozialwissenschaftliche Bilanz der deutschen Einheit 1990-2010*, edited by Peter Krause and Ilona Ostner. Frankfurt am Main: Campus.

Hujer, Reinhard, Marco Caliendo, and Stephan L. Thomsen. 2004. "New Evidence on the Effects of Job Creation Schemes in Germany--a Matching Approach with Threefold Heterogeneity." *Research in Economics* 58 (4): 257-302.

Immergut, Ellen M., Karen M. Anderson, and Isabelle Schulze. 2007. "The Handbook of West European Pension Politics." Oxford: Oxford University Press.

Jarvis, Sarah and Stephen P. Jenkins. 1999. "Marital Splits and Income Changes: Evidence from the British Household Panel Survey." *Population Studies* 53 (2): 237-254.

Jenkins, Stephen P. 2008. "Marital Splits and Income Changes over the Longer Term."25. Institute for Social & Economic Research, Essex.

Jenkins, Stephen P., Lorenzo Cappellari, Peter Lynn, Annette Jäckle, and Emanuela Sala. 2006. "Patterns of Consent: Evidence from a General Household Survey." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169 (4): 701-722.

Johnson, William R. and Jonathan Skinner. 1986. "Labor Supply and Marital Separation." *The American Economic Review* 76 (3): 455-469.

Joshi, Heather and Hugh Davies. 1991. "Pension Splitting and Divorce." *Fiscal Studies* 12 (4): 69-91.

Jürges, Hendrik. 2010. "Provision of Old-Age: National and International Survey Data to Support Research and Policy on Aging", pp. 1094-1105. In: *Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences*, vol. 2, edited by Rat für Sozial- und Wirtschaftsdaten. Opladen: Budrich UniPress Ltd.

Kadane, Joseph B. [1978] 2001. "Some Statistical Problems in Merging Data Files." *Journal of Official Statistics* 17 (3): 423-433.

Kalmijn, Matthijs. 2005. "The Effects of Divorce on Men's Employment and Social Security Histories." *European Journal of Population* 21: 347-366.

Kapteyn, Arie and Jelmer Y. Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics* 25 (3): 513-551.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. 2001. *Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik.* Baden-Baden: Nomos Verlagsgesellschaft.

Köppen, Katja. 2010. "Marriage and Cohabitation in western Germany and France." Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität Rostock, Rostock.

Kortmann, Klaus and Verena Halbherr. 2008. "Alterssicherung in Deutschland 2007 (ASID '07): Zusammenfassung wichtiger Ergebnisse."147. TNS Infratest, München.

Kreuter, Frauke, Gerrit Müller, and Mark Trappmann. 2010. "Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data." *Public Opinion Quarterly* 74 (5): 880-906.

Kreyenfeld, Michaela. 2009. "Uncertainties in Female Employment Careers and the Postponement of Parenthood in Germany." *European Sociological Review* Advance Access Published Online May 4, 2009.

Kreyenfeld, Michaela and Esther Geisler. 2006. "Müttererwerbstätigkeit in Ost- und Westdeutschland." *Zeitschrift für Familienforschung* 18 (3): 333-360.

Kreyenfeld, Michaela and Tatjana Mika. 2008. "Erwerbstätigkeit und Fertilität: Analysen mit der Versicherungskontenstichprobe der deutschen Rentenversicherung " *DRV Schriften* 79 (Die Versicherungskontenstichprobe als Scientific Use File. Workshop des Forschungsdatenzentrums der Rentenversicherung): 71-95.

Kruppe, Thomas. 2009. "Empirical Consequences of Definitions. The Case of Unemployment in German Register Data." *Definitionen und ihre empirische Folgen -- „Arbeitslosigkeit" in deutschen Register Daten.* 34 (3): 138-148.

Kruppe, Thomas, Eva Müller, Laura Wichert, and Ralf A. Wilke. 2008. "On the Definition of Unemployment and its Implementation in Register Data - The Case of Germany." *Schmollers Jahrbuch* 128 (3): 461-488.

Kruse, Edgar. 2007. "Empirische Ergebnisse zur Erwerbsbeteiligung älterer Versicherter anhand der Statistiken der gesetzlichen Rentenversicherung." *Deutsche Rentenversicherung* (11-12): 716-750.

Leacock, Cindi. 2006. "Getting Started with the Health and Retirement Study." Ann Arbor: University of Michigan.

Levenshtein, Vladimir. 1966. "Binary Codes Capable of Correcting Deletions and Insertions, and Reversals." *Soviet Physics Doklady* 10 (8): 707-710.

Lillard, Dean R. and Linda Waite. 2000. "Marriage, Divorce and the Work and Earning Careers of Spouses."24. University of Michigan Retirement Research Center, Ann Arbor.

Little, Roderick J. A. 1988. "Missing-Data Adjustments in Large Surveys." *Journal of Business & Economic Statistics* 6 (3): 287-296.

Loose, Brigitte and Dina Frommert. 2009. "Verstetigung in der Alterssicherung durch Integration ungesicherter Selbständigkeit in die gesetzliche Rentenversicherung." *Deutsche Rentenversicherung* 64 (6): 501-517.

Mahalanobis, Prasanta C. 1936. "On the Generalized Distance in Statistics." *Proceeding of the National Institute of Sciences of India* 2 (1): 49-55.

Manning, Alan and Barbara Petrongolo. 2008. "The Part-Time Pay Penalty for Women in Britain." *The Economic Journal* 118 (2): F28-F51.

Mayer, Karl Ulrich and Michael Wagner. 1996. "Lebenslagen und soziale Ungleichheit im hohen Alter", pp. 251-275. In: *Die Berliner Altersstudie*, edited by Karl Ulrich Mayer and Paul B Baltes. Berlin: Akademie Verlag.

McManus, Patricia and Thomas DiPrete. 2001. "Losers and Winners: The Financial Consequences of Separation and Divorce for Men." *American Sociological Review* 66 (2): 246-268.

Mincer, Jacob and Haim Ofek. 1982. "Interrupted Work Careers: Depreciation and Restoration of Human Capital." *Journal of Human Resources* 17 (1): 3-24.

Mincer, Jacob and Polachek Solomon. 1978. "An Exchange: The Theory of Human Capital and the Earnings of Women: Women's Earnings Reexamined." *The Journal of Human Resources* 13 (1): 118-134.

Misra, Joya, Stephanie Moller, and Michelle J. Budig. 2007. "Work Family Policies and Poverty for Partnered and Single Women in Europe and North America." *Gender & Society* 21 (6): 804-827.

Modigliani, Franco. 1988. "The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth." *The Journal of Economic Perspectives* 2 (2): 15-40.

Morgan, Stephen L. and David J. Harding. 2006. "Matching Estimators of Causal Effects - Prospects and Pitfalls in Theory and Practice." *Sociological Methods & Research* 35 (1): 3-60.

Moriarity, Chris and Fritz Scheuren. 2001. "Statistical Matching: A Paradigm for Assessing the Uncertainty in the Procedure." *Journal of Official Statistics* 17 (3): 407-422.

Munnell, Alicia H. 2004. "Why Are So Many Older Women Poor?" Just the Facts on Retirement Issues.8. Center for Retirement Research at Boston College, Boston.

Needleman, Saul B. and Christian D. Wunsch. 1970. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins." *Journal of Molecular Biology* 48 (3): 443-453.

OECD. 2010. "OECD Family Database 2010." Paris: OECD.

OECD. 2011. "Society at a Glance - OECD Social Indicators." OECD, Paris.

Orlowski, Robert and Regina T. Riphahn. 2009. "The East German wage structure after transition." *Economics of Transition* 17 (4): 629-659.

Pampel, Fred C. 2000. *Logistic Regression Analysis: A Primer*. Thousand Oaks: Sage Publications, Inc.

Polachek, Solomon. 1975. "Discontinuous Labor Force Participation and its Effect on Women's Market Earnings", pp. 90-124. In: *Sex, Discrimination, and the Division of Labor*, edited by Cynthia B. Lloyd. New York: Columbia University Press.

Pollock, Gary. 2007. "Holistic Trajectories: a Study of Combined Employment, Housing and Family Careers by Using Multiple-Sequence Analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (1): 167-183.

Poortman, Anne-Rigt. 2000. "Sex Differences in the Economic Consequences of Separation: A Panel Study of the Netherlands." *European Sociological Review* 16 (4): 367-383.

Rabe, Birgitta. 2007. "Occupational Pensions, Wages, and Job Mobility in Germany." *Scottish Journal of Political Economy* 54 (4): 531-552.

Rasner, Anika. 2006. "Mind the Gap! Einbeziehung internationaler Benchmarks bei der Beurteilung der geschlechtsspezifischen Rentenlücke in Deutschland." *Deutsche Rentenversicherung* (11-12): 737-754.

Rasner, Anika. 2007. "Vorarbeiten für ein Statistisches Matching von Befragungs- und Registerdaten - Das Sozio-oekonomische Panel und der Scientific Use File Vollendete Versichertenleben 2004." Pp. 201-206 in *Vierter Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV)*, vol. 55/2007. Wissenschaftszentrum Berlin für Sozialforschung: Deutsche Rentenversicherung: Berlin.

Rasner, Anika, Joachim R. Frick, and Markus M. Grabka. 2011. "Extending the Empirical Basis for Wealth Inequality Research Using Statistical Matching of Administrative and Survey Data." SOEPpapers on Multidisciplinary Panel Data Research.44. DIW Berlin, Berlin.

Rasner, Anika, Ralf K. Himmelreicher, Markus M. Grabka, and Joachim R. Frick. 2007. "Best of Both Worlds – Preparatory Steps in Matching Survey Data with Administrative Pension Records. The Case of the German Socio-Economic Panel and the Scientific Use File Completed Insurance Biographies 2004.":210. Deutsches Institut für Wirtschaftsforschung, Berlin.

Rässler, Susanne. 2002. *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York: Springer.

Rässler, Susanne, Donald B. Rubin, and Nathaniel Schenker. 2008. "Incomplete Data: Diagnosis, Imputation, and Estimation", pp. 370-386. In: *International Handbook of Survey Methodology*, edited by Edith D. de Leeuw, Joop J. Hox, and Don A. Dillman. Hillsdale: Lawrence Erlbaum Associates.

Rat für Sozial- und Wirtschaftsdaten. 2010. "Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences." vol. 1 & 2. Opladen & Farmington Hills: Budrich UniPress.

Rehfeld, Uwe, Thomas Bütefisch, and Hilmar Hoffmann. 2007. "Gesundheitsbedingte Leistungen der gesetzlichen Rentenversicherung für Frauen und Männer; Indikatoren für die Morbidität", pp. 145-158. In: *Fehlzeiten-Report 2007: Chronische Krankheiten*, edited by Bernhard Badura, Helmut Schellschmidt, and Christian Vetter. Berlin: Springer.

Richter, Maria and Ralf K. Himmelreicher. 2008. "Die Versicherungskontenstichprobe als Datengrundlage für Analysen von Versicherungsbiographien unterschiedlicher Altersjahrgänge." *DRV Schriften* 79: 34-61.

Rodgers, Willard L. 1984. "An Evaluation of Statistical Matching." *Journal of Business & Economic Statistics* 2 (1): 91-102.

Rolf, Gabriele, Gert G. Wagner, and Markus Zwick. 2008. "Fortschritte der informationellen Infrastruktur in Deutschland: Ein Überblick über die Jahre 1998 bis 2008.", pp. 15-35. In: *Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag.*, edited by Gabriele Rolf, Markus Zwick, and Gert G. Wagner. Baden-Baden: Nomos Verlagsgesellschaft.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41-55.

Rubin, Donald B. 1986. "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations." *Journal of Business & Economic Statistics* 4 (1): 87-94.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Sackmann, Reinhold and Matthias Wingens. 2003. "From Transitions to Trajectories: Sequence Types", pp. 93-116. In: *Social Dynamics of the Life Course*, edited by Walter R. Heinz and Victor W. Marshall. Hawthorne: Aldine de Gruyter.

Sainsbury, Diane. 1999. *Gender, Equality, and the Welfare State*. New York: Cambridge.

Schmähl, Winfried. 2011. "Politikberatung und Alterssicherung: Rentenniveau, Altersarmut und das Rentenversicherungssystem." *Vierteljahreshefte zur Wirtschaftsforschung* 80 (1): 159-174.

Schröder, Mathis. 2011. "Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe." in *SHARELIFE Methodology*. Mannheim: MEA.

Schupp, Jürgen. 2009. "25 Jahre Sozio-oekonomisches Panel -- Ein Infrastrukturprojekt der empirischen Sozial- und Wirtschaftsforschung in Deutschland. (German)." *Twenty-Five Years of the German Socio-Economic Panel -- An Infrastructure Project for Empirical Social and Economic Research in Germany. (English)* 38 (5): 350-357.

Schwarze, Johannes, Gert G. Wagner, and Christoph Wunder. 2004. "Alterssicherung: Gesunkene Zufriedenheit und Skepsis gegenüber privater Vorsorge." DIW Wochenbericht 22/2004 DIW, Berlin.

Seitz, Helmut. 2003. "Die Leistungen der neuen Länder im Rahmen des AAÜG - Bestimmungsgründe und Belastungsdynamik - Gutachten im Auftrag der Länder Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen."

Shuey, Kim M. and Angela M. O'Rand. 2006. "Changing Demographics and New Pension Risks." *Research on Aging* 28 (3): 317-340.

Sierminska, Eva. M., Joachim R. Frick, and Markus M. Grabka. 2010. "Examining the Gender Wealth Gap." *Oxford Economic Papers-New Series* 62 (4): 669-690.

Simonson, Julia, Laura Romeu Gordo, and Nadiya Titova. 2011. "Changing Employment Patterns of Women in Germany: How do Baby Boomers Differ from Older Cohorts? A Comparison Using Sequence Analysis." *Advances in Life Course Research* 16 (2): 65-82.

Singh, R., R. J. Petroni, and T. M. Allen. 1994. "Oversampling in Panel Surveys." US Census Bureau, Washington D.C.

Singleton, Royce and Bruce C. Straits. 2010. *Approaches to Social Research*. New York: Oxford University Press.

Smeeding, Timothy M. and Susanna Sandstrom. 2005. "Poverty and Income Maintenance in Old Age: A Cross National View of Low Income Older Women."17. Center for Retirement Research, Boston.

Smock, Pamela J. 1993. "The Economic Costs of Marital Disruption for Young Women over the Past Two Decades." *Demography* 30 (3): 353-371.

Smock, Pamela J. 1994. "Gender and Short-Run Economic Consequences of Marital Disruption." *Social Forces* 73 (1): 243-262.

Smock, Pamela J., Wendy D. Manning, and Sanjiv Gupta. 1999. "The Effect of Marriage and Divorce on Women's Economic Well-Being." *American Sociological Review* 64 (6): 794-812.

Social Security Administration. 2006. "Annual Statistical Supplement to the Social Security Bulletin, 2005." edited by Office of Policy. Baltimore.

Social Security Administration. 2007. "Totalization Agreement with Germany." Retrieved July 10, 2011 (http://www.ssa.gov/international/Agreement_Pamphlets/germany.html).

Social Security Administration. 2011. "National Average Wage Index." Retrieved June 17, 2011, 2011

Spieß, C. Katharina and Katharina Wrohlich. 2008. "The Parental Leave Benefit Reform in Germany: Costs and Labour Market Outcomes of Moving towards the Nordic Model." *Population Research Policy Review* 27 (5): 575-591.

Statistische Ämter des Bundes und der Länder. 2011. "The 2011 Census in Germany." Retrieved July 5, 2011 (https://www.zensus2011.de/english.html).

Statistisches Bundesamt. 2008. "Bevölkerungsfortschreibung." Statistisches Bundesamt, Wiesbaden.

Stegmann, Michael. 2006a. "Aufbereitung der Sondererhebung "Vollendete Versichertenleben 2004" als Scientific Use File für das FDZ-RV." *Deutsche Rentenversicherung* 61 (9-10): 537-553.

Stegmann, Michael. 2006b. "Die Bedeutung des Scientific Use Files FDZ-Biografiedaten- VVL2004 (SUFVVL2004) für die Arbeitsmarktforschung." *Deutsche Rentenversicherung* 61 (9-10): 554-561.

Stegmann, Michael. 2008. "Aufbereitung der Sondererhebung "Versicherungskontenstichprobe (VSKT)" als Scientific Use File für das FDZ-RV." *DRV-Schriften* Band 79.

Stegmann, Michael. 2009. "Das Einkommen aus der gesetzlichen Rentenversicherung von Senioren im Alter über 65 Jahren: Ergebnisse der Rentenzahlbestandsstatistik 2008." *Deutsche Rentenversicherung* (3): 254-269.

Steiner, Viktor and Kersten Wagner. 1998. "Has Earnings Inequality in Germany Changed in the 1980s?" *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 118: 29-59.

Steiner, Viktor and Katharina Wrohlich. 2004. "Household Taxation, Income Splitting and Labor Supply Incentives - A Microsimulation Study for Germany." *CESifo Economic Studies* 50 (3): 541-568.

Steuerle, C. Eugene, Adam Carasso, and Lee Cohen. 2004. "How Progressive is Social Security and Why?" Straight Talk on Social Security and Retirement Policy.2. Urban Institute, Washington

Steuerle, C. Eugene and Christopher Spiero. 1999. "Does Social Security Treat Spouses Fairly?" Retirement Policy Brief Nr. 12. Urban Institute, Washington, D.C. .

Stier, Haya, Noah Lewin-Epstein, and Michael Braun. 2001. "Welfare Regimes, Family-Supportive Policies, and Women's Employment along the Life-Course." *The American Journal of Sociology* 106 (6): 1731-1760.

Tamborini, Christopher R. 2007. "The Never-Married in Old Age: Projections and Concerns for the Near Future." *Social Security Bulletin* 67 (2): 25-40.

Tamborini, Christopher R., Howard M. Iams, and Kevin Whitman. 2009. "Marital History, Race, and Social Security Spouse and Widow Benefit Eligibility in the United States." *Research on Aging* 31 (5): 577-605.

The Urban Institute. 2009. "Reform Options for Social Security: Spouse and Survivor Benefits." Retrieved April 15, 2009, 2009 (http://www.urban.org/retirement_policy/ssspousesurvivor.cfm).

TNS Infratest. 2009. "Alterssicherung in Deutschland 2007 (ASID 2007) – Zusammenfassung wichtiger Untersuchungsergebnisse." Bundesministerium für Arbeit und Soziales, Bonn.

Uunk, Wilfred. 2004. "The Economic Consequences of Divorce for Women in the European Union: The Impact of Welfare State Arrangements." *European Journal of Population* 20: 251-285.

van der Putten, Peter, Joost N. Kok, and Amar Gupta. 2002. "Data Fusion Through Statistical Matching." MIT Sloan School of Management, Cambridge.

Vartanian, Thomas P. and Justine M. McNamara. 2002. "Older Women in Poverty: The Impact of Midlife Factors." *Journal of Marriage and Family* 64 (2): 532-548.

Vaupel, James W. and Elke Loichinger. 2006. "Redistributing Work in Aging Europe." *Science* 312 (5782): 1911-1913.

Verband Deutscher Rentenversicherungsträger. 2005. "Rentenzugang 2004."350. Verband Deutscher Rentenversicherungsträger, Frankfurt am Main.

Wagner, Gert G. 1980. "Der Versicherungsstatus Selbständiger in der Gesetzlichen Rentenversicherung." Frankfurt am Main: Universität Frankfurt.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2006. "Enhancing the Power of Household Panel Studies - The Case of the German Socio-Economic Panel Study (SOEP)." Data Documentation 13. DIW Berlin.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements." *Schmollers Jahrbuch* 127 (1): 139-169.

Wagner, Gert, Jan Göbel, Peter Krause, Rainer Pischner, and Ingo Sieber. 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, Michael and Bernd Weiß. 2006. "On the Variation of Divorce Risks in Europe: Findings from a Meta-Analysis of European Longitudinal Studies." *European Sociological Review* 22 (5): 483-500.

Wilmoth, Janet and Gregor Koso. 2002. "Does Marital History Matter? Marital Status and Wealth Outcomes among Preretirement Adults." *Journal of Marriage and Family* 64 (1): 254-268.

Winkler, William E. 2006. "Overview of Record Linkage and Current Research Directions." U.S. Census Bureau, Statistical Research Division, Washington D.C.

Wirth, Heike and Walter Müller. 2006. "Mikrodaten der Amtlichen Statistik - Ihr Potenzial in der Empirischen Sozialforschung", pp. 93-127. In: *Methoden der Sozialforschung*, edited by Andreas Dieckmann. Wiesbaden: VS Verlag für Sozialwissenschaften.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge MA: The MIT Press.

Appendix

Table A1 Population Counts vs. Active Pension Accounts by Citizenship Status & Birth Cohort

	Population Counts	Pension Accounts	Difference	Population Counts	Pension Accounts	Difference
Cohort	Total	Total	Total	Total	Total	Total
1940-1944	4,737,826	4,597,076	140,750	282,146	755,811	-473,665
1945-1949	4,003,391	3,892,342	111,049	393,007	913,990	-520,983
1950-1954	5,080,904	5,003,533	77,371	417,177	922,950	-505,773
1955-1959	5,646,127	5,657,553	-11,426	466,145	918,765	-452,620
1960-1964	6,493,645	6,542,689	-49,044	580,230	991,741	-411,511
1965-1969	6,192,666	6,255,226	-62,560	711,475	1,097,471	-385,996
1970-1974	4,461,336	4,529,911	-68,575	814,199	1,107,385	-293,186
1975-1979	3,947,114	4,086,032	-138,918	811,234	919,891	-108,657
1980-1984	4,273,585	4,380,788	-107,203	705,915	617,268	88,647
1985-1989	4,363,729	3,800,731	562,998	505,968	348,988	156,980
1990-1992	2,429,682	728,446	1,701,236	258,705	56,461	202,244
Total	51,630,005	49,474,327	2,155,678	5,946,201	8,650,721	-2,704,520

Note: Population counts come from the German Federal Statistical Office (Table B15/A1). Source: Federal Statistical Office 2008 & Federal statutory pension insurance 2007; Author's calculations

Table A2 Distribution of Gender in Population Statistics & Active Pension Accounts by Citizenship

	Men	Women
	Total Population	
Population Counts	29,066,049	28,163,715
Pension Accounts	29,961,333	28,510,157
Difference	-895,284	346,442
	Only Germans	
Population Counts	26,014,788	25,615,217
Pension Accounts	24,947,902	24,526,337
Difference	1,066,886	1,088,880
	Non-German Born	
Population Counts	3,051,261	2,894,940
Pension Accounts	5,013,431	3,637,290
Difference	-1,962,170	-742,350

Note: Population counts come from the German Federal Statistical Office (Table B15/A1). Source: Federal Statistical Office 2008 & Federal statutory pension insurance 2007; Author's calculations

Table A3 Summary Statistics by Status of Account Validation in SAPA data

Variables	Categories	SAPA Total	SAPA Validated with Help	SAPA Validated w/o Help	Non-Validated Accounts
Number of Accounts	Absolute	568,865	214,259	121,810	232,517
	Weighted	58,125,048	26,521,851	13,986,508	17,616,689
	Share	100.0	45.6	24.1	30.3
Gender	Male	51.6	45.8	59.4	54.0
	Female	48.5	54.2	40.6	46.0
Region	Unknown	3.7	3.5	1.6	5.6
	West Germany	77.6	74.5	80.3	80.0
	East Germany	18.8	22.0	18.0	14.5
Origin	German	85.1	91.0	87.7	74.1
	Migrant	14.9	9.0	12.3	25.9
Age	Min	15	16	16	15
	Max	67	67	67	67
	Mean	42.9	49.9	44.4	31.2
Age at Account Validation	Mean	n/a	47	41	n/a
Cohort	1940-1944	9.2	18.1	1.4	2.1
	1945-1949	8.3	12.6	6.7	3.1
	1950-1954	10.2	13.0	12.7	4.0
	1955-1959	11.3	13.4	15.7	4.8
	1960-1964	13	14.4	19.2	5.8
	1965-1969	12.7	12.8	20.1	6.6
	1970-1974	9.7	9.2	15.7	5.6
	1975-1979	8.6	4.9	7.2	15.4
	1980-1984	8.6	1.6	1.2	25.0
	1985-1989	7.1	0.1	0.1	23.3
	1990-1992	1.4	0.0	0.0	4.4
Number of Children (only Women)	No Children	45.4	21.4	49.8	84.9
	One Child	20.6	25.2	24.0	10.0
	Two Children	23.4	36.9	17.7	3.5
	Three Children	7.7	12.0	6.2	1.2
	Four Children	2.1	3.2	1.7	0.3
	Five + Children	0.9	1.5	0.7	0.1
	Mean	1.0	1.6	0.9	0.2
Divorced	Share	8.1	11.5	9.7	1.4
Caregiving	Share	2.4	3.8	2.0	0.5
Retired	Share	10.4	22.7	0.1	0.2

Source: Sample of Active Pension Accounts (SAPA) 2007; Author's calculations

Table A4 Validity and Reliability of Selected Variables in the Sample of Active Pension Accounts

Variable	Description	Relevant for Administration of Statutory Pension Scheme	Cross Sectional/ Longitudinal	Measurement Consistent over Time or Time Dependent
GEH	Gender	Yes, eligibility criteria for certain benefit types depend on gender	Cross-Section	Yes
GBJA/GBMO	Year and month of birth	Yes, retirement age and level of pension benefits depend on year and month of birth	Cross-Section	Yes
KTSD	Status of account validation	No	Cross-Section	Yes
PSGR	Retirement status	Yes	Cross-Section	Yes
TLRT	Partial pension benefit	Yes, determines the amount of pension benefits and earning limits	Cross-Section	Yes
ZTPTRTBEJJ/ZTPTRTBEMM	Year and month of retirement (for benefit currently received	Yes, determine level of benefits	Cross-Section	Yes
GBKIJ/GBKIM	Year and month of birth for each child	Yes determine the number of child care credits	Cross-Section	Yes. The number of child care credits per child changes over time and depends on the year/month of birth.

Source: Codebook Sample of Active Pension Accounts (2007a); Author's Illustration

Table A4 Validity and Reliability of Selected Variables in the Sample of Active Pension Accounts (Continued)

Variable	Description	Relevant for Administration of Statutory Pension Scheme	Cross Sectional/ Longitudinal	Measurement Consistent over Time or Time Dependent
VSKT	Indicates whether a person was ever insured in the miners pension scheme	Yes, because benefit calculation differs for miners	Cross-Section	Consistent
VSAT	Type of insurance (last pension-relevant status)	No	Cross-Section	Not consistent, because last pension-relevant status is not comparable across individuals (for one person it refers to last month for the other variable refers to pension-relevant status five years ago)
TTSC2	Occupational Status	No	Based on information reported by employer for 2007	Inconsistent and unreliable
TTSC3	Educational Attainment	No	Based on information reported by employer for 2007	Inconsistent and unreliable
PFLEGE[1]	Indicates whether person is caregiver	Yes, individuals receive care credits for caregiving	Longitudinal	Pension-relevant since 1995; consistent in measurement

Note: [1] PFLEGE is one of the variables in the variable part of the data. Source: Codebook Sample of Active Pension Accounts (2007a); Author's Illustration

Table A5 Contributions for Periods of Employment and Non-Employment and Legal Changes over time

Pension-Relevant Episode	Time Period	What is the Basis for the Calculation of Earning Points?
Employment subject to social insurance contribution	General Rule	Monthly gross wage (up to some maximum amount). Set in relation to the average monthly gross wage. Social insurance contributions paid by the employer are not counted in.
Sickness	General Rule	During the first six weeks continuation of payments for sick workers through employer. This period is equivalent to employment subject to social insurance contributions and appears as such in the data. After six weeks of sickness, the person receives a sickness allowance. The health insurance provider pays the sickness allowance to the beneficiary, but also contributions to the statutory pension insurance. Typically, sickness allowances are paid for a maximum period of 48 weeks, the duration depends on the statutes of the respective insurance provider.
	October 1974 – December 1983	During the first 12 months of sickness allowance receipt, no contributions are paid. Starting with the 13th month, contributions are paid to the statutory pension insurance. Basis for the calculation of contributions are 100 percent of the last gross wage prior to the receipt of sickness allowance (§2 Abs. 1 Nr. 10a AVG or §247 Abs. 2 SGBVI).
	January 1984 – December 1991	No obligations to pay contributions upon the receipt of sickness allowances. Period of receipt is defined as a creditable period (at that time period of non-contribution). These creditable periods lead to an increase of pension payments as the person retires (§112 Abs. 1 SGBVI or §247 Abs. 1 SGB VI).
	Since 1992	During the receipt of sickness allowances, contributions have to be paid if the person was employed and obliged to pay social insurance contributions. Basis for the assessment of contributions are 80 percent of the last gross wage. If a person was not employed during the 12 months before the receipt of the sickness allowance, no contributions are paid. If a person was unemployed and received unemployment benefits, then the person can pay voluntary contributions (§276 Abs. 1 SGB VI or §3 Satz 1 Nr. 2 SGB VI).

Source: Author's illustration

Table A5 Contributions for Periods of Employment and Non-Employment and Legal Changes over time (Continued)

Pension-Relevant Episode	Time Period	What is the Basis for the Calculation of Earning Points?
Unemployment Compensation (*Arbeitslosengeld*)	General Rule	Unemployment compensation amounts to 67 percent of the last gross wage for individuals with children and 60 percent for persons without any children. The amount of unemployment compensation is not relevant for the determination of contributions, but rather the person's last gross wage.
	July 1978 – July 1983	Unemployment compensation is subject to social insurance contributions. The basis for the determination of contributions is 100 percent of the person's last gross wage.
	August 1983 – December 1991	No obligation to pay contributions upon the receipt of unemployment compensation Period of receipt is defined as a creditable period (at that time *period of non-contribution*). These creditable periods lead to an increase of pension payments as the person retires (§112 Abs. 1 SGB VI or §247 Abs. 1 SGB VI).
	Since 1992	Unemployment assistance is subject to social insurance contributions. The basis for the determination of contributions is 80 percent of the person's last gross wage (since 1992: §3 Satz 1 Nr.3 SGB VI; in addition §166, Abs. 1 Satz 1 Nr. 2 SGB VI).
Unemployment Assistance (*Arbeitslosenhilfe*)	General Rule	The level of unemployment assistance is 52 percent of the last gross wage.
	January 1984 – December 1991	No obligation to pay contributions upon the receipt of unemployment assistance. Period of receipt is defined as creditable period (at that time *period of non-contribution*). These creditable periods lead to an increase of pension payments (§112 Abs. 1 SGB VI or §247 Abs. 1 SGB VI).
	January 1992 – December 1996	Unemployment assistance is subject to social insurance contributions. Basis for the determination of contributions is 80 percent of the last gross wage that was relevant for the payment of unemployment assistance (1992-1994: §276 Abs. 1 SGB VI; 1995-1996: §166 Abs. 1 Satz 1 Nr. 2 SGB VI).

Source: Author's illustration

Table A5 Contributions for Periods of Employment and Non-Employment and Legal Changes over time (Continued)

Pension-Relevant Episode	Time Period	What is the Basis for the Calculation of Earning Points?
Unemployment Assistance (cont'd) (*Arbeitslosenhilfe*)	January 1997 – December 1999	Unemployment assistance is subject to social insurance contributions. Basis for the determination of contributions is 80 percent of last gross wage that was relevant for the payment of unemployment assistance. If unemployment assistance levels are cut because other income is credited against the means-tested benefit, then contributions are cut accordingly (§166 Abs. 1 Satz 1 Nr. 2a SGB VI).
	Since 2000	The actual amount of the unemployment assistance is the reference value for the calculation of benefits §166 Abs. 1 Satz 1 Nr. 2a (incl. footnotes 8 and 11).
Military and Civil Service	General Rule	Contributions are not paid from the monthly pay for soldiers and civil servants. The basis for the determination of contributions is a fictitious wage from which contributions are paid for by the government. Over time, contributions are either determined using the average gross wage.
	1965 - 1981	100 percent of average gross wage
	1982 - 1991	75 percent of average gross wage (§256 Abs. 3 SGB VI)
	1992 - 1999	70 percent of pension insurance reference value *(jährliche Bezugsgröße) (§166 Abs. 1 Satz 1 Nr. 1 SGB VI Footnote 8)*
	Since 2000	60 percent of pension insurance reference value (§166 Abs. 1 Satz 1 Nr. 1 SGB VI)
Vocational Training	General Rule	Contributions depend on the salary of the apprentice.
	Until 1992	Contribution rate is the same as for regular employees. There were no special regulations in place because of below average salaries of apprentices.
	Since 1992	Contribution rate is the same as for regular employees. Below average salaries of apprentices are upgraded (below 630 DM or 325 €). The contribution rate corresponds to one percent of the annual pension insurance reference value.

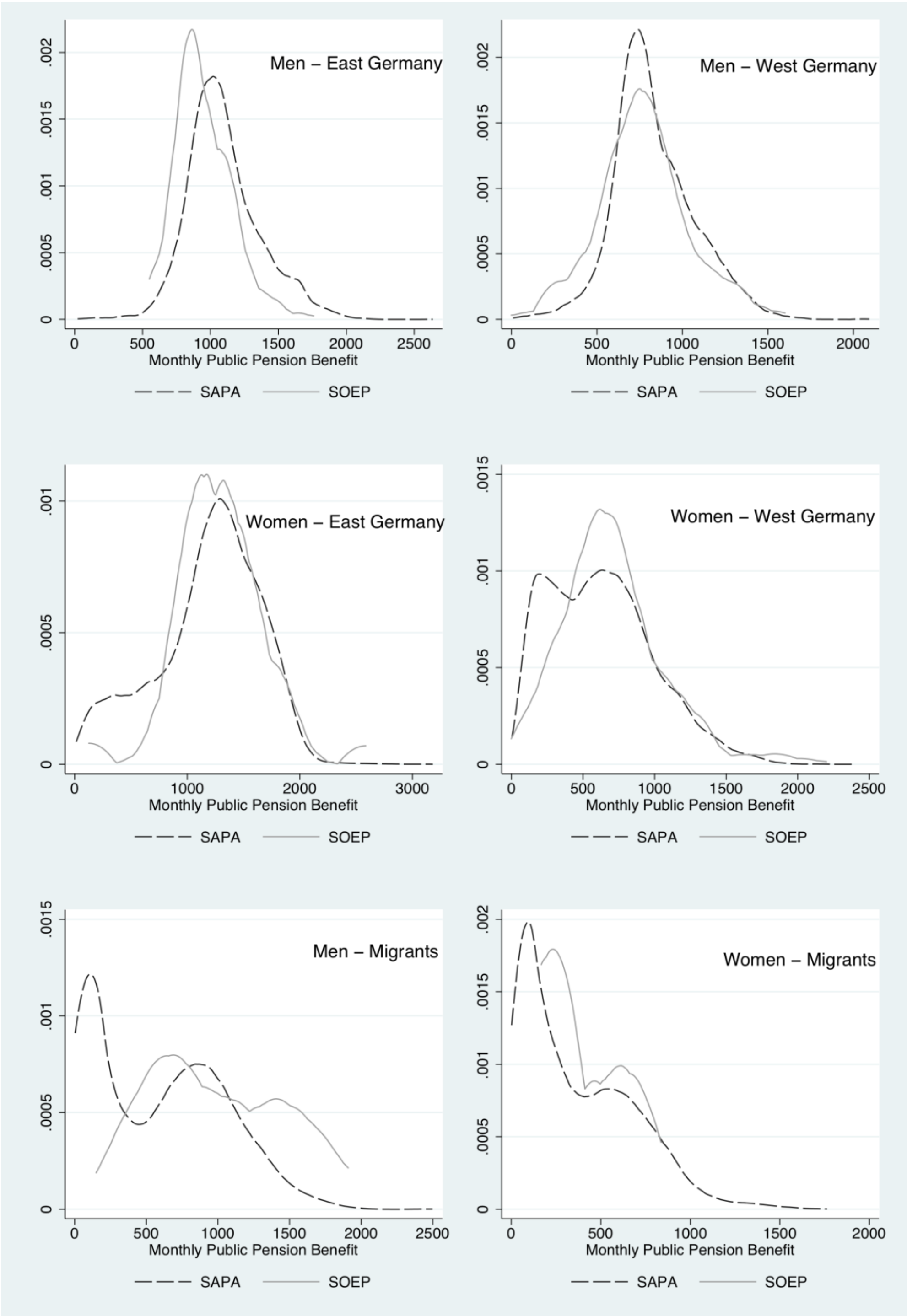
Source: Author's illustration

Table A5 Contributions for Periods of Employment and Non-Employment and Legal Changes over time (Continued)

Pension-Relevant Episode	Time Period	What is the Basis for the Calculation of Earning Points?
Child Care	General Rule	Women who give birth to a child receive child care credits that compensate for the presumably lower labor supply or exit from the labor market after childbirth. The government pays child care credits come from tax revenues.
	Until 1992	A woman receives a one year child care credit per child that corresponds to the average wage in that year (§249 Abs. 1 SGB VI und §70 Abs. 2 SGB VI).
	Since 1992	A woman receives three years of child care credits worth the average wage in that year (§56 Abs. 1 SGB VI und §70 Abs. 2 SGB VI).

Source: Author's illustration

Figure A1 Comparison of Monthly Public Pension Benefit in SOEP and SAPA across Demographic Groups



Source: SAPA 2007 and SOEP 2007; Author's calculations

Notes for the following Tables:

Tables A6 to *A12*: The tables provide results on the average time spent in different activities for all SOEP respondents. The tables compare men with women, Germans with persons with a migration history, and West Germans with East Germans across different populations of retirees. The calculation of the average time is based on all individuals in the respective demographic group, independent of whether they have actually spent time in the respective activity or not. Hence, the denominator is always the total number of individuals in the group.

Tables A13 to *A17*: The tables provide results on the average time spent in different activities for SOEP respondents. The tables compare men with women, Germans with persons with a migration history, and West Germans with East Germans across different populations of retirees. The calculation of the average time is based on all individuals in the respective demographic group, who have spent time in the respective activity. Hence, the denominator includes all individuals who have information > 0 for the respective category.

Retirement status: Information is based on Question 103 in the SOEP Person Questionnaire (*wave v*). For persons reporting an own pension income from the statutory pension insurance in 2005, the variable *retirement status* was coded with 1. If the person was below age 60 in 2005, it is assumed that the person received an invalidity or disability pension (*Erwerbs- oder Berufsunfähigkeitsrente*) or an orphan's pension (*Waisenrente*). In these cases, the variable *retirement status* was recoded to 0.

First-time retiree status: Information is based on the SOEP PBIOSPE file. A person was considered to enter the retiree status either between 2000 and 2004 or 2003 and 2004 if the variable *beginy* was > 1999 (or >2002) and the variable *spelltype* equals 8. Based on the first year of retirement and the year of birth, we checked whether the person was younger than 60. If this was the case, the person was not counted in the population of first-time retirees. It is assumed that these person receive an invalidity or disability pension (*Erwerbs- oder Berufsunfähigkeitsrente*) or an orphan's or survivor's pension (*Waisenrente oder Witwenrente*) and therefore don't belong to the population of interest, e.g. old-age retirees.

Home production: All years a person spent in home production independent of whether other activities were reported.

Home production (modified): Counts only those episodes as years of home production if no other activities were reported during this time.

Migration history: Individuals with a migration history are persons who have no German citizenship in 2005 or if individuals respond that they obtained German citizenship later (and not since birth). Individuals are also considered to be persons with a migration history if they report that they immigrated after 1948.

East vs. West Germany: Based on the variable *Bundesland* (*vbula*) indicating whether a person lives in the old or new Laender.

Civil servants: The occupational status indicates whether a person is a civil servant or not. If a person reported to be a civil servant between 2000 and 2005 (*stib00 – stib05*) in at least one of the last six years, the variable *civil servant* equals one.

Self-employed: The occupational status indicates whether a person is self-employed or not. If a person reported to be self-employed between 2000 and 2005 (*stib00 – stib05*) in at least one of the last six years, the variable *self-employed* equals one.

Table A6 Average Time Spent in Different Activities – German Citizens

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Men		Women		Men		Women		Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	1.75	3.03	1.31	2.13	2.70	4.21	1.37	2.11	2.05	2.86	1.38	2.13
Apprenticeship/Training	2.62	2.64	1.66	1.69	2.64	2.06	2.02	1.63	2.90	1.89	2.23	1.67
Military/Civilian Service	1.30	2.84	0.19	1.37	0.59	1.71	0.18	2.36	0.76	1.25	0.08	0.50
Full-time employed	37.82	7.46	22.87	14.70	38.16	7.11	20.75	14.09	37.42	6.59	20.54	13.72
Part-time employed	0.43	2.05	5.35	8.93	0.43	1.27	7.58	10.04	0.64	1.89	7.46	9.89
Unemployed	1.10	2.30	0.84	2.10	2.20	3.34	1.70	2.89	2.24	2.53	2.13	3.31
Home production	0.23	1.18	16.40	16.12	0.27	1.06	16.54	15.64	0.29	1.06	16.73	14.63
Home production (modified)	0.04	0.50	12.27	14.28	0.01	0.16	12.48	14.30	0.01	0.16	12.77	13.57
Retired	4.54	3.98	4.64	4.35	2.28	1.70	2.03	1.56	1.42	2.10	1.13	0.99
Other	0.44	2.44	0.95	3.58	0.17	0.52	0.59	2.71	0.19	0.47	0.40	1.25
Sum of Activities	51	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.97	3.65	0.91	3.36	1.80	3.17	2.30	4.41	3.37	4.66	2.88	2.67
Mean Age	71.37	6.77	73.07	7.48	65.07	2.17	65.05	2.35	63.59	1.86	63.74	1.93
Sum of Weights	6436520.09		8635570.99		1376688.64		1755168.17		467585.529		725210.011	
Number of observations	1827		2113		392		465		140		180	

Source: SOEP 2005; Author's calculations

Table A6 Average Time Spent in Different Activities – German Citizens (Continued)

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Men		Women		Men		Women		Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	1.70	2.80	1.34	2.15	2.32	3.42	1.35	2.09	2.13	2.89	1.41	2.16
Apprenticeship/Training	2.70	2.63	1.72	1.68	2.64	1.57	2.12	1.61	2.98	1.80	2.35	1.64
Military/Civilian Service	1.27	2.76	0.20	1.33	0.57	1.76	0.19	2.45	0.66	1.17	0.09	0.51
Full-time employed	37.97	7.20	22.63	14.67	39.03	5.59	20.55	14.08	37.41	6.42	20.46	13.78
Part-time employed	0.42	2.10	5.40	8.97	0.44	1.33	7.68	10.21	0.68	1.98	7.45	10.08
Unemployed	0.97	1.98	0.82	2.10	1.70	2.45	1.67	2.89	2.04	2.36	2.17	3.34
Home production	0.25	1.23	16.61	16.15	0.28	1.08	16.48	15.59	0.27	0.98	16.73	14.84
Home production (modified)	0.04	0.51	12.44	14.34	0.01	0.18	12.51	14.33	0.01	0.17	12.67	13.76
Retired	4.54	4.01	4.59	4.37	2.27	1.73	2.01	1.56	1.46	2.22	1.14	1.00
Other	0.43	2.52	0.94	3.66	0.18	0.53	0.59	2.77	0.20	0.49	0.36	1.12
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.96	3.68	0.92	3.42	1.83	3.14	2.33	4.52	3.43	4.50	2.91	2.74
Mean Age	71.41	6.71	73.17	7.54	65.04	2.19	65.02	2.37	63.51	1.86	63.73	1.96
Sum of Weights	5761265.56		8021728.66		1194423.76		1632365.91		412960.638		681393.712	
Number of observations	1631		1933		346		423		122		165	

Source: SOEP 2005; Author's calculations

Table A6 Average Time Spent in Different Activities – Persons with a Migration History (Continued)

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Men		Women		Men		Women		Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	2.20	4.50	1.00	1.80	5.22	7.15	1.53	2.38	1.49	2.58	0.87	1.47
Apprenticeship/Training	1.98	2.59	0.92	1.60	2.66	4.02	0.79	1.38	2.29	2.39	0.45	1.07
Military/Civilian Service	1.59	3.43	0.15	1.78	0.69	1.31	0.00	0.00	1.50	1.59	0.00	0.00
Full-time employed	36.54	9.28	25.96	14.79	32.49	11.95	23.46	14.12	37.47	7.97	21.82	13.20
Part-time employed	0.43	1.55	4.70	8.43	0.40	0.84	6.21	7.32	0.36	1.02	7.58	6.44
Unemployed	2.19	3.99	1.08	2.13	5.49	5.73	2.08	2.99	3.72	3.29	1.49	2.77
Home production	0.12	0.65	13.56	15.46	0.17	0.89	17.34	16.39	0.44	1.58	16.71	11.26
Home production (modified)	0.02	0.30	10.04	13.27	0.00	0.00	12.08	13.99	0.00	0.00	14.29	10.65
Retired	4.49	3.67	5.31	3.95	2.33	1.49	2.18	1.59	1.07	0.65	1.04	0.78
Other	0.54	1.48	1.07	2.33	0.10	0.43	0.70	1.80	0.11	0.27	1.12	2.49
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.01	3.37	0.78	2.37	1.62	3.39	1.96	2.59	2.99	5.85	2.35	1.27
Mean Age	71.10	7.26	71.88	6.55	65.22	2.09	65.34	2.14	64.13	1.82	63.82	1.30
Sum of Weights	675254.532		613842.328		182264.882		122802.26		54624.8904		43816.2995	
Number of observations	196		180		46		43		18		15	

Source: SOEP 2005; Author's calculations

Table A7 Comparison East vs. West Germany – Total

All pensioners in the year 2005	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	1.73	3.08	1.35	2.17	1.80	2.86	1.19	1.99
Apprenticeship/Training	2.64	2.78	1.62	1.68	2.56	2.08	1.79	1.72
Military/Civilian Service	1.26	2.82	0.24	1.54	1.46	2.93	0.05	0.39
Full-time employed	37.85	7.68	20.63	14.78	37.70	6.65	30.18	11.80
Part-time employed	0.44	2.18	5.74	9.23	0.37	1.51	4.07	7.76
Unemployed	1.12	2.45	0.71	1.99	1.00	1.71	1.25	2.37
Home production	0.27	1.30	19.81	16.34	0.12	0.59	5.24	8.61
Home production (modified)	0.05	0.55	14.77	14.95	0.02	0.23	4.11	7.25
Retired	4.51	4.05	4.14	3.99	4.63	3.72	6.27	5.01
Other	0.47	2.50	0.99	3.84	0.34	2.20	0.82	2.53
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.92	3.72	0.80	3.36	1.11	3.40	1.27	3.33
Mean Age	71.65	6.84	73.41	7.43	70.43	6.43	71.97	7.57
Sum of Weights	4981314.42		6611747.11		1455205.67		2023823.88	
Number of observations	1268		1423		559		690	

Source: SOEP 2005; Author's calculations

Table A7 Comparison East vs. West Germany – Total (Cont'd)

First-time pensioners between 2000 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	2.87	4.52	1.39	2.11	2.16	3.06	1.26	2.13
Apprenticeship/Training	2.74	2.22	1.96	1.70	2.36	1.41	2.25	1.27
Military/Civilian Service	0.48	1.81	0.22	2.65	0.91	1.29	0.00	0.00
Full-time employed	38.37	7.44	18.17	13.89	37.54	6.00	30.48	10.00
Part-time employed	0.45	1.33	8.37	10.42	0.37	1.07	4.59	7.76
Unemployed	2.14	3.62	1.10	2.33	2.41	2.26	3.98	3.59
Home production	0.33	1.19	20.09	15.53	0.08	0.38	3.15	5.59
Home production (modified)	0.01	0.19	15.13	14.78	0.00	0.00	2.51	5.11
Retired	2.30	1.81	1.87	1.55	2.23	1.32	2.60	1.48
Other	0.18	0.54	0.64	3.01	0.15	0.47	0.41	0.90
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.45	2.19	2.13	4.78	2.87	4.98	2.94	2.50
Mean Age	65.25	2.18	65.47	2.36	64.48	2.04	63.45	1.43
Sum of Weights	1039575.36		1387319.91		337113.279		367848.26	
Number of observations	266		328		126		137	

Source: SOEP 2005; Author's calculations

Table A7 Comparison East vs. West Germany – Total (Cont'd)

First-time pensioners between 2003 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	1.89	2.88	1.48	2.24	2.48	2.78	0.98	1.55
Apprenticeship/Training	3.03	2.03	2.24	1.75	2.54	1.39	2.21	1.29
Military/Civilian Service	0.67	1.16	0.10	0.55	1.01	1.44	0.00	0.00
Full-time employed	38.56	6.04	18.57	13.84	34.44	7.06	28.60	9.82
Part-time employed	0.74	2.06	8.39	10.52	0.37	1.36	3.63	5.36
Unemployed	1.88	2.47	1.30	2.47	3.17	2.48	5.49	4.11
Home production	0.39	1.23	19.60	14.55	0.00	0.00	5.02	7.32
Home production (modified)	0.02	0.19	14.86	14.04	0.00	0.00	4.25	6.57
Retired	1.51	2.40	1.11	1.04	1.16	0.92	1.25	0.76
Other	0.20	0.50	0.45	1.37	0.16	0.35	0.23	0.53
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	2.50	2.85	2.51	2.40	5.67	7.12	4.37	3.20
Mean Age	63.78	1.88	64.12	1.90	63.08	1.71	62.18	1.09
Sum of Weights	338237.769		582328.701		129347.759		142881.311	
Number of observations	91		124		49		56	

Source: SOEP 2005; Author's calculations

Table A8 Comparison East vs. West Germany – German Citizens

All pensioners in the year 2005	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	1.66	2.79	1.38	2.21	1.80	2.83	1.19	1.96
Apprenticeship/Training	2.75	2.80	1.68	1.67	2.55	2.07	1.83	1.73
Military/Civilian Service	1.21	2.70	0.24	1.51	1.45	2.94	0.06	0.40
Full-time employed	38.04	7.39	20.27	14.67	37.75	6.61	30.12	11.90
Part-time employed	0.44	2.26	5.81	9.28	0.37	1.52	4.11	7.79
Unemployed	0.96	2.06	0.68	1.98	1.00	1.71	1.26	2.39
Home production	0.29	1.37	20.22	16.31	0.13	0.60	5.22	8.60
Home production (modified)	0.05	0.58	15.06	15.01	0.02	0.23	4.16	7.32
Retired	4.52	4.12	4.07	3.98	4.61	3.68	6.23	5.09
Other	0.46	2.62	1.01	3.96	0.34	2.22	0.73	2.45
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.91	3.76	0.79	3.42	1.10	3.41	1.31	3.40
Mean Age	71.72	6.78	73.54	7.47	70.45	6.43	71.98	7.66
Sum of Weights	4330301.22		6094837.39		1430964.34		1926891.27	
Number of observations	1081		1268		550		665	

Source: SOEP 2005; Author's calculations

Table A8 Comparison East vs. West Germany – German Citizens (Cont'd)

First-time pensioners between 2000 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	2.36	3.55	1.40	2.12	2.19	3.08	1.19	1.96
Apprenticeship/Training	2.78	1.63	2.08	1.69	2.29	1.31	2.23	1.28
Military/Civilian Service	0.45	1.90	0.24	2.77	0.90	1.30	0.00	0.00
Full-time employed	39.58	5.31	17.61	13.77	37.59	6.05	30.88	9.59
Part-time employed	0.46	1.41	8.66	10.69	0.38	1.08	4.24	7.36
Unemployed	1.45	2.48	1.02	2.26	2.37	2.27	3.98	3.59
Home production	0.36	1.24	20.26	15.46	0.08	0.38	3.18	5.63
Home production (modified)	0.01	0.21	15.35	14.84	0.00	0.00	2.55	5.15
Retired	2.28	1.87	1.85	1.55	2.26	1.31	2.57	1.47
Other	0.20	0.55	0.64	3.10	0.14	0.47	0.40	0.90
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.43	1.84	2.15	4.93	2.88	5.04	2.97	2.50
Mean Age	65.25	2.21	65.48	2.38	64.51	2.05	63.42	1.42
Sum of Weights	864379.348		1270852.36		330044.409		361513.55	
Number of observations	222		289		124		134	

Source: SOEP 2005; Author's calculations

Table A8 Comparison East vs. West Germany – German Citizens (Cont'd)

First-time pensioners between 2003 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	1.94	2.92	1.52	2.29	2.58	2.81	0.98	1.56
Apprenticeship/Training	3.23	1.97	2.38	1.72	2.37	1.13	2.21	1.30
Military/Civilian Service	0.53	1.00	0.11	0.57	0.98	1.46	0.00	0.00
Full-time employed	38.68	5.60	18.29	13.88	34.39	7.25	28.74	9.77
Part-time employed	0.80	2.18	8.50	10.80	0.39	1.39	3.45	5.07
Unemployed	1.59	2.14	1.30	2.45	3.11	2.54	5.50	4.13
Home production	0.38	1.15	19.78	14.81	0.00	0.00	5.07	7.34
Home production (modified)	0.02	0.20	14.86	14.31	0.00	0.00	4.29	6.59
Retired	1.58	2.57	1.11	1.06	1.18	0.93	1.25	0.76
Other	0.22	0.53	0.40	1.23	0.15	0.35	0.21	0.50
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	2.41	1.80	2.53	2.47	5.85	7.28	4.37	3.22
Mean Age	63.71	1.88	64.14	1.94	63.06	1.75	62.18	1.09
Sum of Weights	290681.749		539963.221		122278.889		141430.491	
Number of observations	75		110		47		55	

Source: SOEP 2005; Author's calculations

Table A9 Comparison East vs. West Germany – Persons with Migration History

All Pensioners in the Year 2005	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	2.20	4.51	0.96	1.61	2.12	4.45	1.17	2.60
Apprenticeship/Training	1.94	2.59	0.90	1.64	3.13	2.55	1.01	1.41
Military/Civilian Service	1.57	3.47	0.18	1.94	2.19	1.88	0.00	0.00
Full-time employed	36.62	9.32	24.95	15.38	34.48	8.40	31.37	9.67
Part-time employed	0.45	1.58	4.96	8.62	0.08	0.19	3.30	7.38
Unemployed	2.22	4.04	1.07	2.16	1.32	2.02	1.14	2.05
Home production	0.12	0.67	15.03	16.00	0.00	0.00	5.76	8.92
Home production (modified)	0.02	0.30	11.36	13.87	0.00	0.00	2.96	5.64
Retired	4.44	3.58	4.96	4.06	5.85	5.86	7.18	2.64
Other	0.55	1.51	0.82	1.96	0.19	0.26	2.42	3.50
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.99	3.40	0.84	2.50	1.65	2.33	0.46	1.47
Mean Age	71.19	7.28	71.89	6.74	68.70	6.61	71.81	5.55
Sum of Weights	651013.20		516909.71		24241.33		96932.61	
Number of observations	187		155		9		25	

Source: SOEP 2005; Author's calculations

Table A9 Comparison East vs. West Germany – Persons with Migration History (Cont'd)

First-time pensioners between 2000 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	5.40	7.23	1.33	1.92	0.78	2.06	5.33	6.27
Apprenticeship/Training	2.55	4.04	0.67	1.31	5.50	2.65	2.88	1.12
Military/Civilian Service	0.65	1.31	0.00	0.00	1.55	1.18	0	0
Full-time employed	32.38	12.18	24.32	13.97	35.23	3.24	7.68	4.93
Part-time employed	0.42	0.85	5.22	6.07	0	0	24.45	2.15
Unemployed	5.54	5.84	1.96	2.90	4.17	0.88	4.26	4.44
Home production	0.18	0.91	18.24	16.37	0	0	0.96	1.22
Home production (modified)	0.00	0.00	12.74	14.08	0	0	0	0
Retired	2.39	1.49	2.08	1.52	0.83	0.88	4.11	1.82
Other	0.09	0.44	0.68	1.84	0.39	0.29	1.10	0.92
Sum of Activities	51.00	0.00	51.00	0.00	51	0	51	0
Years Missing	1.59	3.45	2.00	2.63	2.55	1.18	1.21	1.93
Mean Age	65.30	2.09	65.37	2.17	63.45	1.18	64.79	1.93
Sum of Weights	175196.012		116467.55		7068.87		6334.71	
Number of observations	44		39		2		3	

Source: SOEP 2005; Author's calculations

Table A9 Comparison East vs. West Germany – Persons with Migration History (Cont'd)

First-time pensioners between 2003 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	1.60	2.69	0.86	1.50	0.78	2.06	1	
Apprenticeship/Training	1.81	2.04	0.39	1.05	5.50	2.65	2	
Military/Civilian Service	1.49	1.68	0.00	0.00	1.55	1.18	0	
Full-time employed	37.81	8.47	22.06	13.39	35.23	3.24	14.83	
Part-time employed	0.41	1.08	7.10	5.99	0	0	21.33	
Unemployed	3.65	3.53	1.39	2.77	4.17	0.88	4.33	
Home production	0.50	1.69	17.28	11.00	0	0	0	
Home production (modified)	0.00	0.00	14.78	10.49	0	0	0	
Retired	1.11	0.65	1.02	0.79	0.83	0.88	1.50	
Other	0.07	0.25	1.09	2.53	0.39	0.29	2	
Sum of Activities	51.00	0.00	51.00	0.00	51	0	51	
Years Missing	3.05	6.28	2.30	1.25	2.55	1.18	4	
Mean Age	64.23	1.87	63.88	1.28	63.45	1.18	62	
Sum of Weights	47556.0202		42365.4796		7068.87		1450.82	
Number of observations	16		14		2		1	

Source: SOEP 2005; Author's calculations

Table A10 Comparison Civil Servant in Last Job vs. Others

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Civil Servants		Others		Civil Servants		Others		Civil Servants		Others	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	3.29	3.98	1.47	2.52	6.82	4.99	1.85	3.16	5.19	4.16	1.60	2.36
Apprenticeship/Training	2.51	1.88	2.06	2.20	4.61	10.05	2.27	1.84	2.31	1.77	2.47	1.76
Military/Civilian Service	1.53	2.97	0.65	2.18	0.57	0.88	0.34	2.08	0.96	1.05	0.32	0.88
Full-time employed	35.68	9.37	29.15	14.26	34.89	8.80	27.82	14.60	37.84	2.38	26.52	14.30
Part-time employed	0.67	2.64	3.29	7.35	0.71	2.08	4.70	8.71	0.77	1.61	5.00	8.67
Unemployed	0.15	0.62	0.96	2.21	0.06	0.29	1.92	3.10	0.06	0.25	2.20	3.05
Home production	2.26	8.29	9.61	14.66	1.71	7.26	10.15	14.80	0.16	0.44	10.79	14.15
Home production (modified)	1.83	7.46	7.13	12.43	0.00	0.00	7.44	12.74	0.00	0.00	8.08	12.40
Retired	5.05	5.70	4.59	4.16	1.60	1.38	2.18	1.77	1.47	1.66	1.36	1.97
Other	0.13	0.46	0.74	3.18	0.13	0.57	0.40	2.05	0.00	0.00	0.32	1.02
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.17	0.71	0.94	3.51	1.60	2.20	2.09	3.89	2.40	2.71	3.12	3.57
Mean Age	73.46	7.39	72.33	7.23	66.56	1.97	65.05	2.30	65.57	1.56	63.63	1.89
Sum of Weights	235778.969		14836312.1		264487.34		3193841.36		120773.061		1200670.21	
Number of observations	59		3883		43		867		17		324	

Source: SOEP 2005; Author's calculations

Table A11 Comparison Self-Employed in Last Job vs. Others

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Self-Employed		Others		Self-Employed		Others		Self-Employed		Others	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	3.13	3.67	1.46	2.51	3,35	3,89	2,26	3,46	3,46	4,58	2,28	3,02
Apprenticeship/Training	2.54	1.91	2.06	2.20	1,95	1,47	2,37	2,91	1,83	1,57	2,35	1,72
Military/Civilian Service	0.44	1.26	0.67	2.21	0,89	4,22	0,38	1,74	0,31	0,67	0,40	0,90
Full-time employed	35.69	11.64	29.09	14.24	29,84	13,84	26,28	13,99	27,99	15,64	24,62	13,95
Part-time employed	3.99	7.95	3.23	7.29	6,16	10,63	3,88	7,64	7,81	11,57	3,83	7,24
Unemployed	0.52	1.49	0.96	2.20	0,81	2,14	1,66	2,85	1,25	2,49	1,77	2,79
Home production	5.29	11.03	9.60	14.67	5,44	10,60	7,93	13,19	6,40	11,28	7,75	12,42
Home production (modified)	1.60	5.61	7.18	12.48	0,93	2,68	5,72	11,19	0,80	1,71	5,83	10,81
Retired	1.46	2.09	4.67	4.20	1,72	1,63	2,21	1,87	1,56	2,11	1,38	2,03
Other	1.00	3.74	0.73	3.14	0,76	3,58	0,64	2,41	0,16	0,40	0,59	1,47
Sum of Activities	51.00	0.00	51.00	0.00	51	0,00	51	0,00	51	0,00	51	0,00
Years Missing	0.64	1.19	0.94	3.52	4,57	7,29	5,60	8,45	5,84	8,94	7,95	9,82
Mean Age	69.42	5.94	72.42	7.25	65.93	2.27	65.12	2.31	64.70	1.80	63.75	1.94
Sum of Weights	365091.839		14706999.20		292883,82		5063520,53		115466,85		2130737,99	
Number of observations	93		3847		79		1272		32		513	

Source: SOEP 2005; Author's calculations

Table A12 Average Time Spent in Different Activities - Total

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Men		Women		Men		Women		Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	3.33	3.16	2.53	2.40	4.31	3.64	2.41	2.29	3.63	2.97	2.17	2.36
Apprenticeship/Training	3.14	2.59	2.74	1.31	3.01	1.30	2.78	1.24	3.20	1.67	2.93	1.27
Military/Civilian Service	3.70	3.64	2.63	4.16	2.17	2.90	11.58	21.49	2.06	1.15	3.00	.
Full-time employed	38.01	7.08	24.11	13.92	39.03	5.59	21.17	13.82	37.41	6.42	20.99	13.55
Part-time employed	2.45	4.52	11.29	10.08	1.73	2.18	11.63	10.58	2.51	3.19	11.12	10.54
Unemployed	2.68	2.50	3.37	3.07	3.17	2.56	4.13	3.23	3.54	2.08	4.43	3.58
Home production	2.79	3.14	21.23	15.34	2.33	2.21	20.42	14.86	1.89	2.00	20.01	14.05
Home production (modified)	3.66	3.13	18.69	13.86	2.69	1.35	18.15	13.98	2.00	.	16.74	13.48
Retired	4.92	3.95	5.18	4.30	2.27	1.73	2.01	1.56	1.46	2.22	1.14	1.00
Other	2.48	5.63	4.12	6.74	1.08	0.83	3.02	5.70	0.94	0.66	1.83	1.95
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.96	3.68	0.92	3.42	1.83	3.14	2.33	4.52	3.43	4.50	2.91	2.74
Mean Age	71.41	6.71	73.17	7.54	65.04	2.19	65.02	2.37	63.51	1.86	63.73	1.96
Sum of Weights	5761265.56		8021728.66		1194423.76		1632365.91		412960.638		681393.712	
Number of observations	1631		1933		346		423		122		165	

Source: SOEP 2005; Author's calculations

Table A12 Average Time Spent in Different Activities – German Citizens (Cont'd)

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Men		Women		Men		Women		Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	3.48	3.50	2.53	2.38	4.92	4.63	2.43	2.31	3.55	2.97	2.17	2.33
Apprenticeship/Training	3.15	2.59	2.74	1.32	3.10	1.88	2.78	1.23	3.22	1.70	2.93	1.27
Military/Civilian Service	3.69	3.75	2.73	4.41	2.24	2.74	11.58	21.49	2.18	1.16	3.00	.
Full-time employed	37.86	7.36	24.33	13.94	38.16	7.11	21.34	13.84	37.42	6.59	21.04	13.50
Part-time employed	2.40	4.35	11.20	10.08	1.67	2.05	11.39	10.39	2.54	3.10	10.91	10.28
Unemployed	2.98	2.96	3.35	3.03	3.98	3.61	4.11	3.22	3.77	2.24	4.35	3.57
Home production	2.71	3.07	21.09	15.34	2.32	2.22	20.56	14.87	2.08	2.18	20.06	13.77
Home production (modified)	3.52	3.09	18.74	13.78	2.69	1.35	18.32	13.89	2.00	.	16.81	13.21
Retired	4.92	3.91	5.23	4.27	2.28	1.70	2.03	1.56	1.42	2.10	1.13	0.99
Other	2.41	5.26	4.08	6.49	1.03	0.85	2.94	5.45	0.91	0.64	1.99	2.15
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.97	3.65	0.91	3.36	1.80	3.17	2.30	4.41	3.37	4.66	2.88	2.67
Mean Age	71.37	6.77	73.07	7.48	65.07	2.17	65.05	2.35	63.59	1.86	63.74	1.93
Sum of Weights	6436520.09		8635570.99		1376688.64		1755168.17		467585.529		725210.011	
Number of observations	1827		2113		392		465		140		180	

Source: SOEP 2005; Author's calculations

Table A12 Average Time Spent in Different Activities – Individuals with Migration History (Cont'd)

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Men		Women		Men		Women		Men		Women	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	4.88	5.67	2.48	2.09	8.34	7.55	2.65	2.64	2.88	3.04	2.24	1.58
Apprenticeship/Training	3.26	2.63	2.86	1.57	3.87	4.36	2.87	0.94	3.43	2.17	2.61	0.89
Military/Civilian Service	3.57	4.40	7.92	10.88	2.73	1.05			2.66	1.15		
Full-time employed	36.54	9.28	27.18	13.99	32.49	11.95	23.54	14.08	37.47	7.97	21.82	13.20
Part-time employed	2.06	2.87	9.95	9.93	1.35	1.07	8.51	7.33	3.02	0.58	8.38	6.27
Unemployed	5.08	4.71	3.15	2.60	8.21	5.15	3.82	3.12	5.18	2.68	3.08	3.34
Home production	1.83	1.95	18.99	15.21	2.27	2.60	22.51	15.20	3.95	3.90	20.78	8.25
Home production (modified)	1.98	2.58	19.67	12.45			21.06	12.22			17.77	8.72
Retired	4.92	3.56	5.89	3.72	2.33	1.49	2.18	1.59	1.07	0.65	1.04	0.78
Other	2.02	2.30	3.67	3.03	0.68	0.98	2.29	2.69	0.65	0.28	3.56	3.38
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.01	3.37	0.78	2.37	1.62	3.39	1.96	2.59	2.99	5.85	2.35	1.27
Mean Age	71.10	7.26	71.88	6.55	65.22	2.09	65.34	2.14	64.13	1.82	63.82	1.30
Sum of Weights	675254.532		613842.328		182264.882		122802.26		54624.8904		43816.2995	
Number of observations	196		180		46		43		18		15	

Source: SOEP 2005; Author's calculations

Table A13 Comparison East vs. West Germany – Total

All pensioners in 2005	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	3.43	3.59	2.53	2.41	3.65	3.13	2.52	2.25
Apprenticeship/Training	3.22	2.75	2.78	1.27	2.93	1.96	2.64	1.46
Military/Civilian Service	3.75	3.78	2.93	4.66	3.51	3.66	1.38	1.51
Full-time employed	37.90	7.55	22.28	14.11	37.70	6.65	30.60	11.34
Part-time employed	2.57	4.72	11.95	10.15	1.88	2.97	8.67	9.41
Unemployed	3.06	3.22	3.26	3.14	2.70	1.82	3.54	2.79
Home production	2.84	3.29	23.72	15.07	2.03	1.37	8.90	9.66
Home production (modified)	3.76	3.32	20.68	13.81	2.30	1.05	8.93	8.45
Retired	4.89	3.99	4.78	3.92	5.05	3.61	6.55	4.94
Other	2.40	5.22	4.28	7.04	2.45	5.50	3.43	4.25
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.92	3.72	0.80	3.36	1.11	3.40	1.27	3.33
Mean Age	71.65	6.84	73.41	7.43	70.43	6.43	71.97	7.57
Sum of Weights	4981314.42		6611747.11		1455205.67		2023823.88	
Number of observations	1268		1423		559		690	

Source: SOEP 2005; Author's calculations

Table A13 Comparison East vs. West Germany – Total (Cont'd)

First-time pensioners between 2000 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	5.13	4.99	2.39	2.29	4.20	3.10	2.63	2.43
Apprenticeship/Training	3.22	2.06	2.84	1.30	2.73	1.13	2.62	0.96
Military/Civilian Service	2.27	3.40	11.58	21.49	2.19	1.10		
Full-time employed	38.37	7.44	18.83	13.70	37.54	6.00	30.48	10.00
Part-time employed	1.77	2.16	12.64	10.50	1.38	1.70	6.79	8.62
Unemployed	4.36	4.13	3.54	3.00	3.20	2.06	4.92	3.37
Home production	2.37	2.32	22.98	14.46	1.81	0.57	5.82	6.51
Home production (modified)	2.69	1.35	19.99	13.83			6.33	6.49
Retired	2.30	1.81	1.87	1.55	2.23	1.32	2.60	1.48
Other	1.07	0.87	3.40	6.25	0.92	0.80	1.63	1.14
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.45	2.19	2.13	4.78	2.87	4.98	2.94	2.50
Mean Age	65.25	2.18	65.47	2.36	64.48	2.04	63.45	1.43
Sum of Weights	1039575.36		1387319.91		337113.279		367848.26	
Number of observations	266		328		126		137	

Source: SOEP 2005; Author's calculations

Table A13 Comparison East vs. West Germany – Total (Cont'd)

First-time pensioners between 2003 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	3.43	3.14	2.24	2.44	3.81	2.60	1.83	1.71
Apprenticeship/Training	3.45	1.80	3.02	1.34	2.67	1.30	2.62	0.94
Military/Civilian Service	2.00	1.16	3.00	.	2.55	1.10		
Full-time employed	38.56	6.04	19.13	13.66	34.44	7.06	28.60	9.82
Part-time employed	2.77	3.24	12.48	10.66	1.77	2.59	4.98	5.73
Unemployed	3.67	2.31	3.31	2.98	3.95	2.13	6.24	3.81
Home production	2.08	2.18	22.10	13.54			8.10	7.86
Home production (modified)	2.00	.	18.57	13.31			7.14	7.23
Retired	1.51	2.40	1.11	1.04	1.16	0.92	1.25	0.76
Other	1.00	0.71	2.19	2.34	0.72	0.40	1.14	0.59
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	2.50	2.85	2.51	2.40	5.67	7.12	4.37	3.20
Mean Age	63.78	1.88	64.12	1.90	63.08	1.71	62.18	1.09
Sum of Weights	338237.769		582328.701		129347.759		142881.311	
Number of observations	91		124		49		56	

Source: SOEP 2005; Author's calculations

Table A14 Comparison East vs. West Germany – German Citizens

All pensioners in 2005	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	3.24	3.18	2.55	2.45	3.63	3.09	2.47	2.20
Apprenticeship/Training	3.22	2.76	2.77	1.25	2.92	1.96	2.64	1.47
Military/Civilian Service	3.78	3.62	2.81	4.40	3.53	3.70	1.38	1.51
Full-time employed	38.10	7.23	21.93	14.01	37.75	6.61	30.56	11.42
Part-time employed	2.66	4.99	12.16	10.14	1.89	2.98	8.54	9.39
Unemployed	2.68	2.69	3.28	3.20	2.68	1.82	3.54	2.82
Home production	2.94	3.38	23.85	15.07	2.03	1.37	9.06	9.68
Home production (modified)	3.96	3.37	20.67	13.89	2.30	1.05	8.91	8.52
Retired	4.89	4.06	4.71	3.92	5.04	3.56	6.50	5.03
Other	2.47	5.65	4.39	7.32	2.55	5.61	3.24	4.30
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.91	3.76	0.79	3.42	1.10	3.41	1.31	3.40
Mean Age	71.72	6.78	73.54	7.47	70.45	6.43	71.98	7.66
Sum of Weights	4330301.22		6094837.39		1430964.34		1926891.27	
Number of observations	1081		1268		550		665	

Source: SOEP 2005; Author's calculations

Table A14 Comparison East vs. West Germany – German Citizens (Cont'd)

First-time pensioners between 2000 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	4.35	3.82	2.39	2.31	4.21	3.12	2.50	2.20
Apprenticeship/Training	3.13	1.38	2.84	1.31	2.66	1.01	2.61	0.96
Military/Civilian Service	2.15	3.73	11.58	21.49	2.20	1.12		
Full-time employed	39.58	5.31	18.30	13.58	37.59	6.05	30.88	9.59
Part-time employed	1.88	2.36	13.17	10.70	1.38	1.70	6.33	8.23
Unemployed	3.17	2.83	3.50	3.00	3.17	2.08	4.93	3.36
Home production	2.38	2.32	22.96	14.44	1.81	0.57	5.88	6.54
Home production (modified)	2.69	1.35	19.91	13.95			6.33	6.49
Retired	2.28	1.87	1.85	1.55	2.26	1.31	2.57	1.47
Other	1.11	0.83	3.55	6.61	0.97	0.83	1.64	1.17
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.43	1.84	2.15	4.93	2.88	5.04	2.97	2.50
Mean Age	65.25	2.21	65.48	2.38	64.51	2.05	63.42	1.42
Sum of Weights	864379.348		1270852.36		330044.409		361513.55	
Number of observations	222		289		124		134	

Source: SOEP 2005; Author's calculations

Table A14 Comparison East vs. West Germany – German Citizens (Cont'd)

First-time pensioners between 2003 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	3.53	3.16	2.24	2.47	3.82	2.63	1.84	1.72
Apprenticeship/Training	3.51	1.80	3.02	1.34	2.50	1.02	2.62	0.94
Military/Civilian Service	1.77	1.06	3.00	.	2.61	1.15		
Full-time employed	38.68	5.60	18.90	13.69	34.39	7.25	28.74	9.77
Part-time employed	2.75	3.37	12.97	10.97	1.77	2.59	4.75	5.42
Unemployed	3.27	1.98	3.34	2.96	3.94	2.20	6.26	3.83
Home production	1.89	2.00	22.20	13.86			8.10	7.86
Home production (modified)	2.00	.	18.64	13.65			7.14	7.23
Retired	1.58	2.57	1.11	1.06	1.18	0.93	1.25	0.76
Other	1.01	0.73	2.01	2.14	0.77	0.43	1.10	0.57
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	2.41	1.80	2.53	2.47	5.85	7.28	4.37	3.22
Mean Age	63.71	1.88	64.14	1.94	63.06	1.75	62.18	1.09
Sum of Weights	290681.749		539963.221		122278.889		141430.491	
Number of observations	75		110		47		55	

Source: SOEP 2005; Author's calculations

Table A15 Comparison East vs. West Germany – Persons with a Migration History

All Pensioners in the Year 2005	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	4.87	5.69	2.27	1.78	2.12	4.45	4.13	3.50
Apprenticeship/Training	3.22	2.65	2.94	1.68	3.13	2.55	2.55	1.00
Military/Civilian Service	3.61	4.52	7.92	10.88	2.19	1.88		
Full-time employed	36.62	9.32	26.35	14.58	34.48	8.40	31.37	9.67
Part-time employed	2.10	2.89	9.61	9.97	0.08	0.19	13.79	9.12
Unemployed	5.12	4.77	3.08	2.69	1.32	2.02	3.59	2.11
Home production	1.83	1.95	21.82	14.93	0.00	0.00	6.78	9.38
Home production (modified)	1.98	2.58	20.70	12.50	0.00	0.00	9.75	6.14
Retired	4.88	3.45	5.59	3.88	5.85	5.86	7.36	2.40
Other	2.10	2.33	3.12	2.73	0.19	0.26	5.35	3.45
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.99	3.40	0.84	2.50	1.65	2.33	0.46	1.47
Mean Age	71.19	7.28	71.89	6.74	68.70	6.61	71.81	5.55
Sum of Weights	651013.20		516909.71		24241.33		96932.61	
Number of observations	187		155		9		25	

Source: SOEP 2005; Author's calculations

Table A15 Comparison East vs. West Germany – Persons with a Migration History (Cont'd)

First-time pensioners between 2000 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	8.41	7.59	2.32	2.05	3.50	.	7.53	6.40
Apprenticeship/Training	3.77	4.45	2.87	0.94	5.50	2.65	2.88	1.12
Military/Civilian Service	2.83	1.08			2.00	.		
Full-time employed	32.38	12.18	24.40	13.92	35.23	3.24	7.68	4.93
Part-time employed	1.35	1.07	7.30	6.02			24.45	2.15
Unemployed	8.46	5.20	3.77	3.06	4.17	0.88	4.26	4.44
Home production	2.27	2.60	23.19	14.97			2.00	.
Home production (modified)			21.06	12.22				
Retired	2.39	1.49	2.08	1.52	0.83	0.88	4.11	1.82
Other	0.72	1.11	2.39	2.86	0.50	.	1.55	0.44
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	1.59	3.45	2.00	2.63	2.55	1.18	1.21	1.93
Mean Age	65.30	2.09	65.37	2.17	63.45	1.18	64.79	1.93
Sum of Weights	175196.012		116467.55		7068.87		6334.71	
Number of observations	44		39		2		3	

Source: SOEP 2005; Author's calculations

Table A15 Comparison East vs. West Germany – Persons with a Migration History (Cont'd)

First-time pensioners between 2003 and 2004	Men West Mean	Std.	Women West Mean	Std.	Men East Mean	Std.	Women East Mean	Std.
School/University	2.84	3.16	2.35	1.62	3.50	.	1.00	.
Apprenticeship/Training	2.93	1.89	2.76	0.97	5.50	2.65	2.00	.
Military/Civilian Service	2.80	1.23			2.00	.		
Full-time employed	37.81	8.47	22.06	13.39	35.23	3.24	14.83	.
Part-time employed	3.02	0.58	7.89	5.82			21.33	.
Unemployed	5.40	2.91	2.98	3.46	4.17	0.88	4.33	.
Home production	3.95	3.90	20.78	8.25				
Home production (modified)			17.77	8.72				
Retired	1.11	0.65	1.02	0.79	0.83	0.88	1.50	.
Other	0.86	0.31	3.75	3.57	0.50	.	2.00	.
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	.
Years Missing	3.05	6.28	2.30	1.25	2.55	1.18	4.00	.
Mean Age	64.23	1.91	63.88	1.28	63.45	1.18	62.00	.
Sum of Weights	47556.0202		42365.4796		7068.87		1450.82	
Number of observations	16		14		2		1	

Source: SOEP 2005; Author's calculations

Table A16 Comparison Civil Servant in Last Job vs. Others

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Civil Servants		Others		Civil Servants		Others		Civil Servants		Others	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	4.40	4.05	2.89	2.90	7.55	4.70	3.40	3.61	6.13	3.82	2.61	2.54
Apprenticeship/Training	3.30	1.42	2.94	2.08	6.50	11.46	2.92	1.56	2.92	1.46	3.03	1.44
Military/Civilian Service	5.06	3.39	3.44	3.92	1.39	0.88	2.86	5.42	1.44	1.01	2.18	1.08
Full-time employed	35.68	9.37	30.22	13.36	34.89	8.80	28.29	14.26	37.84	2.38	26.91	14.04
Part-time employed	4.85	5.70	9.32	9.85	1.77	3.03	9.52	10.39	2.08	2.12	9.49	10.02
Unemployed	1.48	1.45	3.17	3.00	1.05	0.72	4.04	3.42	1.00	.	4.14	3.08
Home production	17.10	17.14	19.69	15.55	12.56	17.08	19.36	15.47	1.18	0.54	18.53	14.16
Home production (modified)	22.71	15.98	18.53	13.80			18.57	14.09			16.63	13.21
Retired	5.84	5.74	5.08	4.09	1.60	1.38	2.18	1.77	1.47	1.66	1.36	1.97
Other	1.47	0.76	3.48	6.14	1.87	1.37	2.17	4.35			1.57	1.79
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.17	0.71	0.94	3.51	1.60	2.20	2.09	3.89	2.40	2.71	3.12	3.57
Mean Age	73.46	7.39	72.33	7.23	66.56	1.97	65.05	2.30	65.57	1.56	63.63	1.89
Sum of Weights	235778.969		14836312.1		264487.34		3193841.36		120773.061		1200670.21	
Number of observations	59		3883		43		867		17		324	

Source: SOEP 2005; Author's calculations

Table A17 Comparison Self-Employed in Last Job vs. Others

	All pensioners in the year 2005				First time pensioners b/w 2000-2004				First time pensioners b/w 2003-2004			
Type of Activity	Self-Employed		Others		Self-Employed		Others		Self-Employed		Others	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
School/University	4.60	3.60	2.87	2.90	5.22	3.89	3.40	3.68	6.09	4.98	2.49	2.35
Apprenticeship/Training	3.18	1.59	2.94	2.08	2.59	1.02	2.95	1.61	2.39	1.23	3.09	1.48
Military/Civilian Service	3.14	1.76	3.49	3.94	1.57	0.83	2.91	5.41	1.40	0.69	2.29	1.13
Full-time employed	35.69	11.64	30.17	13.34	34.50	13.02	28.51	14.08	32.27	14.79	27.27	13.74
Part-time employed	7.49	9.65	9.36	9.84	8.98	9.84	9.13	10.06	13.96	12.70	8.93	9.60
Unemployed	2.03	2.41	3.18	3.01	2.37	2.75	4.12	3.43	1.91	1.59	4.24	3.08
Home production	14.60	14.23	19.78	15.56	12.16	11.03	19.06	15.28	20.74	13.03	18.22	14.24
Home production (modified)	8.04	10.59	18.68	13.79	3.93	3.61	18.88	13.85	3.33	1.16	17.26	13.20
Retired	1.93	2.21	5.16	4.12	1.40	1.33	2.18	1.63	1.01	1.52	1.26	1.53
Other	5.84	7.42	3.42	6.08	6.16	9.80	1.97	3.80	0.90	0.73	1.60	1.82
Sum of Activities	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
Years Missing	0.64	1.19	0.94	3.52	0.92	1.22	2.15	4.01	1.23	1.40	3.18	3.65
Mean Age	69.42	5.94	72.42	7.25	65.97	2.12	65.00	2.27	65.05	1.70	63.59	1.88
Sum of Weights	365091.839		14706999.20		174859.70		2956997.11		68277.96		1124517.58	
Number of observations	93		3847		48		809		17		303	

Source: SOEP 2005; Author's calculations

Notes for the following Tables:

Tables A18 to *A23*: The tables provide results on the average time spent in different activities for all old-age retirees in the SUF VVL 2004. The tables compare men with women, Germans with persons with a migration history, and West Germans with East Germans. The calculation of the average time is based on all individuals in the respective demographic group, independent of whether they have actually spent time in a certain activity or not. Hence, the denominator is always the total number of individuals in the group.

Tables A24 to *A29*: The tables provide results on the average time spent in different activities for all old-age retirees in the SUF VVL 2004. The tables compare men with women, Germans with persons with a migration history, and West Germans with East Germans. The calculation of the average time is based on all individuals in the respective demographic group, who have spent time in a certain activity. Hence, the denominator includes all individuals who have information > 0 for the respective category.

Table A18 Average Time Spent in Different Activities – Total

First-Time Old-Age Pensioners in 2004 (West and East Germany)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	0.99	2.09	0.45	1.32
Apprenticeship/Training	2.01	1.41	0.80	1.25
Non-Professional Caretaking	0.02	0.29	0.37	1.40
Childcare/Household	0.02	0.44	8.02	6.61
Invalidity and Sickness	0.29	0.58	0.22	0.50
Unemployed	1.73	2.83	1.62	2.78
Military and Civilian Service	0.44	0.68	0.00	0.00
Marginal Employment	0.06	0.45	0.41	1.22
Self-Employed	0.29	1.94	0.07	0.96
Other	1.86	5.69	0.86	3.35
Employment Subject to Insurance Contributions	33.03	12.32	21.01	13.86
Invalidity	0.32	0.63	0.26	0.74
Pension	0.30	0.53	0.34	0.44
Missings	10.64	11.04	17.57	10.88
Mean Age	62.86	1.89	62.79	2.14
Sum	52	0	52	0
Number of Observations	14274		17200	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A19 Average Time Spent in Different Activities – Total, East Germany vs. West Germany

First-Time Old-Age Pensioners in 2004 (West Germany)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	0.92	2.04	0.37	1.22
Apprenticeship/Training	2.06	1.46	0.64	1.19
Non-Professional Caretaking	0.02	0.29	0.41	1.48
Childcare/Household	0.03	0.45	9.31	6.60
Invalidity and Sickness	0.24	0.56	0.13	0.38
Unemployed	1.44	2.77	1.17	2.38
Military and Civilian Service	0.43	0.67	0	0.00
Marginal Employment	0.07	0.50	0.49	1.33
Self-Employed	0.32	2.12	0.06	0.99
Other	2.18	6.28	0.97	3.61
Employment Subject to Insurance Contributions	31.74	13.36	17.91	13.34
Invalidity	0.31	0.68	0.26	0.79
Pension	0.31	0.51	0.33	0.43
Missings	11.94	12.01	19.94	10.65
Mean Age	63.00	1.89	63.19	2.10
Sum	52	0	52	0
Number of Observations	10727		13486	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A19 Average Time Spent in Different Activities – Total, East Germany vs. West Germany (Cont'd)

First-Time Old-Age Pensioners in 2004 (East Germany)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	1.18	2.21	0.74	1.62
Apprenticeship/Training	1.87	1.27	1.39	1.26
Non-Professional Caretaking	0.02	0.29	0.20	1.04
Childcare/Household	0.01	0.43	3.34	4.03
Invalidity and Sickness	0.45	0.62	0.58	0.70
Unemployed	2.59	2.84	3.26	3.42
Military and Civilian Service	0.49	0.71	0.00	0.00
Marginal Employment	0.03	0.25	0.12	0.63
Self-Employed	0.20	1.24	0.09	0.84
Other	0.89	3.10	0.44	2.15
Employment Subject to Insurance Contributions	36.94	7.18	32.25	9.06
Invalidity	0.36	0.43	0.26	0.51
Pension	0.26	0.56	0.38	0.47
Missings	6.71	5.82	8.95	6.47
Mean Age	62.44	1.80	61.34	1.59
Sum	52	0	52	0
Number of Observations	3547		3714	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A20 Average Time Spent in Different Activities – Total, German Citizens

First-Time Old-Age Pensioners in 2004 (Only Germans)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	0.99	2.09	0.45	1.32
Apprenticeship/Training	2.09	1.39	0.82	1.26
Non-Professional Caretaking	0.02	0.30	0.37	1.40
Childcare/Household	0.02	0.40	8.15	6.62
Invalidity and Sickness	0.29	0.59	0.22	0.50
Unemployed	1.68	2.78	1.59	2.74
Military and Civilian Service	0.44	0.65	0.00	0.00
Marginal Employment	0.06	0.46	0.41	1.23
Self-Employed	0.30	1.97	0.07	0.97
Other	1.91	5.78	0.87	3.40
Employment Subject to Insurance Contributions	33.10	12.41	20.89	13.92
Invalidity	0.31	0.62	0.24	0.66
Pension	0.30	0.52	0.35	0.43
Missings	10.50	11.09	17.56	10.90
Mean Age	62.86	1.88	62.81	2.14
Sum	52	0	52	0
Number of Observations	13696		16668	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A21 Average Time Spent in Different Activities – East vs. West Germany, German Citizens

First-Time Old-Age Pensioners in 2004 (West Germany, only Germans)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	0.92	2.05	0.36	1.20
Apprenticeship/Training	2.15	1.42	0.66	1.21
Non-Professional Caretaking	0.02	0.30	0.42	1.49
Childcare/Household	0.03	0.46	9.51	6.57
Invalidity and Sickness	0.24	0.57	0.12	0.37
Unemployed	1.37	2.70	1.11	2.31
Military and Civilian Service	0.42	0.63	0.00	0.00
Marginal Employment	0.07	0.51	0.50	1.34
Self-Employed	0.34	2.16	0.06	1.00
Other	2.25	6.41	0.99	3.66
Employment Subject to Insurance Contributions	31.73	13.51	17.66	13.35
Invalidity	0.30	0.67	0.24	0.71
Pension	0.32	0.50	0.34	0.42
Missings	11.84	12.13	20.03	10.66
Mean Age	63.01	1.89	63.22	2.09
Sum	52	0	52	0
Number of Observations	10196		12996	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A21 Average Time Spent in Different Activities – East vs. West Germany, German Citizens (Cont'd)

First-Time Old-Age Pensioners in 2004 (East Germany, only Germans)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	1.18	2.20	0.74	1.62
Apprenticeship/Training	1.89	1.26	1.40	1.26
Non-Professional Caretaking	0.02	0.29	0.20	1.03
Childcare/Household	0.00	0.07	3.35	4.04
Invalidity and Sickness	0.45	0.62	0.58	0.70
Unemployed	2.57	2.81	3.27	3.42
Military and Civilian Service	0.50	0.71	0.00	0.00
Marginal Employment	0.03	0.25	0.12	0.63
Self-Employed	0.20	1.25	0.09	0.84
Other	0.90	3.12	0.44	2.16
Employment Subject to Insurance Contributions	37.08	7.00	32.33	9.00
Invalidity	0.36	0.43	0.25	0.46
Pension	0.26	0.56	0.38	0.46
Missings	6.57	5.63	8.85	6.32
Mean Age	62.43	1.80	61.33	1.59
Sum	52	0	52	0
Number of Observations	3500		3672	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A22 Average Time Spent in Different Activities - Total, Persons with Migration History

First-Time Old-Age Pensioners in 2004 (Only Persons with Migration History)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	0.90	2.04	0.59	1.48
Apprenticeship/Training	0.23	0.72	0.12	0.44
Non-Professional Caretaking	0.02	0.25	0.30	1.41
Childcare/Household	0.07	1.05	4.01	4.87
Invalidity and Sickness	0.30	0.53	0.30	0.53
Unemployed	2.90	3.75	2.62	3.61
Military and Civilian Service	0.59	1.15	0.00	0.00
Marginal Employment	0.04	0.36	0.21	0.83
Self-Employed	0.09	1.19	0.02	0.36
Other	0.71	2.42	0.39	1.38
Employment Subject to Insurance Contributions	31.45	10.07	24.66	11.39
Invalidity	0.57	0.86	0.92	1.93
Pension	0.15	0.67	0.19	0.55
Missings	13.98	9.25	17.68	10.24
Mean Age	62.94	1.91	62.20	1.96
Sum	52	0	52	0
Number of Observations	578		532	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A23 Average Time Spent in Different Activities - East vs. West Germany, Persons with Migration History

First-Time Old-Age Pensioners in 2004 (West Germany, Only Persons with Migration History)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	0.86	1.97	0.59	1.47
Apprenticeship/Training	0.23	0.72	0.10	0.41
Non-Professional Caretaking	0.02	0.23	0.29	1.38
Childcare/Household	0.01	0.10	4.14	4.97
Invalidity and Sickness	0.29	0.53	0.29	0.52
Unemployed	2.75	3.68	2.59	3.61
Military and Civilian Service	0.62	1.17	0.00	0.00
Marginal Employment	0.04	0.37	0.21	0.85
Self-Employed	0.09	1.24	0.02	0.37
Other	0.74	2.49	0.39	1.42
Employment Subject to Insurance Contributions	31.90	9.82	24.59	11.38
Invalidity	0.58	0.88	0.89	1.92
Pension	0.15	0.68	0.17	0.50
Missings	13.73	9.18	17.72	10.10
Mean Age	62.96	1.90	62.20	1.96
Sum	52	0	52	0
Number of Observations	531		490	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A23 Average Time Spent in Different Activities - East vs. West Germany, Persons with Migration History (Cont'd)

First-Time Old-Age Pensioners in 2004 (East Germany, Only Persons with Migration History)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	1.37	2.68	0.54	1.62
Apprenticeship/Training	0.22	0.69	0.32	0.64
Non-Professional Caretaking	0.06	0.44	0.38	1.73
Childcare/Household	0.76	3.62	2.53	3.10
Invalidity and Sickness	0.38	0.54	0.41	0.61
Unemployed	4.58	4.22	3.00	3.62
Military and Civilian Service	0.25	0.80	0.00	0.00
Marginal Employment	0.01	0.06	0.14	0.53
Self-Employed	0.00	0.00	0.00	0.00
Other	0.37	1.31	0.33	0.78
Employment Subject to Insurance Contributions	26.46	11.54	25.53	11.51
Invalidity	0.52	0.55	1.33	1.93
Pension	0.18	0.54	0.34	0.97
Missings	16.83	9.69	17.15	11.89
Mean Age	62.79	2.03	62.14	1.97
Sum	52	0	52	0
Number of Observations	47		42	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A24 Average Time Spent in Different Activities – Total

First-Time Old-Age Pensioners in 2004 (Total)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.26	2.64	2.30	2.16
Apprenticeship/Training	2.82	0.73	2.49	0.79
Non-Professional Caretaking	2.46	2.37	3.20	2.83
Childcare/Household	3.51	4.15	9.04	6.32
Invalidity and Sickness	0.62	0.72	0.61	0.67
Unemployed	3.38	3.18	3.60	3.17
Military and Civilian Service	1.28	0.51	0.00	0.00
Marginal Employment	1.64	1.72	2.82	1.88
Self-Employed	5.99	6.59	5.42	6.84
Other	9.36	9.64	4.99	6.70
Employment Subject to Insurance Contributions	33.12	12.22	21.97	13.41
Invalidity	0.61	0.75	0.64	1.05
Pension	0.64	0.62	0.59	0.43
Missings	10.64	11.04	17.57	10.88
Mean Age	52.00	0.00	52.00	0.00
Sum	62.86	1.89	62.79	2.14
Number of Observations	14274		17200	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A25 Average Time Spent in Different Activities – Total, East Germany vs. West Germany

First-Time Old-Age Pensioners in 2004 (Total, West Germany)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.12	2.70	2.21	2.18
Apprenticeship/Training	2.92	0.70	2.62	0.79
Non-Professional Caretaking	2.39	2.46	3.26	2.83
Childcare/Household	3.22	3.65	10.55	6.02
Invalidity and Sickness	0.62	0.77	0.51	0.61
Unemployed	3.13	3.38	3.03	3.02
Military and Civilian Service	1.23	0.56		
Marginal Employment	1.82	1.79	2.94	1.86
Self-Employed	7.66	7.12	8.30	8.34
Other	10.07	10.15	5.10	6.88
Employment Subject to Insurance Contributions	31.85	13.25	18.97	12.98
Invalidity	0.61	0.85	0.63	1.13
Pension	0.63	0.58	0.58	0.42
Missings	11.94	12.01	19.94	10.65
Mean Age	52.00	0.00	52.00	0.00
Sum	63.00	1.89	63.19	2.10
Number of Observations	10727		13486	

First-Time Old-Age Pensioners in 2004 (Total, East Germany)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.65	2.46	2.48	2.11
Apprenticeship/Training	2.51	0.74	2.30	0.74
Non-Professional Caretaking	2.66	2.11	2.76	2.77
Childcare/Household	9.98	9.10	3.70	4.08
Invalidity and Sickness	0.61	0.65	0.72	0.71
Unemployed	3.89	2.66	4.78	3.14
Military and Civilian Service	1.44	0.31		
Marginal Employment	0.89	1.14	1.76	1.67
Self-Employed	2.87	3.87	2.97	3.83
Other	6.14	5.86	4.26	5.32
Employment Subject to Insurance Contributions	36.94	7.18	32.28	9.02
Invalidity	0.61	0.40	0.66	0.63
Pension	0.65	0.74	0.62	0.45
Missings	6.71	5.82	8.95	6.47
Mean Age	52.00	0.00	52.00	0.00
Sum	62.44	1.80	61.34	1.59
Number of Observations	3547		3714	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A26 Average Time Spent in Different Activities – Total, German Citizens

First-Time Old-Age Pensioners in 2004 (Total, Only German Citizens)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.26	2.64	2.29	2.16
Apprenticeship/Training	2.83	0.71	2.51	0.77
Non-Professional Caretaking	2.45	2.42	3.17	2.81
Childcare/Household	3.27	3.66	9.19	6.31
Invalidity and Sickness	0.62	0.73	0.61	0.67
Unemployed	3.33	3.13	3.58	3.13
Military and Civilian Service	1.25	0.45		
Marginal Employment	1.66	1.72	2.83	1.88
Self-Employed	5.95	6.57	5.40	6.85
Other	9.78	9.72	5.21	6.81
Employment Subject to Insurance Contributions	33.18	12.31	21.87	13.46
Invalidity	0.60	0.74	0.60	0.93
Pension	0.63	0.60	0.59	0.42
Missings	10.50	11.09	17.56	10.90
Mean Age	52.00	0.00	52.00	0.00
Sum	62.86	1.88	62.81	2.14
Number of Observations	13696		16668	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A27 Average Time Spent in Different Activities – East vs. West Germany, German Citizens

First-Time Old-Age Pensioners in 2004 (West Germany, Only German Citizens)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.11	2.70	2.19	2.18
Apprenticeship/Training	2.94	0.67	2.65	0.76
Non-Professional Caretaking	2.39	2.51	3.24	2.82
Childcare/Household	3.29	3.69	10.77	5.94
Invalidity and Sickness	0.62	0.77	0.50	0.61
Unemployed	3.06	3.33	2.96	2.95
Military and Civilian Service	1.18	0.48		
Marginal Employment	1.85	1.79	2.95	1.86
Self-Employed	7.61	7.11	8.30	8.39
Other	10.61	10.24	5.34	7.00
Employment Subject to Insurance Contributions	31.84	13.41	18.73	12.99
Invalidity	0.60	0.85	0.59	1.01
Pension	0.63	0.55	0.58	0.41
Missings	11.84	12.13	20.03	10.66
Mean Age	52.00	0.00	52.00	0.00
Sum	63.01	1.89	63.22	2.09
Number of Observations	10196		12996	

First-Time Old-Age Pensioners in 2004 (East Germany, Only German Citizens)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.63	2.45	2.48	2.10
Apprenticeship/Training	2.52	0.74	2.31	0.74
Non-Professional Caretaking	2.65	2.15	2.73	2.74
Childcare/Household	2.21	2.53	3.71	4.09
Invalidity and Sickness	0.61	0.65	0.72	0.71
Unemployed	3.86	2.62	4.79	3.14
Military and Civilian Service	1.44	0.30		
Marginal Employment	0.90	1.15	1.76	1.68
Self-Employed	2.87	3.87	2.97	3.83
Other	6.21	5.87	4.34	5.36
Employment Subject to Insurance Contributions	37.08	7.00	32.36	8.96
Invalidity	0.61	0.40	0.64	0.54
Pension	0.65	0.74	0.62	0.44
Missings	6.57	5.63	8.85	6.32
Mean Age	52.00	0.00	52.00	0.00
Sum	62.43	1.80	61.33	1.59
Number of Observations	3500		3672	

Source: FDZ-RV - SUFVVL2004, Author's calculations

Table A28 Average Time Spent in Different Activities - Total, Persons with Migration History

First-Time Old-Age Pensioners in 2004 (Total, Persons with Migration History)	Men Mean	Std. Dev.	Women Mean	Std. Dev.
School/University	3.40	2.68	2.58	2.12
Apprenticeship/Training	1.40	1.24	0.92	0.85
Non-Professional Caretaking	2.53	1.07	4.83	3.24
Childcare/Household	7.80	9.11	4.57	4.95
Invalidity and Sickness	0.63	0.62	0.67	0.61
Unemployed	4.21	3.86	4.16	3.78
Military and Civilian Service	2.44	0.98		
Marginal Employment	1.17	1.59	2.21	1.73
Self-Employed	12.40	8.30	8.25	.
Other	2.51	4.02	1.24	2.27
Employment Subject to Insurance Contributions	31.56	9.92	25.04	11.05
Invalidity	0.70	0.90	1.17	2.10
Pension	0.75	1.36	0.84	0.91
Missings	13.98	9.25	17.68	10.24
Mean Age	52.00	0.00	52.00	0.00
Sum	62.94	1.91	62.20	1.96
Number of Observations	578		532	

Source: FDZ-RV - SUFVVL2004, Author’s calculations

Table A29 Average Time Spent in Different Activities - East vs. West Germany, Persons with Migration History

First-Time Old-Age Pensioners in 2004 (West Germany, Persons with Migration History)	Men Mean	 Std. Dev.	Women Mean	 Std. Dev.
School/University	3.25	2.63	2.58	2.07
Apprenticeship/Training	1.39	1.24	0.88	0.87
Non-Professional Caretaking	2.42	1.20	4.63	3.23
Childcare/Household	1.17	0.90	4.71	5.05
Invalidity and Sickness	0.63	0.62	0.65	0.61
Unemployed	4.05	3.83	4.17	3.80
Military and Civilian Service	2.44	0.99		
Marginal Employment	1.21	1.63	2.26	1.78
Self-Employed	12.40	8.30	8.25	.
Other	2.53	4.08	1.23	2.32
Employment Subject to Insurance Contributions	32.02	9.64	25.00	11.02
Invalidity	0.70	0.93	1.14	2.11
Pension	0.75	1.40	0.79	0.81
Missings	13.73	9.18	17.72	10.10
Mean Age	52.00	0.00	52.00	0.00
Sum	62.96	1.90	62.20	1.96
Number of Observations	531		490	

First-Time Old-Age Pensioners in 2004 (East Germany, Persons with Migration History)	Men Mean	 Std. Dev.	Women Mean	 Std. Dev.
School/University	4.96	2.89	2.53	2.81
Apprenticeship/Training	1.48	1.23	1.11	0.74
Non-Professional Caretaking	3.00	.	8.00	1.06
Childcare/Household	17.75	0.82	2.95	3.16
Invalidity and Sickness	0.64	0.56	0.91	0.60
Unemployed	5.82	3.92	4.07	3.66
Military and Civilian Service	2.38	0.96		
Marginal Employment	0.42		1.52	1.00
Self-Employed				
Other	2.19	2.61	1.38	1.08
Employment Subject to Insurance Contributions	26.46	11.54	25.53	11.51
Invalidity	0.69	0.53	1.55	2.00
Pension	0.76	0.92	1.41	1.60
Missings	16.83	9.69	17.15	11.89
Mean Age	52.00	0.00	52.00	0.00
Sum	62.79	2.03	62.14	1.97
Number of Observations	47		42	

Source: FDZ-RV - SUFVVL2004; Author's calculations

Notes for the following Table:

Tables A30 compares the average time spent in different activities after streamlining the activities in SOEP and SUF VVL 2004 data. The average values only considers individuals who have a values > 0 in the respective category, individuals with a zero are left out of the calculations.

Table A30 Comparison of Average Time Spent in Different Activities after Streamlining, Total Men

Original Categories	New Categories	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	School/University	3.26	4314	4.92	192	-1.66	3.26	3.55	78	-0.29
Apprenticeship/Training	Apprenticeship/Training	2.82	10191	3.1	340	-0.28	2.82	3.22	122	-0.41
Home production	Home production	3.51	95	2.69	2	0.81	3.51	2	1	1.51
Unemployed	Unemployed	3.38	7287	3.98	210	-0.6	3.38	3.77	73	-0.4
Military/ Civil Service	Military/ Civil Service	1.28	4948	2.24	129	-0.96	1.28	2.18	61	-0.89
Other	Other Activities	12.43	2835	0.17	73	12.26	12.43	0.91	28	11.52
Care giving			107							
Invalidity/Sickness			6768							
Full-time	Employment	40.75	14237	39.83	392	0.91	40.75	39.96	140	0.79
Part-time					192				36	
Marginal			531							
Self-Employment			696							
Invalidity/disability	Retirement	1.24	7609	2.28	392	-1.04	1.24	1.42	140	-0.18
Old-Age			6642							
Years Missing	Years Missing	10.64	14274	1.03	392	9.61	10.64	3.37	140	7.26
Average Age	Average Age	62.86	14274	65.07	392	-2.2	62.86	63.59	140	-0.72

Notes: △ = average time spent in activities x_{VVL} – average time spent in activity x_{SOEP}. Source: SOEP 2005 and FDZ-RV - SUFVVL2004, Author's calculations

Table A31 Comparison of Average Time Spent in Different Activities after Streamlining, Total Women (Cont'd)

Original Categories	New Categories	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	School/University	2.3	3373	2.43	252	-0.13	2.3	2.17	99	0.12
Apprenticeship/Training	Apprenticeship/Training	2.49	5550	2.78	335	-0.3	2.49	2.93	139	-0.44
Home production	Home production	9.04	15255	18.32	309	-9.28	9.04	16.81	133	-7.76
Unemployed	Unemployed	3.6	7725	4.11	208	-0.5	3.6	4.35	97	-0.74
Military/ Civil Service	Military/ Civil Service	0	0	11.58	2	-11.58	0	3	1	-3
Other	Other Activities	8.79	3956	2.94	87	5.85	8.79	1.99	39	6.8
Care giving			1985							
Invalidity/Sickness			6342							
Full-time	Employment	30.21	16443	32.74	448	-2.53	30.21	31.95	176	-1.74
Part-time					306				120	
Marginal			2493							
Self-Employment			209							
Invalidity/disability	Retirement	1.23	7136	2.03	465	-0.8	1.23	1.13	180	0.09
Old-Age			9941							
Years Missing	Years Missing	17.57	17200	2.3	465	15.26	17.57	2.88	180	14.69
Average Age	Average Age	62.79	17200	65.05	465	-2.26	62.79	63.74	180	-0.95

Notes: △ = average time spent in activities x_{VVL} – average time spent in activity x_{SOEP}. Source: SOEP 2005 and FDZ-RV - SUFVVL2004, Author's calculations

Table A31 Comparison of Average Time Spent in Different Activities after Streamlining, Men West (Cont'd)

Original Categories	New Categories	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	School/University	3.12	3164	5.13	128	-2.01	3.12	3.43	48	-0.31
Apprenticeship/Training	Apprenticeship/Training	2.92	7559	3.22	225	-0.3	2.92	3.45	76	-0.53
Home production	Home production	3.22	91	2.69	2	0.53	3.22	2	1	1.22
Unemployed	Unemployed	3.13	4920	4.36	116	-1.23	3.13	3.67	38	-0.53
Military/ Civil Service	Military/ Civil Service	1.23	3730	2.27	68	-1.04	1.23	2	91	-0.77
Other	Other Activities	13.08	2322	1.07	47	12.01	13.08	1	17	12.08
Care giving			80							
Invalidity/Sickness			4161							
Full-time	Employment	41.33	10690	40.14	266	1.19	41.33	41.32	91	0.01
Part-time					73				22	
Marginal			427							
Self-Employment			453							
Invalidity/disability	Retirement	1.24	5483	2.3	266	-1.06	1.24	1.51	91	-0.28
Old-Age			5231							
Years Missing	Years Missing	11.94	10727	1.45	266	10.48	11.94	2.5	91	9.44
Average Age	Average Age	63	10727	65.25	266	-2.25	63	63.78	91	-0.78

Notes: △ = average time spent in activities x_{VVL} – average time spent in activity x_{SOEP}. Source: SOEP 2005 and FDZ-RV - SUFVVL2004, Author's calculations

Table A31 Comparison of Average Time Spent in Different Activities after Streamlining, Women West (Cont'd)

Original Categories	New Categories	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	School/University	2.21	2266	2.39	175	-0.18	2.21	2.24	67	-0.04
Apprenticeship/Training	Apprenticeship/Training	2.62	3311	2.84	217	-0.22	2.62	3.02	88	-0.4
Home production	Home production	10.55	11905	19.99	253	-9.44	10.55	18.57	102	-8.03
Unemployed	Unemployed	3.03	5191	3.54	104	-0.51	3.03	3.31	50	-0.28
Military/ Civil Service	Military/ Civil Service	0	0	11.58	2	-11.58	0	3	1	-3
Other	Other Activities	8.87	2570	3.4	59	5.47	8.87	2.19	26	6.68
Care giving			1711							
Invalidity/Sickness			3362							
Full-time	Employment	30.21	12732	31.47	311	-1.26	30.21	31.61	120	-1.4
Part-time										
Marginal			2233							
Self-Employment			96							
Invalidity/disability	Retirement	1.21	5672	1.87	328	-0.66	1.21	1.11	124	0.11
Old-Age			7706							
Years Missing	Years Missing	19.94	13486	2.13	328	17.81	19.94	2.51	124	17.43
Average Age	Average Age	63.19	13486	65.47	328	-2.28	63.19	64.12	124	-0.93

Notes: △ = average time spent in activities x_{VVL} – average time spent in activity x_{SOEP}. Source: SOEP 2005 and FDZ-RV - SUFVVL2004, Author's calculations

Table A31 Comparison of Average Time Spent in Different Activities after Streamlining, Men East (Cont'd)

Original Categories	New Categories	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	School/University	3.65	1150	4.2	64	-0.55	3.65	3.81	30	-0.16
Apprenticeship/Training	Apprenticeship/Training	2.51	2632	2.73	115	-0.21	2.51	2.67	46	-0.15
Home production	Home production	9.98	4	0	0	9.98	9.98	0	0	9.98
Unemployed	Unemployed	3.89	2367	3.2	94	0.69	3.89	3.95	35	-0.07
Military/ Civil Service	Military/ Civil Service	1.44	1218	2.19	61	-0.75	1.44	2.55	25	-1.11
Other	Other Activities	9.42	518	0.92	26	8.5	9.42	0.72	11	8.7
Care giving			27							
Invalidity/Sickness			2607							
Full-time	Employment	40.71	3547	38.92	126	1.79	40.71	36.21	49	4.5
Part-time					39				14	
Marginal			104							
Self-Employment			243							
Invalidity/disability	Retirement	1.26	2126	2.23	126	-0.97	1.26	1.16	49	0.1
Old-Age			1411							
Years Missing	Years Missing	6.71	3547	2.87	126	3.84	6.71	5.67	49	1.04
Average Age	Average Age	62.44	3547	64.48	126	-2.05	62.44	63.08	49	-0.64

Notes: △ = average time spent in activities x_{VVL} – average time spent in activity x_{SOEP}. Source: SOEP 2005 and FDZ-RV - SUFVVL2004, Author's calculations

Table A31 Comparison of Average Time Spent in Different Activities after Streamlining, Women East (Cont'd)

Original Categories	New Categories	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	School/University	2.48	1107	2.63	77	-0.15	2.48	1.83	32	0.65
Apprenticeship/Training	Apprenticeship/Training	2.3	2239	2.62	118	-0.31	2.3	2.62	51	-0.32
Home production	Home production	3.7	3350	6.33	56	-2.63	3.7	7.14	31	-3.44
Unemployed	Unemployed	4.78	2534	4.92	104	-0.14	4.78	6.24	47	-1.45
Military/ Civil Service	Military/ Civil Service	0	0	0	0	0	0	0	0	0
Other	Other Activities	7.75	386	1.63	28	6.12	7.75	1.14	13	6.61
Care giving			274							
Invalidity/Sickness			2980							
Full-time	Employment	37	3711	37.27	137	-0.27	37	33.58	56	3.42
Part-time					97				41	
Marginal			260							
Self-Employment			113							
Invalidity/disability	Retirement	1.28	1464	2.6	137	-1.31	1.28	1.25	56	0.04
Old-Age			2235							
Years Missing	Years Missing	8.95	3714	2.94	137	6.01	8.95	4.37	56	4.57
Average Age	Average Age	61.34	3714	63.45	137	-2.11	61.34	62.18	56	-0.84

Notes: △ = average time spent in activities x_{VVL} – average time spent in activity x_{SOEP}. Source: SOEP 2005 and FDZ-RV - SUFVVL2004, Author's calculations

Table A31 Regression Results without Modifications for Various SOEP Samples, No additional Controls

Dependent Variable: Monthly Public Pension Benefit (topcode)	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Years in School (topcode)	38.437 (6.26)**	32.845 (4.15)**	37.041 (4.17)**	45.763 (5.02)**	18.641 (2.24)*	48.698 (3.95)**	33.509 (2.93)**	9.446 (0.92)	47.813 (3.05)**
Years in Training (topcode)	28.756 (7.75)**	25.037 (2.61)**	34.272 (2.90)**	19.932 (1.68)	30.243 (3.21)**	26.341 (1.67)	9.723 (0.57)	25.332 (2.23)*	62.655 (3.37)**
Years in Employment	4.350 (2.89)	1.106 (0.31)	10.392 (2.14)*	5.154 (1.05)	-4.118 (1.22)	5.047 (0.76)	12.103 (1.99)*	-4.421 (1.16)	13.773 (1.31)
Years in Unemployment	-25.665 (4.82)**	-34.147 (4.82)**	-13.203 (1.77)	-41.671 (4.35)**	-24.019 (4.13)**	-42.005 (3.39)**	-28.898 (2.17)*	-22.218 (2.77)**	0.640 (0.05)
Years in Home production	-13.412 (4.75)**	-14.661 (4.30)**	-8.769 (1.69)		-18.402 (5.85)**			-18.325 (5.10)**	-4.697 (0.47)
Retired (dummy)	-2.791 (0.04)	-100.616 (1.11)	133.231 (1.47)	-219.778 (1.65)	32.109 (0.44)	-329.964 (1.89)	38.557 (0.23)	-27.867 (0.30)	101.332 (0.94)
Other (dummy)	-186.666 (2.49)*	-196.845 (2.33)*	-31.622 (0.14)	-424.818 (2.22)*	-175.032 (2.44)*	-380.221 (1.73)	dropped	-167.303 (2.11)*	62.353 (0.27)
Missing (dummy)	-56.724 (1.70)	-47.902 (1.00)	-87.217 (2.32)*	-100.974 (1.87)	-60.537 (1.57)	-113.654 (1.45)	-40.067 (0.67)	4.337 (0.08)	-109.666 (2.25)*
Receives Civil Servant Pension (dummy)	-671.797 (6.35)**	-782.488 (6.13)**	-230.859 (1.34)	-721.529 (6.26)**	dropped	-833.168 (5.82)**	-212.360 (1.17)	dropped	dropped
Receives Private Pension (dummy)	-110.076 (1.71)	-135.850 (1.80)	-52.813 (0.44)	-210.942 (2.42)*	79.264 (0.85)	-224.659 (2.17)*	-180.922 (0.99)	49.441 (0.44)	38.052 (0.24)
Worked as Civil Servant or Self-employed (dummy)	-258.515 (6.63)**	-300.494 (6.23)**	-118.463 (2.05)*	-335.448 (6.43)**	-159.982 (2.75)**	-386.817 (5.82)**	-184.381 (2.49)*	-181.444 (2.52)*	13.858 (0.15)

Table A31 Regression Results without Modifications for Various SOEP Samples, No additional Controls (Cont'd)

Dependent Variable: Monthly Public Pension Benefit (topcode)	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Eduational Attainment: low Reference category: medium	-41.808 (1.27)	-37.169 (0.92)	-53.743 (1.06)	-108.931 (1.71)	-8.766 (0.26)	-106.680 (1.30)	-68.358 (0.80)	-20.394 (0.50)	-30.535 (0.49)
Educational Attainment: high Reference category: medium	104.383 (2.19)*	147.519 (2.39)*	86.075 (1.37)	92.803 (1.52)	34.846 (0.45)	152.602 (1.87)	24.105 (0.32)	-25.940 (0.26)	188.165 (1.55)
Educational Attainment: missing Reference category: medium	36.228 (0.51)	-133.669 (1.19)	196.652 (2.69)**	-124.008 (1.15)	199.664 (2.33)*	-265.026 (1.50)	34.184 (0.34)	-13.828 (0.10)	397.929 (3.63)**
Educational Attainment: unknown Reference category: medium	-106.643 (2.18)*	-89.247 (1.64)	-133.362 (0.58)	-220.776 (2.70)**	3.815 (0.07)	-216.211 (2.35)*	dropped	-5.116 (0.08)	-61.265 (0.27)
Sex – Female=1 (Dummy)	-349.582 (13.69)**	-458.074 (13.09)**	-187.980 (6.23)**						
West – West=1 (dummy)	100.678 (3.96)**			222.938 (5.94)**	-30.647 (0.96)				
Constant	895.672 (6.43)**	1203.751 (7.00)**	523.258 (2.37)*	857.542 (3.67)**	939.924 (5.89)**	1067.965 (3.29)**	578.985 (2.08)*	941.145 (5.27)**	56.850 (0.12)
Observations	949	662	287	443	506	304	139	358	148
R-Squared	0.59	0.62	0.55	0.40	0.44	0.40	0.33	0.40	0.55

Notes: Absolute value of t-statistic in parenthesis. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. Source: SOEP 2005; Author's calculations

Table A32 Regression Results without Modifications for Various SUF VVL 2004 Samples, no Additional Controls

Dependent Variable Monthly Public Pension Benefit	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Years in School (topcode)	39.602 (48.71)***	31.365 (31.37)***	55.205 (42.44)***	45.465 (42.74)***	38.244 (32.93)***	39.619 (30.66)***	54.717 (30.10)***	59.172 (31.33)***	29.904 (20.73)***
Years in Training (topcode)	24.709 (26.13)***	23.518 (21.17)***	20.141 (11.78)***	41.319 (29.90)***	5.209 (4.43)***	43.507 (27.34)***	30.397 (11.27)***	11.110 (5.26)***	4.720 (3.41)***
Years in Employment	26.615 (186.00)***	27.049 (169.43)***	22.734 (61.76)***	31.711 (159.25)***	20.750 (112.52)***	32.207 (145.93)***	24.977 (43.23)***	20.958 (45.59)***	20.856 (100.84)***
Years in Unemployment	0.389 (0.85)	-0.323 (0.57)	-1.415 (1.91)*	1.939 (2.79)***	1.533 (2.79)***	1.777 (2.17)**	-3.920 (3.01)***	0.647 (0.76)	3.180 (4.51)***
Years in Home production	8.999 (29.28)***	10.573 (31.14)***	4.263 (4.92)***		1.933 (6.19)***			1.850 (2.06)**	2.071 (6.14)***
Retired (Dummy)	229.809 (11.99)***	231.493 (9.92)***	227.662 (7.65)***	268.592 (9.88)***	178.167 (7.40)***	246.880 (7.56)***	322.420 (7.12)***	132.841 (3.50)***	198.437 (6.70)***
Other (Dummy)	96.500 (25.46)***	101.958 (23.82)***	47.791 (6.10)***	154.445 (26.39)***	64.185 (14.57)***	171.341 (25.65)***	50.984 (4.28)***	44.513 (4.38)***	67.527 (13.78)***
Missing (Dummy)	-125.583 (23.94)***	-153.178 (23.78)***	-77.427 (9.59)***	-61.586 (10.71)***	-180.386 (10.51)***	-76.170 (10.91)***	-48.799 (5.13)***	-104.194 (4.55)***	-232.392 (9.99)***
Sex – Female = 1 (Dummy)	-197.208 (59.37)***	-231.164 (55.77)***	-132.136 (26.41)***						
West – West =1 (Dummy)	166.819 (53.67)***			233.646 (52.75)***	66.463 (16.87)***				

Table A33 Regression Results without Modifications for Various SUF VVL 2004 Samples, no Additional Controls (Cont'd)

Dependent Variable Monthly Public Pension Benefit	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Educational Attainment: low (Reference Category: medium)	-114.930 (21.73)***	-128.184 (21.19)***	-52.728 (5.11)***	-151.069 (16.50)***	-69.389 (12.12)***	-163.455 (15.93)***	-79.718 (3.97)***	-39.853 (3.52)***	-73.893 (11.19)***
Educational Attainment: high (Reference Category: medium)	188.765 (28.14)***	200.363 (21.84)***	178.025 (20.61)***	146.457 (17.40)***	191.018 (17.90)***	150.438 (13.86)***	163.576 (13.27)***	180.243 (14.78)***	142.427 (8.04)***
Educational Attainment: missing (Reference Category: medium)	-60.144 (17.52)***	-52.591 (12.81)***	-64.516 (11.40)***	-105.538 (20.69)***	-37.767 (9.08)***	-106.693 (17.54)***	-81.710 (9.29)***	-56.231 (7.86)***	-31.230 (6.29)***
Educational Attainment: unknown (Reference Category: medium)	-81.834 (18.99)***	-85.979 (16.63)***	-57.922 (8.29)***	-90.819 (13.71)***	-70.801 (14.06)***	-97.525 (12.25)***	-57.996 (5.22)***	-64.143 (7.46)***	-72.439 (12.04)***
Constant	73.134 (7.69)***	269.830 (26.65)***	138.025 (6.97)***	-225.425 (17.72)***	193.267 (9.91)***	5.316 (0.40)	28.677 (0.95)	98.061 (3.07)***	306.550 (12.27)***
Observations	30965	23749	7216	14036	16929	10504	3532	3684	13245
R-Square	0.78	0.80	0.70	0.74	0.70	0.76	0.61	0.67	0.67

Notes: Absolute value of t-statistic in parenthesis. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Source: FDZ-RV - SUFVVL2004; Author's calculations

Table A33 Regression Results after Modifications for Various SOEP Samples, no Additional Controls

Dependent Variable: Monthly Public Pension benefit	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Years in School (Topcode)	35.346 (6.07)**	38.135 (5.03)**	24.194 (3.18)**	37.354 (4.00)**	20.407 (2.64)**	44.559 (3.48)**	13.578 (1.27)	19.370 (1.98)*	33.343 (2.88)**
Years in Training (Topcode)	31.312 (4.17)**	28.653 (3.06)**	32.149 (2.94)**	16.527 (1.38)	42.474 (4.61)**	24.836 (1.56)	2.891 (0.18)	36.861 (3.22)**	59.432 (4.24)**
Years in Employment	13.353 (7.99)**	12.625 (6.17)**	14.826 (5.43)**	3.468 (0.71)	12.060 (7.59)**	4.200 (0.63)	11.840 (2.04)*	12.220 (6.48)**	13.967 (4.90)**
Years in Unemployment	-12.894 (2.95)**	-18.604 (3.10)**	-5.629 (1.05)	-46.251 (4.90)**	-3.895 (0.81)	-47.075 (3.83)**	-27.002 (2.12)*	-0.315 (0.04)	2.910 (0.50)
Years in Homeproduction	-10.200 (2.70)**	-7.720 (1.71)	8.880 (0.53)		-6.590 (1.89)			-9.618 (2.45)*	
Retired (Dummy) if years in retirement > 4	58.941 (0.85)	8.632 (0.10)	170.575 (1.99)*	-232.418 (1.75)	142.457 (1.98)*	-340.300 (1.93)	56.282 (0.35)	114.260 (1.23)	174.965 (1.79)
Other (Dummy) if years in other activities > 3	-94.753 (1.28)	-84.926 (1.01)	-8.716 (0.04)	-431.427 (2.25)*	-29.606 (0.43)	-376.950 (1.69)	0.000 (0.00)**	-23.229 (0.30)	72.744 (0.33)
Missing (Dummy) if years missing > 3	-20.311 (0.65)	-12.407 (0.28)	-74.798 (2.13)*	-94.195 (1.74)	-11.565 (0.34)	-105.246 (1.33)	-42.620 (0.76)	41.066 (0.89)	-81.432 (1.90)
Worked as Civil Servant or Self-Employed (Dummy)	-302.669 (7.66)**	-351.942 (7.07)**	-141.289 (2.55)*	-341.986 (6.56)**	-151.734 (2.57)*	-392.859 (5.88)**	-212.287 (3.04)**	-179.715 (2.43)*	-0.089 (0.00)

Table A34 Regression Results after Modifications for Various SOEP Samples, no Additional Controls (Cont'd)

Dependent Variable: Monthly Public Pension benefit	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Educational Attainment: Low Reference Category: Intermediate	-54.301 (1.64)	-55.801 (1.36)	-46.355 (0.94)	-98.479 (1.55)	-32.320 (0.94)	-96.917 (1.18)	-48.785 (0.59)	-50.987 (1.23)	-17.064 (0.29)
Educational Attainment: High Reference Category: Intermediate	162.520 (4.27)**	147.190 (2.71)**	212.110 (5.05)**	139.762 (2.70)**	219.483 (4.17)**	155.443 (2.12)*	168.495 (3.02)**	92.197 (1.16)	295.065 (4.78)**
Educational Attainment: Missing Reference Category: Intermediate	-94.716 (0.69)	-123.466 (0.75)	138.267 (0.60)	-231.194 (1.31)	0.536 (0.00)	-231.010 (1.01)	60.772 (0.26)	-4.455 (0.02)	0.000 (0.00)**
Educational Attainment: Unknown Reference Category: Intermediate	-130.087 (1.74)	-124.274 (1.49)		-189.058 (1.54)	-42.205 (0.50)	-173.114 (1.25)	0.000 (0.00)**	-62.835 (0.68)	0.000 (0.00)**
Gender (Female = 1)	-331.502 (12.35)**	-431.668 (11.27)**	-197.502 (6.67)**						
West (West = 1)	119.825 (4.48)**			229.683 (6.11)**	12.727 (0.37)				
Receives Civil Servant Pension (Dummy)				-705.688 (6.14)**		-819.666 (5.68)**			
Receives Private Pension (Dummy)				-203.679 (2.33)*		-205.209 (1.97)*			
Constant	458.298 (5.41)**	659.207 (6.36)**	312.191 (2.49)*	921.928 (3.94)**	171.991 (2.14)*	1094.230 (3.34)**	586.616 (2.22)*	197.825 (2.15)*	20.804 (0.16)
Observations	949	662	287	443	506	304	139	358	148
R-squared	0.58	0.59	0.57	0.40	0.42	0.39	0.36	0.36	0.56

Notes: Absolute value of t-statistic in parenthesis. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Source: SOEP 2005; Author's calculations

Table A34 Regression Results After Modifications for Various SOEP Samples (Excluding Civil Servants and Self-Employed), no Additional Controls

Dependent Variable: VP10301 Monthly Public Pension Benefit	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Years in School (Topcode)	44.250 (7.48)**	42.099 (5.57)**	39.580 (4.86)**	41.944 (4.39)**	34.030 (4.23)**	43.860 (3.36)**	30.186 (2.72)**	25.002 (2.48)*	48.568 (3.79)**
Years in Training (Topcode)	34.034 (4.49)**	30.739 (3.27)**	38.997 (3.45)**	15.670 (1.26)	45.912 (4.97)**	15.522 (0.94)	12.378 (0.74)	39.987 (3.56)**	62.686 (4.14)**
Years in Employment	13.496 (8.02)**	12.691 (6.19)**	16.250 (5.80)**	7.825 (1.54)	11.972 (7.48)**	6.300 (0.90)	13.886 (2.33)*	12.116 (6.42)**	16.589 (5.70)**
Years in Unemployment	-13.684 (3.10)**	-17.480 (2.86)**	-7.263 (1.32)	-37.122 (3.71)**	-6.481 (1.34)	-39.569 (3.02)**	-26.005 (1.97)	-1.127 (0.16)	0.786 (0.13)
Years in Homeproduction	-10.580 (2.79)**	-8.015 (1.78)	4.110 (0.24)		-7.521 (2.16)*			-10.242 (2.62)**	
Retired (Dummy)	70.113 (1.00)	15.140 (0.17)	150.334 (1.67)	-173.955 (1.25)	136.472 (1.86)	-275.144 (1.49)	45.846 (0.27)	115.205 (1.23)	107.443 (1.01)
Other (Dummy)	-109.370 (1.48)	-98.603 (1.18)	-6.405 (0.03)	-429.380 (2.14)*	-39.679 (0.57)	-412.433 (1.76)	DROPPED	-25.666 (0.34)	62.339 (0.28)
Missing (Dummy)	-27.863 (0.89)	-15.039 (0.34)	-79.148 (2.18)*	-87.926 (1.55)	-19.078 (0.56)	-97.922 (1.18)	-44.531 (0.76)	41.842 (0.90)	-84.678 (1.93)
Worked as Civil Servant or Self-Employed (Dummy)	-296.824 (7.46)**	-344.948 (6.92)**	-136.585 (2.38)*	-384.271 (7.09)**	-154.646 (2.59)**	-438.296 (6.26)**	-206.449 (2.85)**	-179.291 (2.41)*	9.533 (0.10)

Table A35 Regression Results After Modifications for Various SOEP Samples (Excluding Civil Servants and Self-Employed), no Additional Controls (Cont'd)

Dependent Variable: VP10301 Monthly Public Pension Benefit	Total	Total West	Total East	Total Men	Total Women	Men West	Men East	Women West	Women East
Educational Attainment: Low Reference Category: Intermediate	-50.431 (1.49)	-52.679 (1.26)	-55.983 (1.10)	-113.741 (1.70)	-20.059 (0.57)	-117.760 (1.35)	-72.646 (0.85)	-40.652 (0.96)	-28.647 (0.46)
Educational Attainment: High Reference Category: Intermediate	87.280 (1.79)	128.102 (2.00)*	82.280 (1.31)	67.772 (1.06)	67.578 (0.84)	118.590 (1.37)	29.533 (0.40)	17.606 (0.17)	206.773 (1.73)
Educational Attainment: Missing Reference Category: Intermediate	64.348 (0.88)	-99.095 (0.84)	211.290 (2.91)**	-94.741 (0.84)	232.144 (2.63)**	-227.085 (1.21)	53.241 (0.53)	16.141 (0.12)	401.257 (3.90)**
Educational Attainment: Unknown Reference Category: Intermediate	-93.645 (1.86)	-81.985 (1.45)	-125.925 (0.54)	-221.699 (2.59)**	23.266 (0.41)	-220.096 (2.25)*		10.525 (0.17)	-67.248 (0.30)
Sex – Female = 1 (Dummy)	-332.450 (12.28)**	-432.380 (11.29)**	-193.994 (6.27)**						
West – West = 1	106.753 (3.97)**			212.988 (5.43)**	-11.133 (0.32)				
Constant	465.900 (5.46)**	657.393 (6.32)**	263.034 (2.04)*	754.023 (3.10)**	188.411 (2.34)*	1032.331 (3.04)**	499.318 (1.84)	190.462 (2.07)*	-56.540 (0.42)
Observations	949	662	287	443	506	304	139	358	148
R-squared	0.57	0.59	0.54	0.34	0.41	0.32	0.32	0.36	0.55

Notes: Absolute value of t-statistic in parenthesis. Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Source: SOEP 2005; Author's calculations

Table A35 Regression Results After Modifications for Various SUF VVL 2004 Samples, no Additional Controls

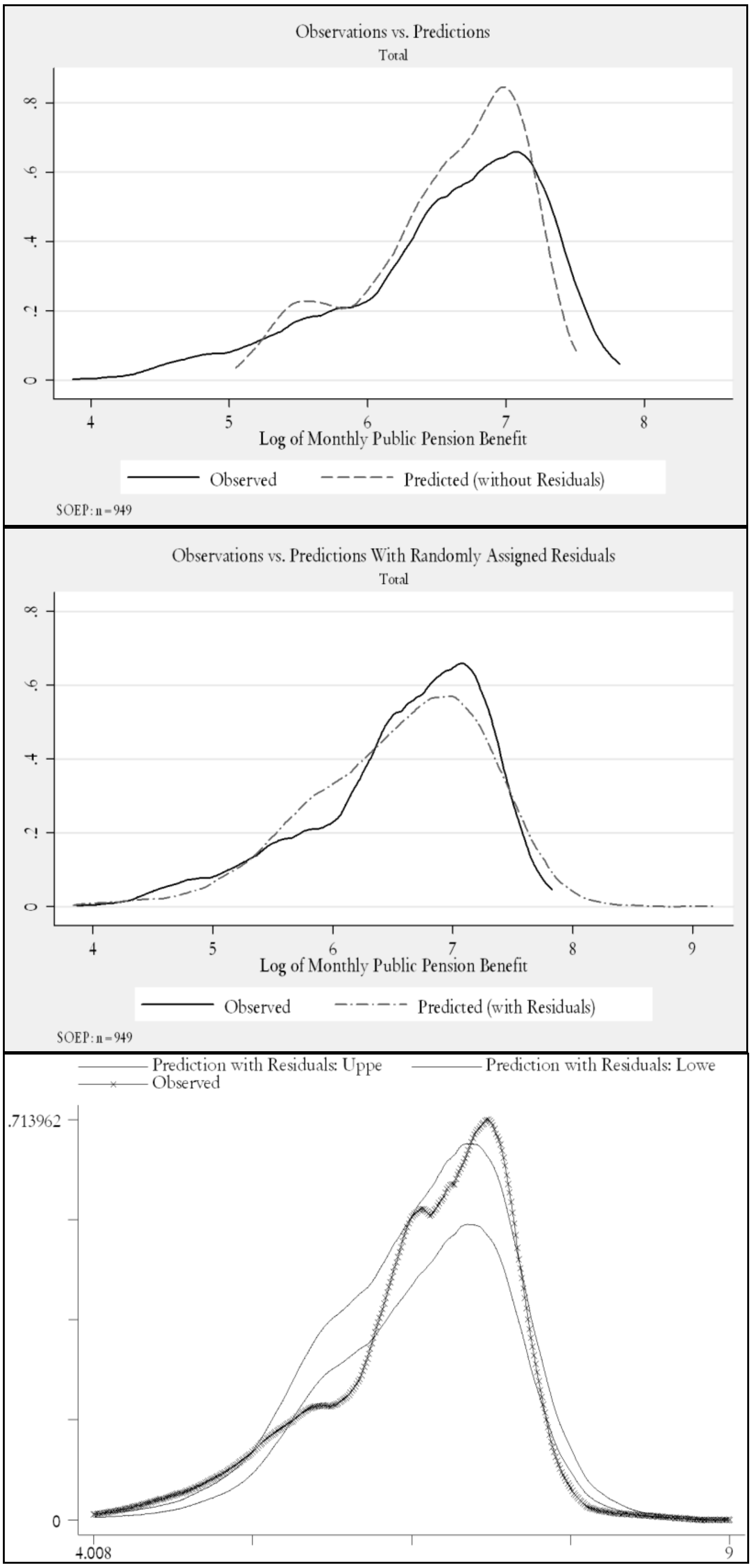
Dependent Variable: Monthly Public Pension Benefit (Log)	Total	Total West	Total East	Total Men	Total Women	West Men	East Men	West Women	East Women
Years in School (Topcode)	36.738 (42.39)**	27.999 (26.01)**	53.514 (39.92)**	43.477 (37.20)**	35.416 (29.53)**	37.347 (25.92)**	53.417 (28.32)**	26.772 (17.94)**	56.823 (29.40)**
Years in Training (Topcode)	23.907 (23.70)**	23.457 (19.59)**	18.805 (10.66)**	41.755 (27.46)**	4.547 (3.75)**	44.205 (24.85)**	28.721 (10.29)**	4.359 (3.05)**	10.483 (4.83)**
Years in Employment	25.214 (165.44)**	25.247 (147.92)**	22.627 (58.50)**	30.046 (138.19)**	20.102 (105.08)**	30.014 (123.99)**	24.910 (40.84)**	20.047 (93.68)**	21.084 (43.69)**
Years in Unemployment	6.579 (13.54)**	6.934 (11.42)**	2.856 (3.78)**	11.770 (15.67)**	4.473 (7.94)**	12.964 (14.37)**	2.821 (2.13)*	6.493 (8.98)**	3.304 (3.79)**
Years in Homeproduction	9.703 (29.14)**	11.070 (29.83)**	3.748 (4.18)**		2.514 (7.65)**			2.705 (7.60)**	1.432 (1.55)
Retired (Dummy)	147.405 (7.20)**	134.762 (5.37)**	190.753 (6.15)**	209.226 (7.02)**	81.895 (3.28)**	170.889 (4.70)**	307.068 (6.56)**	92.148 (3.02)**	70.571 (1.77)
Other (Dummy)	-37.382 (4.75)**	-33.788 (3.86)**	-73.801 (4.01)**	-62.771 (1.57)	-50.971 (7.58)**	-82.847 (1.68)	-10.038 (0.16)	-49.106 (6.67)**	-83.670 (4.76)**
Missing (Dummy)	-120.168 (21.41)**	-147.032 (21.11)**	-72.427 (8.67)**	-58.608 (9.24)**	-110.073 (6.18)**	-73.517 (9.40)**	-47.086 (4.76)**	-124.925 (5.19)**	-79.062 (3.34)**

Table A36 Regression Results After Modifications for Various SUF VVL 2004 Samples, no Additional Controls (Cont'd)

Dependent Variable: Monthly Public Pension Benefit (Log)	Total	Total West	Total East	Total Men	Total Women	West Men	East Men	West Women	East Women
Educational Attainment: Low Reference Category: Intermediate	-128.253 (22.73)**	-141.582 (21.73)**	-68.926 (6.31)**	-159.917 (15.91)**	-77.006 (13.02)**	-172.511 (15.09)**	-100.136 (4.76)**	-80.590 (11.82)**	-51.233 (4.32)**
Educational Attainment: High Reference Category: Intermediate	201.585 (28.08)**	222.587 (22.49)**	183.626 (20.51)**	152.399 (16.47)**	195.338 (17.53)**	162.829 (13.44)**	171.279 (13.39)**	149.346 (8.09)**	181.670 (14.34)**
Educational Attainment: Missing Reference Category: Intermediate	-129.566 (36.99)**	-134.930 (32.03)**	-100.023 (17.75)**	-208.322 (39.40)**	-72.479 (17.40)**	-228.653 (36.03)**	-128.653 (14.64)**	-70.015 (14.09)**	-79.887 (11.17)**
Educational Attainment: Unknown Reference Category: Intermediate	-108.725 (23.75)**	-120.558 (21.75)**	-67.889 (9.44)**	-120.235 (16.56)**	-87.901 (17.00)**	-132.011 (14.90)**	-67.158 (5.85)**	-93.385 (15.12)**	-73.947 (8.42)**
Gender (Female = 1)	-198.275 (55.85)**	-229.388 (51.26)**	-129.814 (25.11)**						
West (West = 1)	144.817 (43.91)**			200.766 (41.57)**	53.087 (13.11)**				
Constant	118.616 (11.59)**	307.647 (27.82)**	126.633 (6.06)**	-157.161 (11.23)**	141.170 (6.88)**	67.117 (4.51)**	15.000 (0.47)	209.057 (7.97)**	62.570 (1.86)
Number of Observations	30829	23656	7173	13983	16846	10463	3520	13193	3653
R-Squared	0.78	0.80	0.70	0.74	0.70	0.76	0.61	0.67	0.67

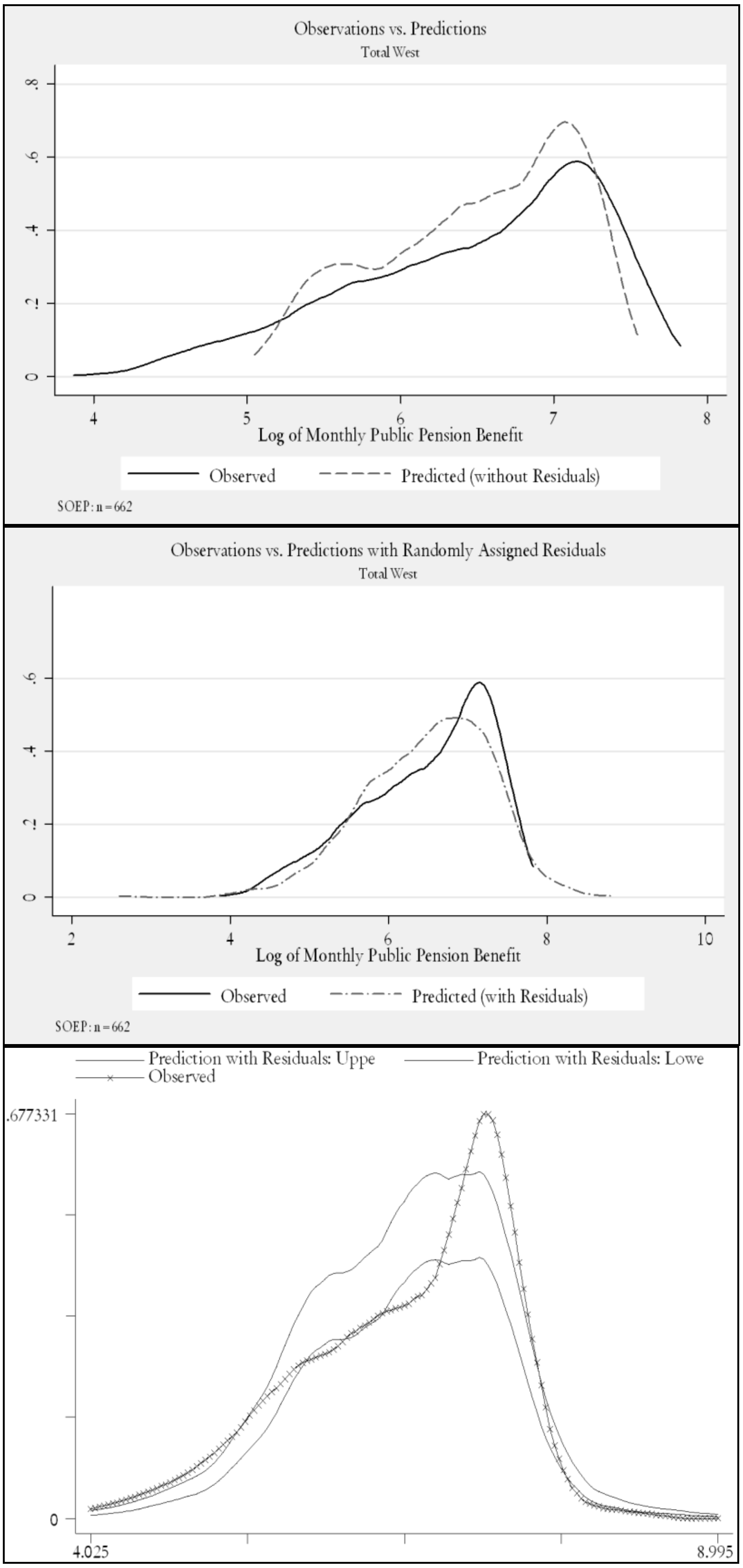
Notes: Absolute value of t-statistic in parenthesis. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Source: FDZ-RV - SUFVVL2004; Author's calculations

Figure A2 Observations, In-Sample Predictions, Total



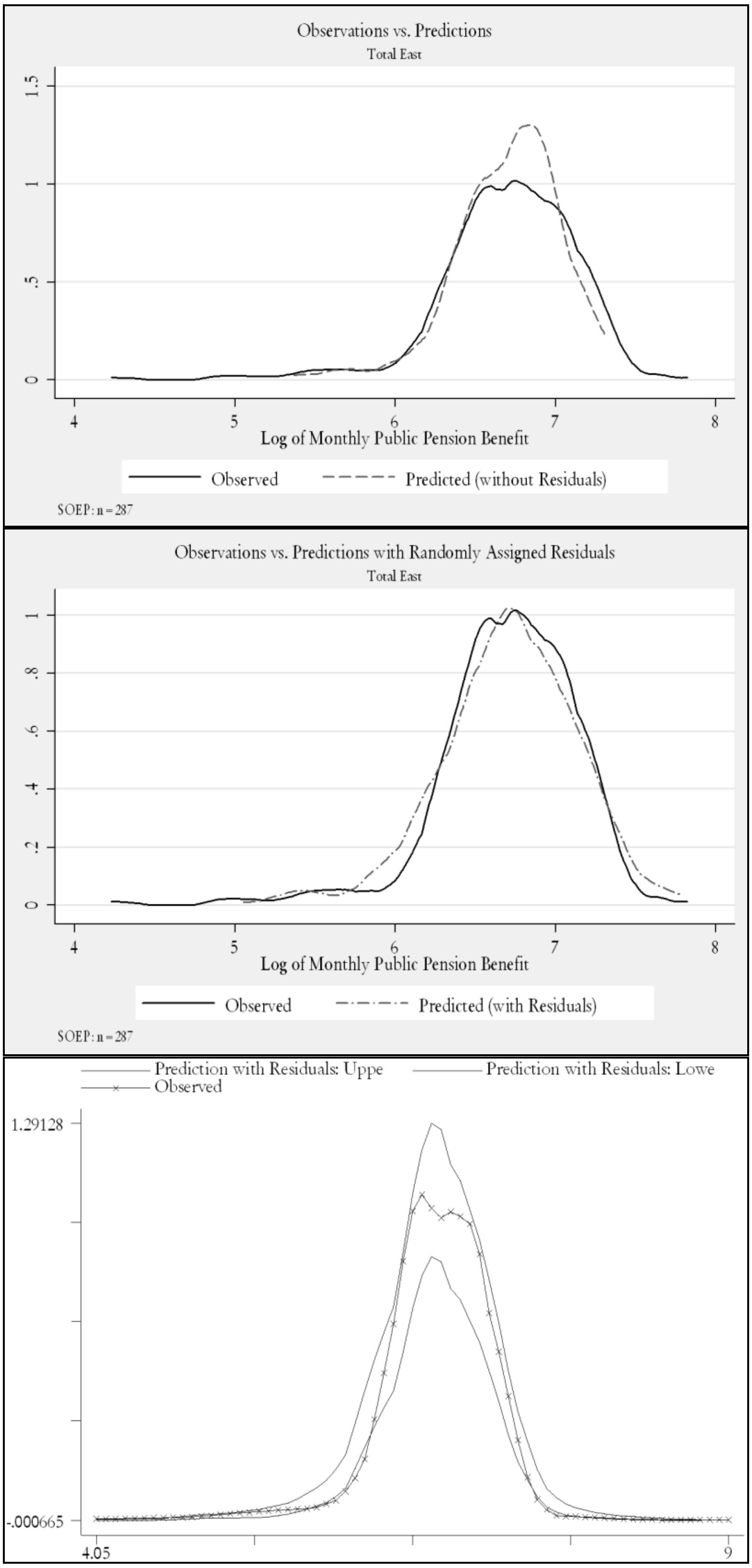
Source: SOEP 2005; Author's calculations

Figure A3 Observations and In-Sample Predictions, Total West Germany



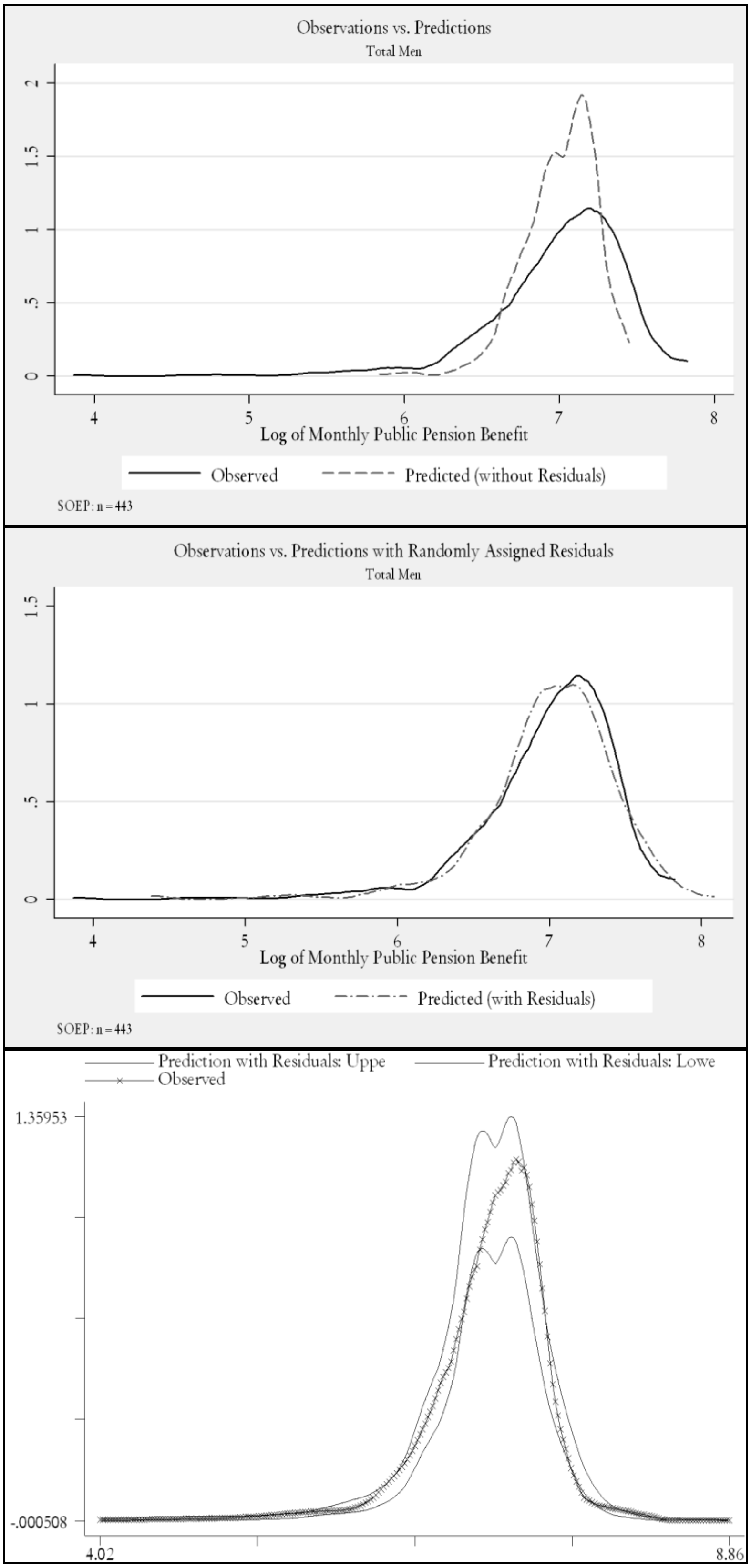
Source: SOEP 2005; Author's calculations

Figure A4 Observations and In-Sample Predictions, Total East



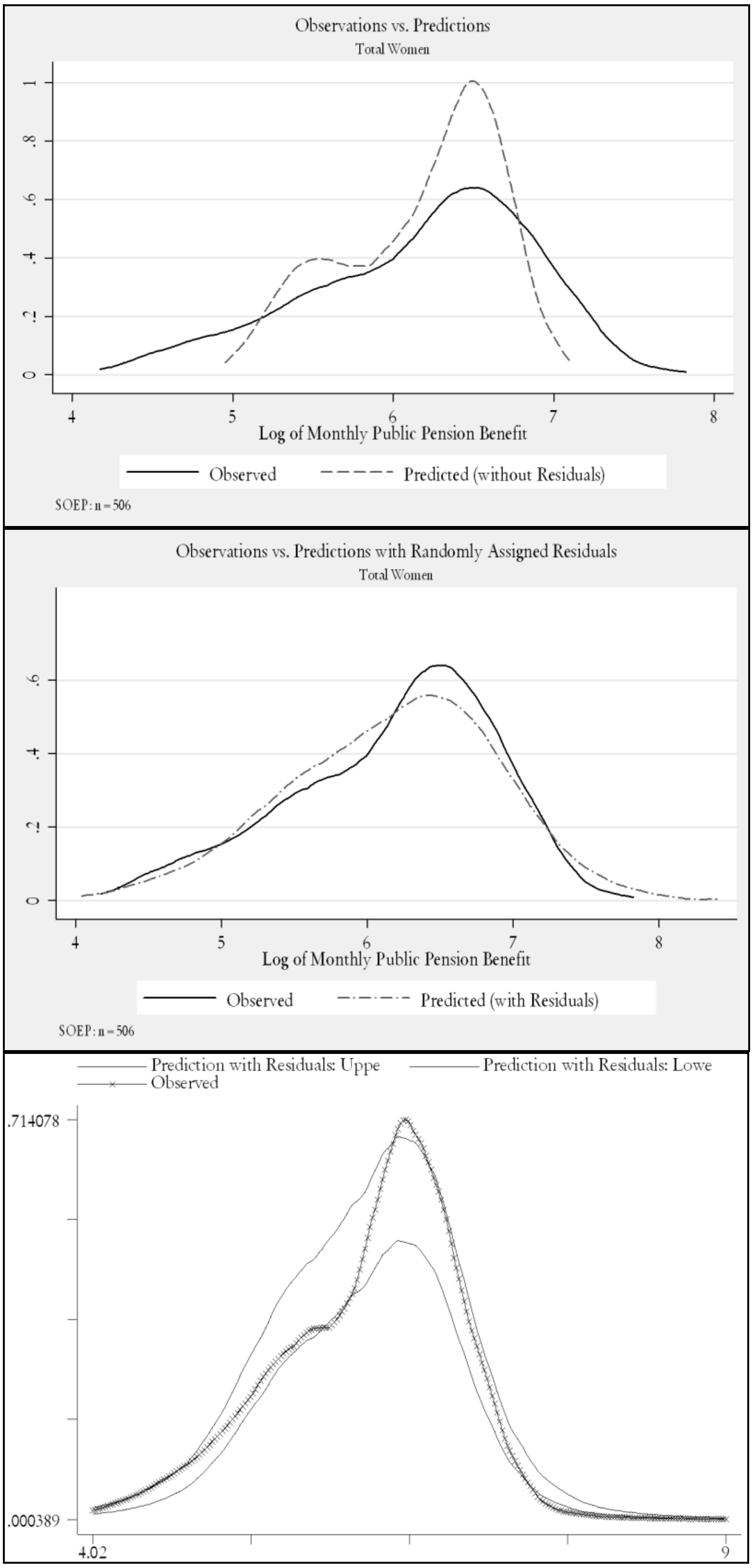
Source: SOEP 2005; Author's calculations

Figure A5 Observations and In-Sample Predictions, Total Men



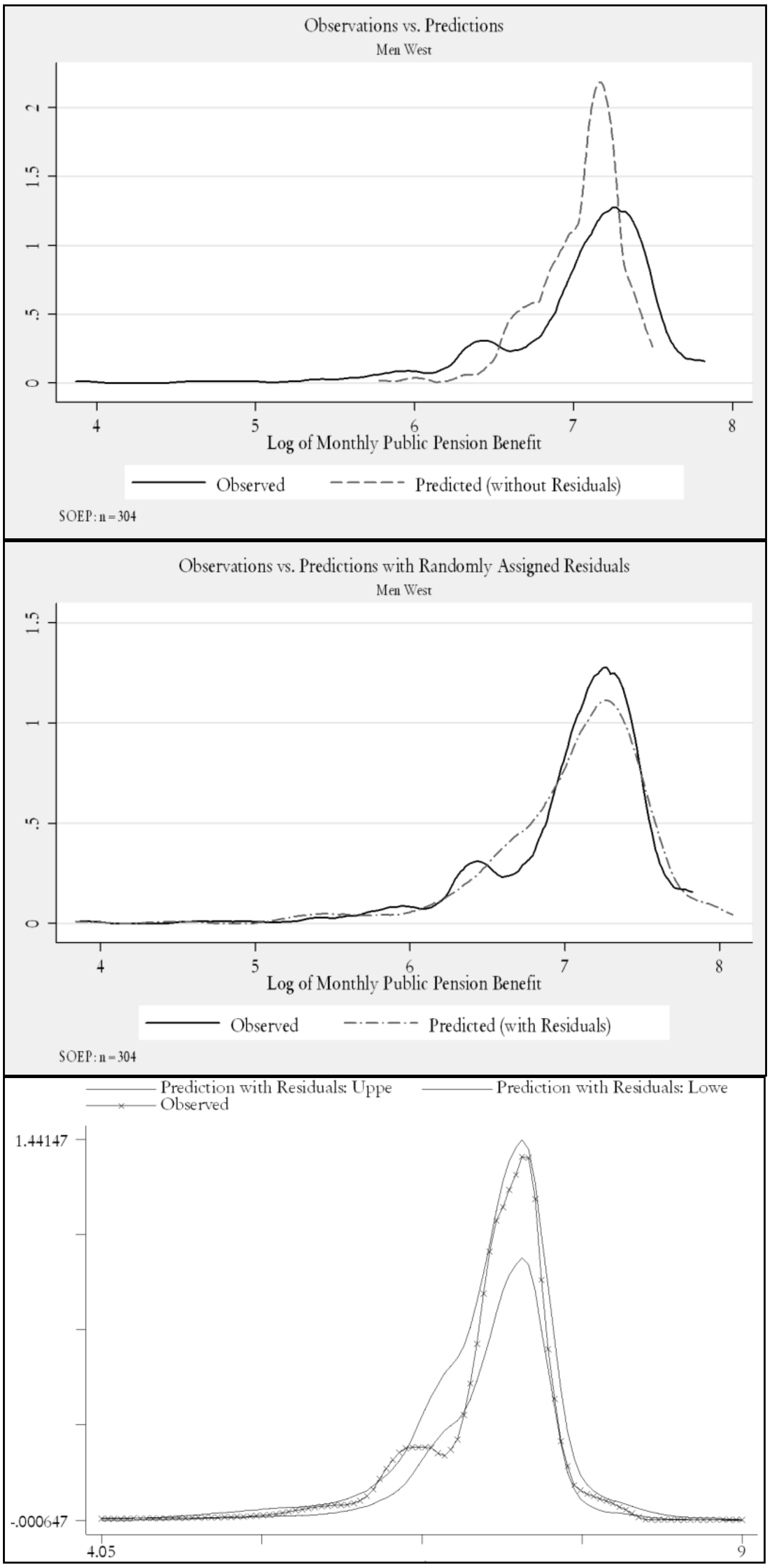
Source: SOEP 2005; Author's calculations

Figure A6 Observations and In-Sample Predictions, Total Women



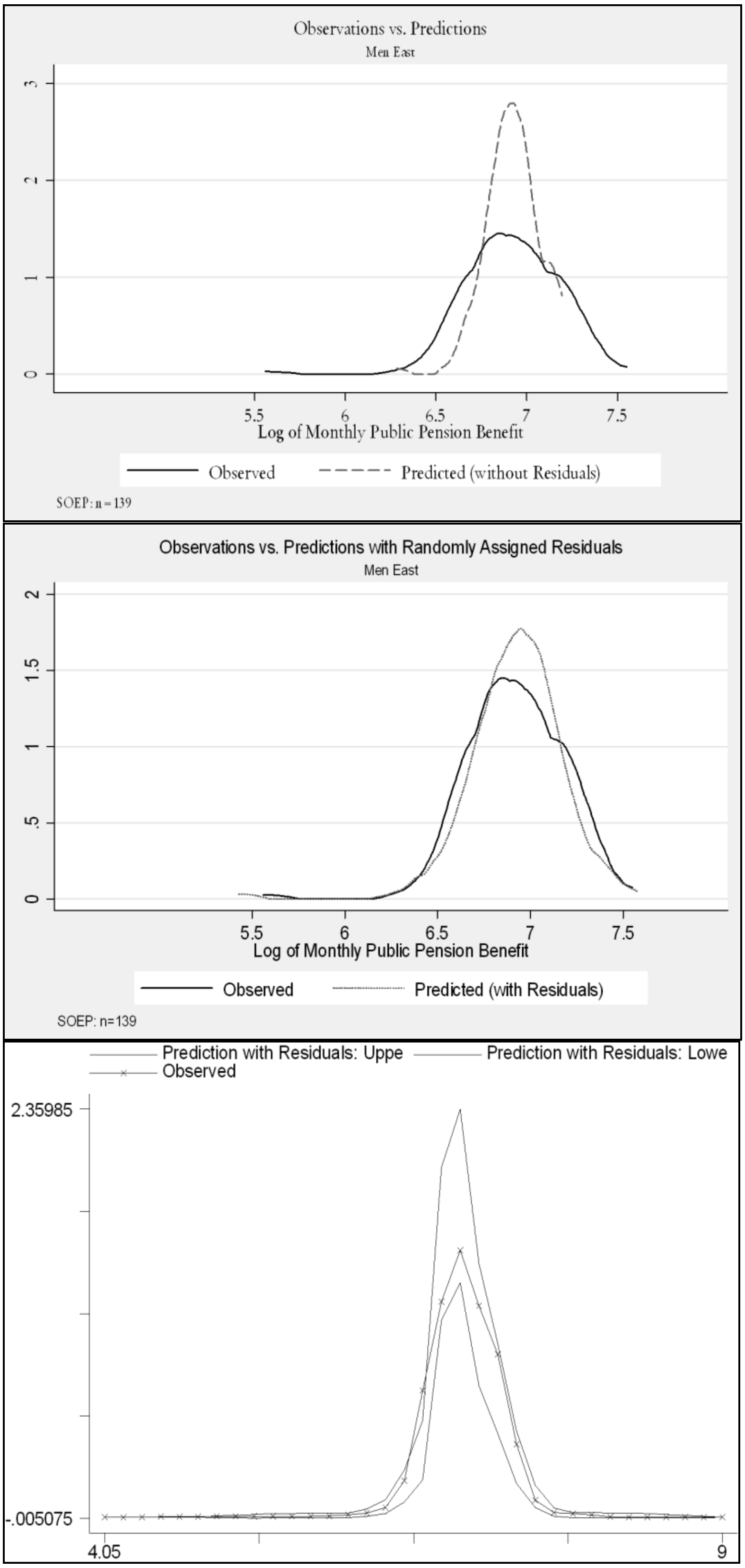
Source: SOEP 2005; Author's calculations

Figure A7 Observations and In-Sample Predictions, Men West



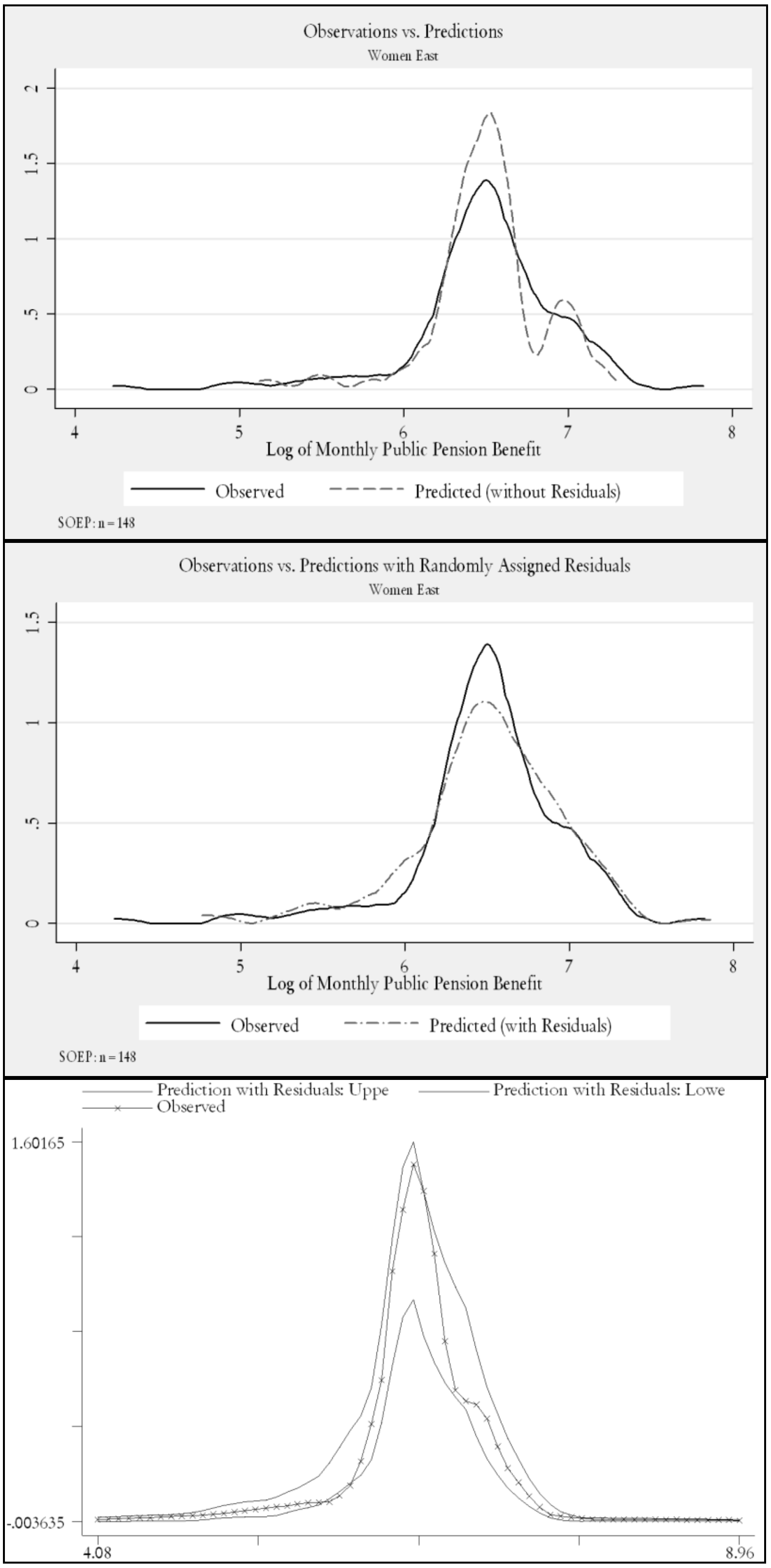
Source: SOEP 2005; Author's calculations

Figure A8 Observations and In-Sample Predictions, Men East



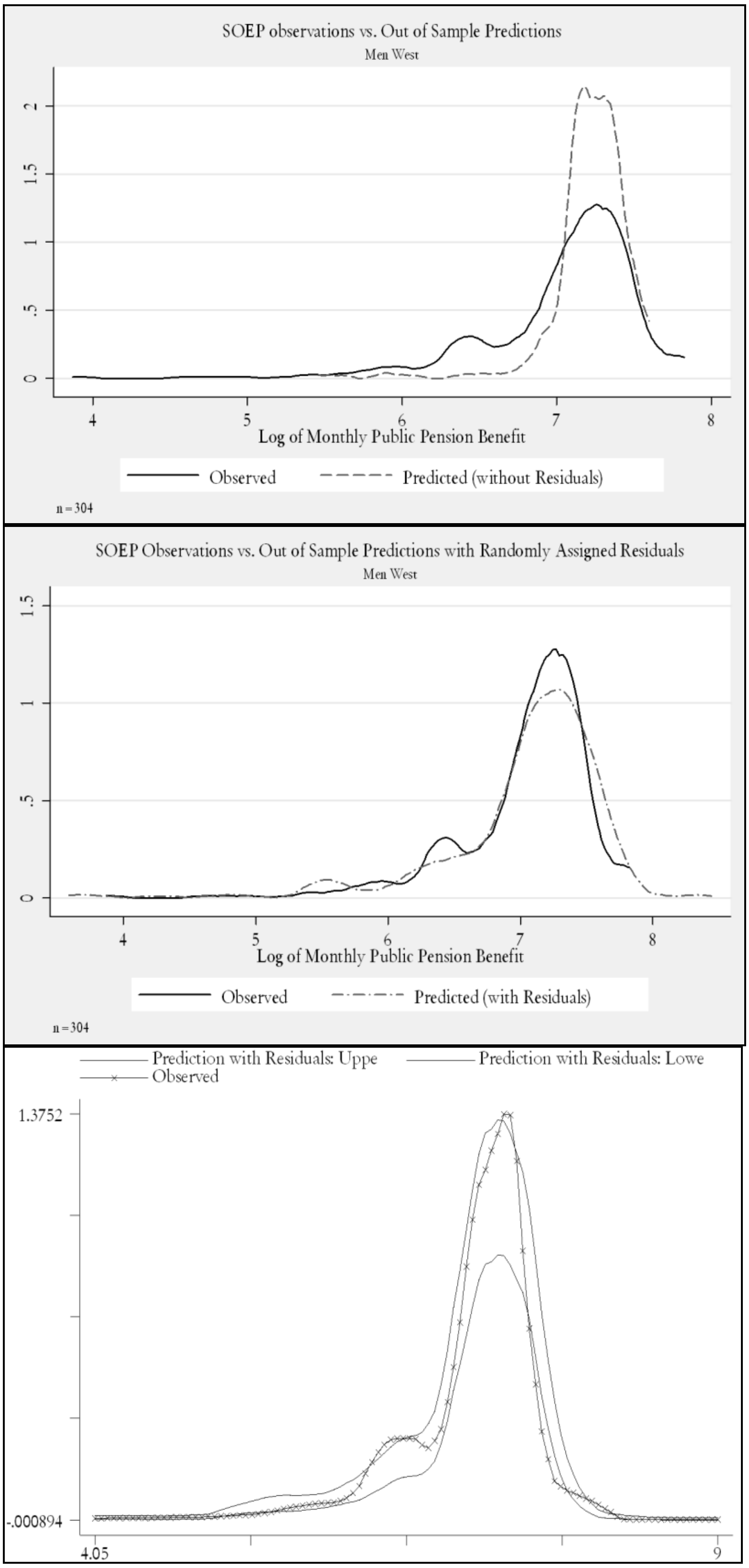
Source: SOEP 2005; Author's calculations

Figure A9 Observations and In-Sample Predictions, Women West



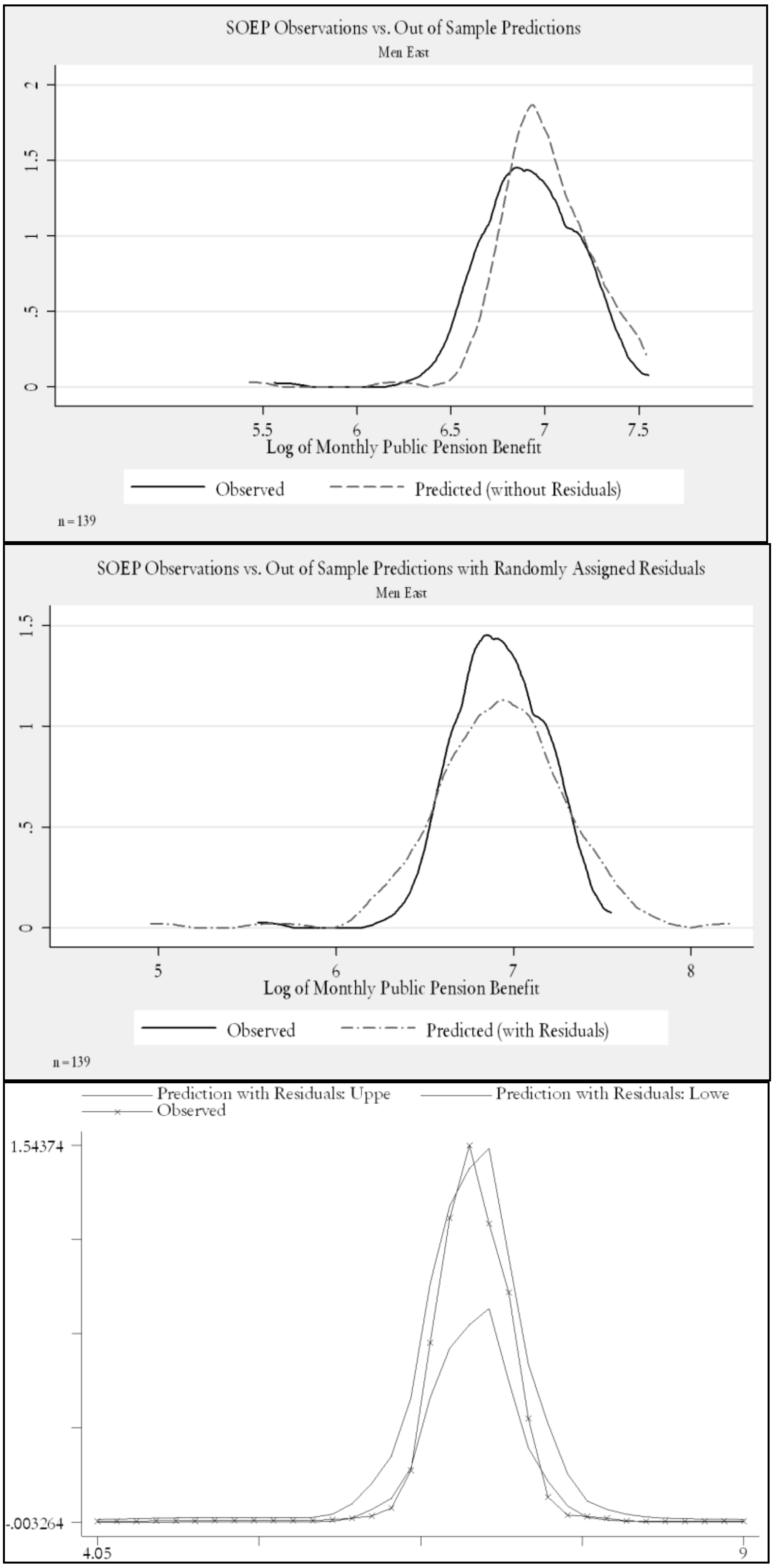
Source: SOEP 2005; Author's calculations

Figure A10 Observations and Out-of-Sample Predictions, Men West



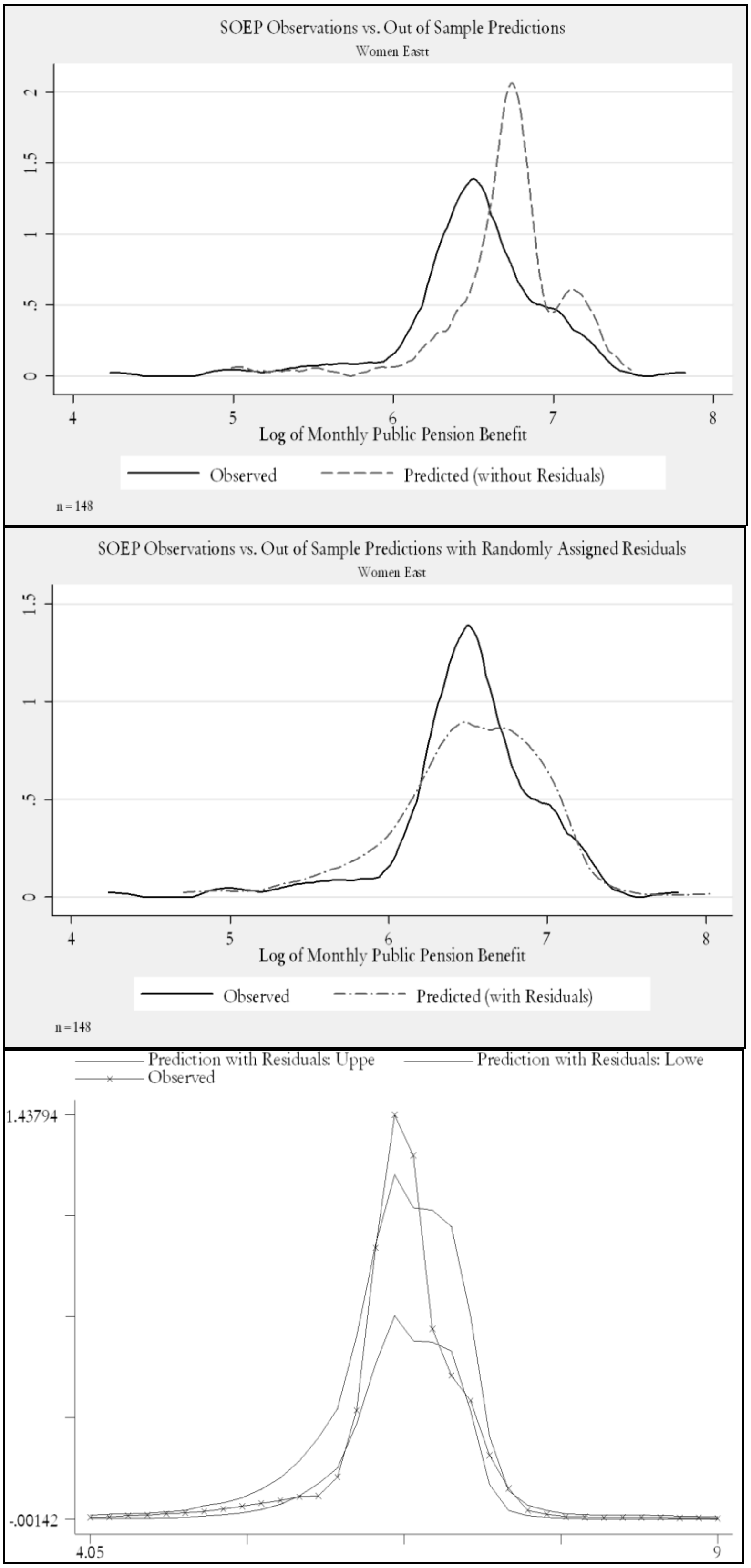
Source: SOEP 2005 and FDZ-RV - SUFVVL2004; Author's calculations

Figure A11 Observations and Out-of-Sample Predictions, Men East



Source: SOEP 2005 and FDZ-RV - SUFVVL2004; Author's calculations

Figure A12 Observations and Out-of-Sample Predictions, Women East



Source: SOEP 2005 and FDZ-RV - SUFVVL2004; Author's calculations

Table A36 Testing Matching Algorithms

	#1	#2	#3	#4	#5	#6	#7	#8
Income 1983-2006	✓		✓	✓	✓	✓	✓	✓
Years with Creditable Periods	✓	✓	✓	✓	✓	✓	✓	
Number of Child Credits	✓	✓	✓	✓	✓	✓	✓	
Years in Unemployment	✓	✓	✓	✓	✓	✓	✓	
Years in Employment								
Years in Education							✓	
Years with Non-Creditable Periods			✓	✓				
Home Production		✓			✓	✓	✓	
Age in Years	✓	✓	✓	✓	✓	✓	✓	

Notes: Creditable periods add up years in employment, school, training, military or civilian service, and non-professional care giving. Non-creditable periods sum up years in home production and years missing. Model #1 differs from all other options, because it includes disability pensioners. Model #2 is the only model that leaves out the income variables. Model #3 and Model #4 are different from each other because Model #3 sets home production for men equal to zero. Model #5 differs from Model #6 because it only considers home production for women and not for men. Model #7 is the only model that considers years in education as a separate matching variable. Model #8 only uses income information as a matching variable. Source: Author's illustration.

Table A37 Correlation Coefficients for Five Random Samples (Sampling with Replacement)

Sampling with Replacement	Sample 1 $Pension_{SOEP}$	Sample 2 $Pension_{SOEP}$	Sample 3 $Pension_{SOEP}$	Sample 4 $Pension_{SOEP}$	Sample 5 $Pension_{SOEP}$
Hotdeck	0,1413	0,3335	0,2309	0,3037	0,1560
UVIS	0,6667	0,6057	0,6784	0,7067	0,7208
Regression	0,6858	0,6314	0,6886	0,6531	0,6082
Mahalanobis	0,6043	0,6615	0,6771	0,6519	0,6218

Notes: Five 20 percent joint random samples from validated accounts in SAPA data. Observations can occur multiple times in the five samples or not at all. It is also possible that the best matching partner is not part of the five random samples. Source: SOEP 2007 and SAPA 2007; Author's calculations

Table A38 Correlation Coefficients for Five Disjoint Samples (Sampling without Replacement)

Sampling without Replacement	Sample 1 PensionSOEP	Sample 2 PensionSOEP	Sample 3 PensionSOEP	Sample 4 PensionSOEP	Sample 5 PensionSOEP
Hotdeck	0,3305	0,2553	0,2637	0,2678	0,3152
UVIS	0,6725	0,7108	0,6209	0,6762	0,6336
Regression	0,6375	0,6494	0,6144	0,6812	0,637
Mahalanobis	0,6150	0,6366	0,6903	0,6992	0,6472

Notes: Five 20 percent disjoint random samples from validated accounts in SAPA data. The samples are disjoint which implies that each observation can only occur in one of the five samples. Source: SOEP 2007 and SAPA 2007; Author's calculations

Table A39 Average Distance between Observed and Matched Benefit under Four Imputation and Matching Techniques, Total Population

Average Distance between Benefit$_{OBSERVED}$ and Benefit$_{MATCHED}$		Total Population (n=634)
Hot deck	Median	-14,30887
	Mean	16,38116
	Std. Dev.	542,8117
Regression	Median	-103,8114
	Mean	-105,5211
	Std. Dev.	328,781
UVIS	Median	-112,5852
	Mean	-101,5857
	Std. Dev.	327,4796
Mahalanobis	Median	-78,16614
	Mean	-74,7595
	Std. Dev.	320,5958

Notes: For the calculation of the average distance, we subtract the benefit amount obtained from the matching and imputation from the benefit amount reported in the SOEP. Source: SAPA 2007 and SOEP 2007; Author's calculations

Table A40 Average Distance between Observed and Matched Benefit under Four Imputation and Matching Techniques, Six Matching Strata

		Men East (n=126)	Men West (n=138)	Men Migrants (n=47)	Women East (n=141)	Women West (n=154)	Women Migrants (n=28)
Hot deck	Median	-67,08	28,32	193,95	-82,26	43,86	178,70
	Mean	-124,24	112,23	202,97	-53,56	39,08	90,88
	Std. Dev.	399,72	701,35	681,73	338,59	566,00	486,18
Regression	Median	-97,84	-34,74	-280,27	-101,25	-123,05	-137,09
	Mean	-114,72	-42,81	-163,02	-125,80	-111,41	-142,05
	Std. Dev.	290,75	386,51	408,82	298,02	314,92	218,07
UVIS	Median	-107,17	-67,71	-174,92	-81,51	-138,43	-165,30
	Mean	-105,78	-70,020	-131,47	-90,56	-119,46	-145,24
	Std. Dev.	256,29	447,69	388,51	250,39	312,04	218,96
Mahalanobis	Median	-100,66	-19,84	-170,33	-81,91	-88,88	-55,54
	Mean	-106,86	-34,77	-90,68	-70,46	-76,91	-110,36
	Std. Dev.	315,89	375,54	381,78	259,96	317,35	218,50

Notes: For the calculation of the average distance, we subtract the benefit amount obtained from the matching and imputation from the benefit amount reported in the SOEP. Source: SAPA 2007 and SOEP 2007; Author's calculations

Table A41 Correlation Coefficients for Different Matching Algorithms Under Four Matching and Imputation Techniques

	#1	#2	#3	#4	#5	#6	#7	#8	#9
Hot deck	0.17	0.24	0.21	0.21	0.21	0.21	0.21	0.21	0.21
UVIS	0.66	0.52	0.67	0.67	0.70	0.72	0.64	0.67	0.67
Regression	0.70	0.55	0.68	0.68	0.64	0.64	0.65	0.67	0.67
Mahalanobis	0.69	0.58	0.69	0.69	0.66	0.66	0.66	0.65	0.69

Note: For an overview of the nine matching algorithms go to *Table A36*. Model #9 is the final matching algorithm applied in Chapter 4. Source: SOEP 2007 and SAPA 2007; Author's calculations.

Table A42 Average Distance between Observed and Matched Benefit for Different Matching Algorithms under four Imputation and Matching Techniques, Total Population

		#1	#2	#3	#4	#5	#6	#7	#8
Hot	Median	2.00	-8.87	-11.62	-11.62	-14.30	-14.30	-16.49	-19.37
	Mean	11.79	11.91	17.03	17.038	17.12	17.12	17.35	14.40
	Std. Dev.	509.57	528.78	542.82	542.82	543.88	543.88	542.93	543.98
Reg	Median	-103.72	-146.18	-112.37	-112.37	-95.90	-95.90	-120.46	-71.38
	Mean	-107.72	-147.84	-112.57	-112.57	-100.90	-100.74	-111.69	-66.58
	Std. Dev.	326.89	392.078	325.05	325.05	345.16	344.87	335.74	342.07
UVIS	Median	-112.60	-132.52	-115.64	-115.64	-95.87	-93.40	-114.93	-75.53
	Mean	-108.36	-134.05	-104.52	-104.52	-97.00	-77.30	-115.48	-50.38
	Std. Dev.	338.25	387.37	333.08	333.08	328.72	304.39	325.048	328.37
Maha	Median	-85.81	-152.95	-81.65	-82.15	-99.54	-99.54	-92.11	-44.18
	Mean	-77.42	-151.97	-83.69	-84.17	-96.09	-96.74	-87.26	-44.33
	Std. Dev.	331.95	379.66	322.92	323.07	323.74	324.16	322.84	347.67

Note: For an overview of the nine matching algorithms go to *Table A36*. Model #9 is not displayed for lack of space. Source: SOEP 2007 and SAPA 2007; Author's calculations.

Table A43 Average Distance between Observed and Matched Benefit across Four Imputation and Matching Techniques – Five Joint Random Samples

Sampling with Replacement		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Regression	Median	-106,83	-108,75	-110,89	-110,73	-111,55
	Mean	-104,58	-116,74	-98,79	-110,15	-106,71
	Std. Dev.	328,86	350,44	323,95	331,96	348,83
UVIS	Median	-108,40	-119,01	-99,71	-125,62	-111,38
	Mean	-105,77	-125,70	-99,73	-133,75	-114,81
	Std. Dev.	343,78	359,89	327,03	320,57	325,57
Mahalanobis	Median	-70,26	-61,54	-81,91	-67,08	-71,31
	Mean	-69,58	-57,82	-63,35	-59,69	-72,65
	Std. Dev.	353,71	341,78	325,47	340,61	351,74

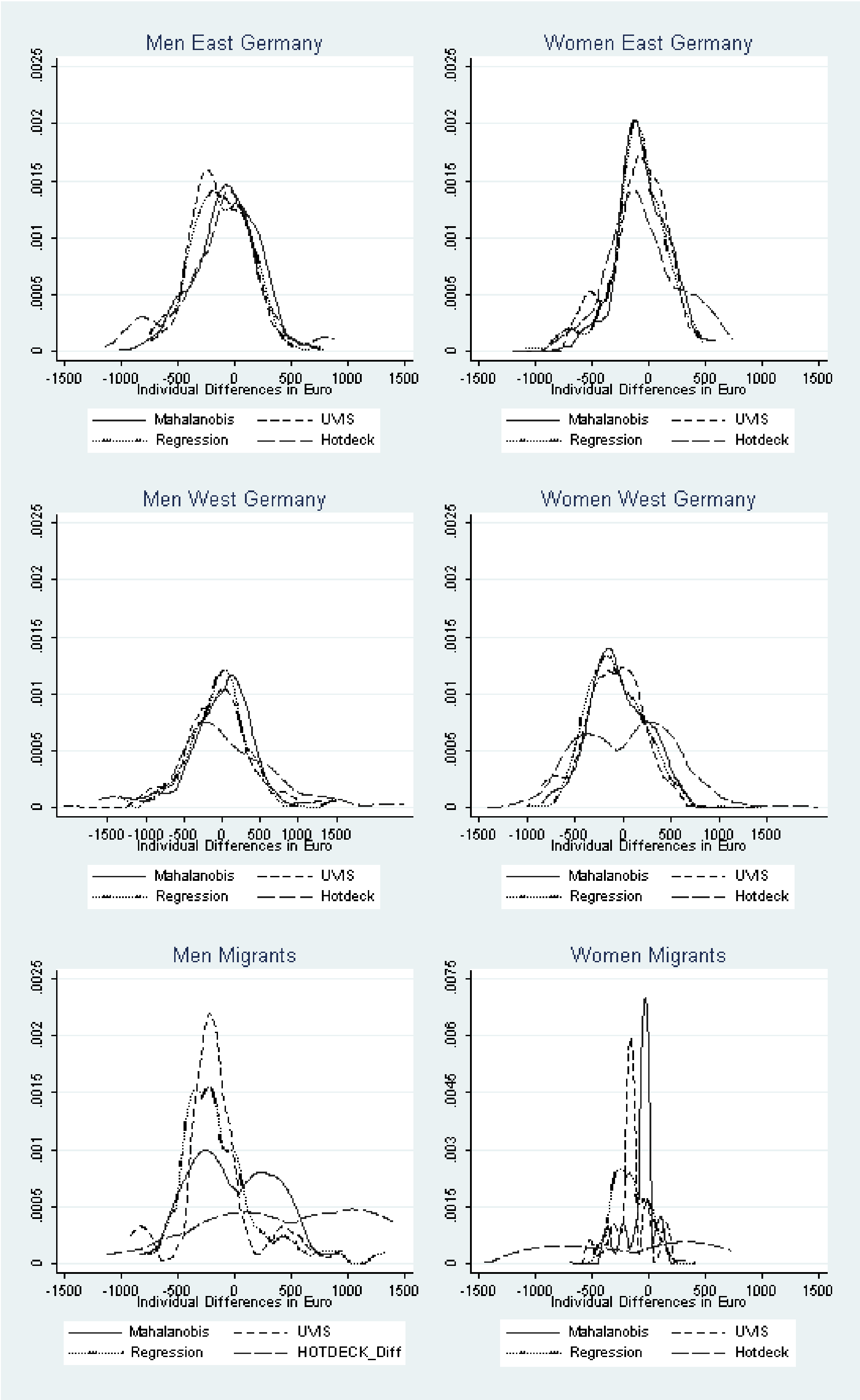
Source: Five 20 percent joint random samples from validated accounts in SAPA data. Observations can occur multiple times in the five samples or not at all. It is also possible that the best matching partner is not part of the five random samples. Source: SOEP 2007 and SAPA 2007; Author's calculations

Table A44 Average Distance between Observed and Matched Benefit across Four Imputation and Matching Techniques – Five Disjoint Random Samples

Sampling without Replacement		Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Regression	Median	-134,43	-103,74	-128,65	-104,13	-134,76
	Mean	-121,81	-105,32	-110,94	-101,76	-116,51
	Std. Dev.	347,56	340,41	349,87	338,14	340,53
UVIS	Median	-109,36	-109,88	-112,76	-101,27	-103,37
	Mean	-110,56	-96,909	-109,53	-103,89	-90,26
	Std. Dev.	329,85	319,96	346,56	346,89	344,26
Mahalanobis	Median	-63,91	-69,54	-81,39	-54,78	-70,39
	Mean	-58,00	-68,66	-70,40	-56,30	-67,12
	Std. Dev.	340,94	340,67	335,67	324,06	354,11

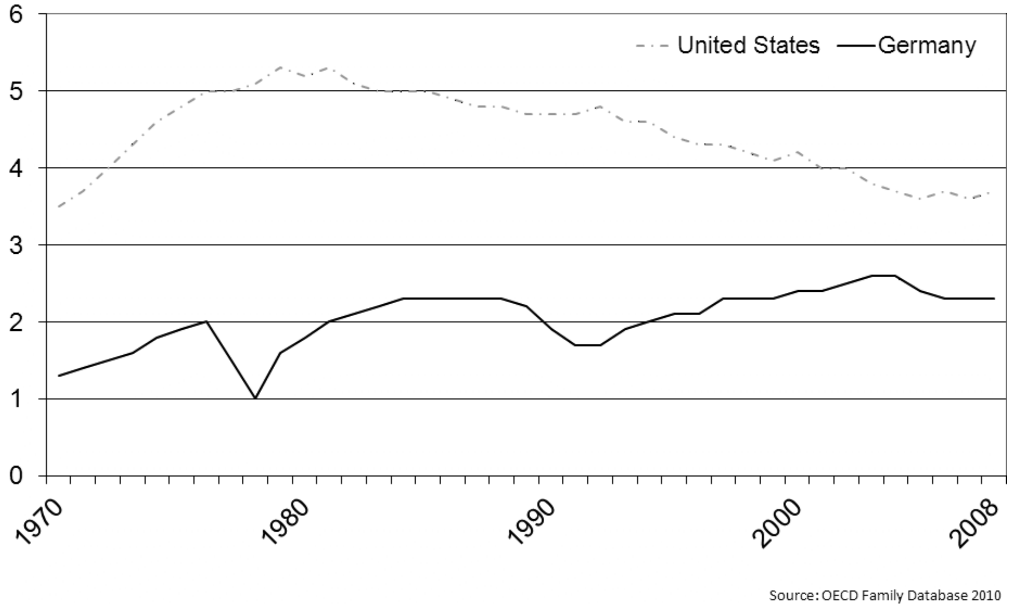
Notes: Five 20 percent disjoint random samples from validated accounts in SAPA data. The samples are disjoint which implies that each observation can only occur in one of the five samples. Source: SOEP 2007 and SAPA 2007; Author's calculations

Figure A13 Kernel Density Plots for Individual Differences between Observed and Matched Benefit Information – Total Population



Source: SOEP 2007 and SAPA 2007; Author's calculations

Figure A14 Crude Divorce Rate in Germany and the U.S. between 1970 and 2008



Source: OECD Family Database (2010)

Table A45 Synopsis of the German and U.S. Public Pension Scheme

Criteria	Germany	United States
Type of Pension Scheme	Defined benefit (pay-as-you-go)	Defined benefit (pay-as-you go)
Insurance	Old-age, disability, and survivor's pensions	Old-age, disability, and survivor's pensions
Access	All employees (except for self-employed and civil servants)	All employees (including the self-employed)
Financing	Contributions (2010: 9.95 percent of monthly earnings paid by employee and employer – total of 19.9 percent)	Payroll taxes (2010: 6.2 percent of monthly earnings paid by employee and employer – total of 12.4 percent)
Maximum Contribution Ceiling/ Taxable Maximum	€64,800	$106,800
Eligibility	Minimum of five years (different rules for special pension schemes)	Minimum of ten years (40 quarters in total)
Basis for Benefit Calculation	All years	Best 35 years
Redistribution	Credits for caregiving (children and elderly family members in need of care) Upgrade of below average contributions (for periods of child-rearing or low earnings [only temporary])	Progressive benefit calculation formula beneficial to low earners (bending points)
Administration	statutory pension insurance	Social Security Administration

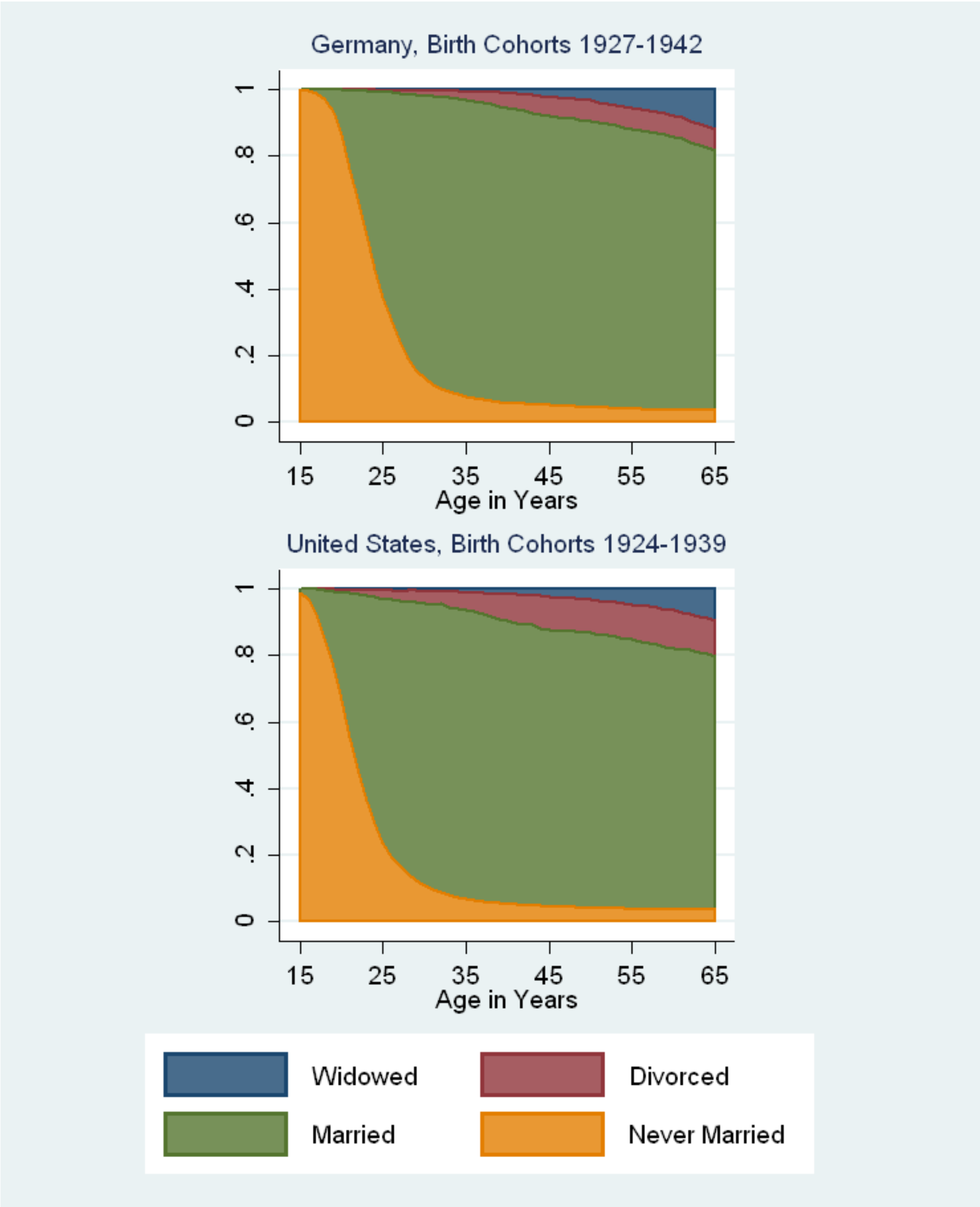
Source: Author's illustration

Table A46 Pension Benefit Calculation in the Public Pension Schemes in Germany and the U.S.

Criteria	Germany	United States
Benefit Calculation Formula	$PB = EP_i * PV_t * AA_i * PTF_i$	$PIA_i = 0.9 * 1^{st}\,BP_t + 0.32 * 2^{nd}\,BP_t + 0.15 * 3^{rd}\,BP_t$ BP_t *refer to* $AIME_i$
Factors of the Benefit Calculation Formula	Pension Benefit (PB): Is the monthly pension benefit a person receives. Sum of Earning Points (EPs): EPs describe the individual's earning position relative to the average earnings of all individuals that pay contributions into the public pension scheme in a given month. The EPs are summed up over the entire working life. Pension Type Factor (PTF): Depends on the type of pension a person applies for. The PTF ranges from one (old-age pension) to 0.25 (orphan's pension). Pension Value (PV): The PV serves the adjustment of past earnings to today's wage levels. The PV is equal for all individuals but is adjusted annually. For 2009, the PV amounts to € 27.20 in West and € 24.13 in East Germany. Actuarial Adjustment Factor (AAF): Factor reflects the age of retirement. The factor equals one if a person retires at the normal retirement age. For each month the person retires earlier the factor decreases by 0.3 percent. For each month the person retires later the factor increases by 0.5 percent.	Average Indexed Monthly Earnings (AIME): The AIME considers the 35 highest earning years indexed to growth in wages up to age 60. Best 35 years are first divided by 35 and then 12 in order to get workers average monthly earnings in today's wage levels. Primary Insurance Amount: Is the monthly benefit a person receives if he/she starts to draw benefits upon reaching the normal retirement age (no actuarial adjustment for early or delayed retirement). Bend points (BPs): Serve the purpose of calculating the primary insurance amount (PIA). The bend points decompose the AIME into three parts and assign weights to each respective part. The BPs are adjusted annually. For 2009 they amount to First BP: 90 percent of $0 to $744 of AIME Second BP: 32 percent of $745 to $4483 of AIME Third BP: 15 percent of $4484 > of AIME

Source: Author's illustration

Figure A15 Marital Status across Age in Germany and the U.S., Retiree Population



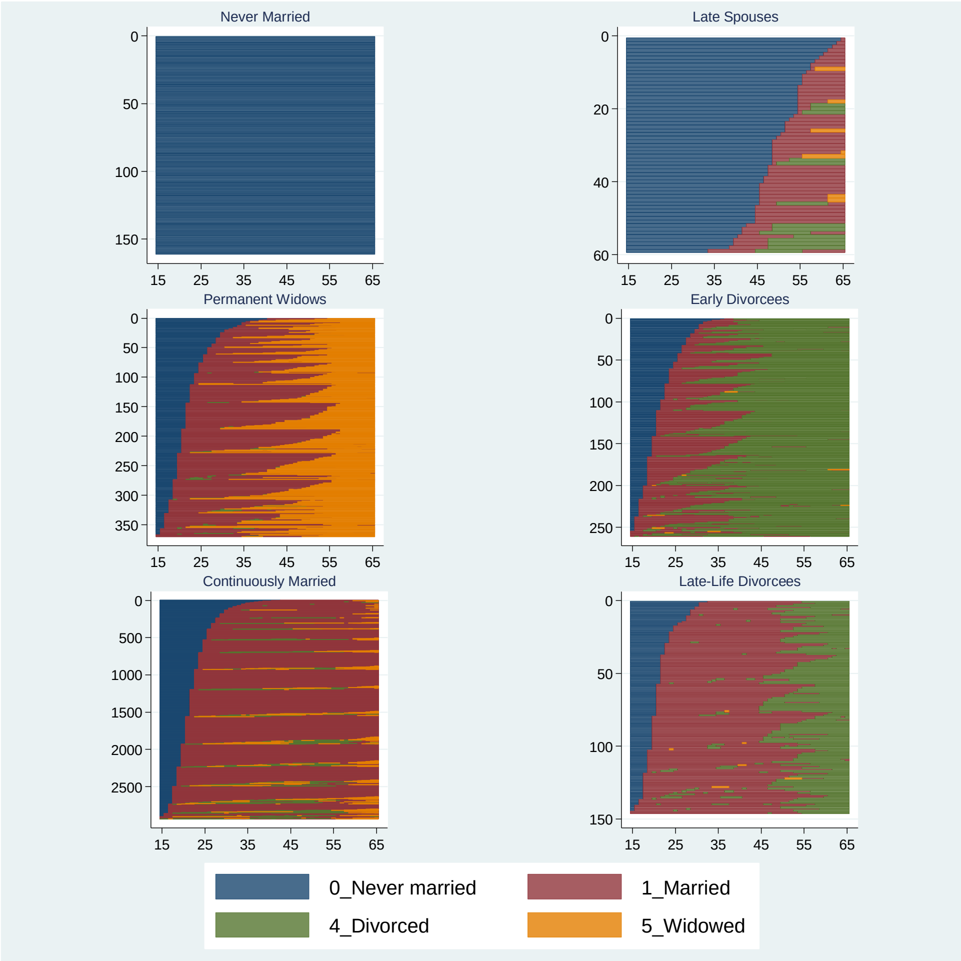
Note: Results are weighted with wave-specific individual-level weights in HRS and SOEP data. Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

Table A47 Comparing Summary Statistics for HRS and SSA Permissions 2004

Variable	Label	HRS 2004 – No Permission	HRS 2004 – Permission
Gender (in percent)	Male	46.9	49.7
	Female	53.2	50.3
Race (in percent)	Non-Hispanic White	80.9	81.0
	Non-Hispanic Black	9.4	9.4
	Hispanic	7.5	7.1
	Other	3.0	2.6
Age in Years	Mean	62.3	62.3
Census Region (in percent)	North East	16.9	17.9
	Midwest	26.1	24.7
	South	38.0	37.2
	West	18.8	20.2
	Other	0.1	0.1
Years in Education	Mean	12.9	12.9
Education (in percent)	Less than High School	16.7	16.7
	GED	4.5	4.9
	High School Graduate	29.6	29.7
	Some College	23.8	24.6
	College and Above	25.4	25.2
Longest Job Tenure	Mean	18.3	18.0
Children (Average)	Number of Children	2.9	3.1
Marriages (Average)	Number of Marriages	1.4	1.4
Longest Marriage in Years	Mean	28.7	28.7
	Median	30.1	29.8
Never Married (in percent)	Ever Married	95.7	94.9
	Never Married	4.3	5.1
Number of Divorces (in percent)	Zero	63.9	60.2
	One	26.2	28.6
	Two	7.8	8.7
	Three or more	1.8	2.5
Living in Poverty (in percent)		0.09	0.08
Total Household Income	Mean	473,531	450,799
	Median	187,770	170,100
Social Security	Mean	7,734	7,965
	Median	7,800	8,358
Pensions and Annuities	Mean	8,636	8,841
Number of Observations	n	9,174	7,685

Notes: Deviations from 100 percent are due to rounding. The mean value for income from social security as well as pension and annuities only considers persons who report to be fully retired. The median for pension and annuities equals 0, indicating that more than 50 percent of all retired persons don't receive any income from these sources. Source: HRS 2004 and SSA Permissions 2004; Author's calculations

Figure A16 Clusters of Marital Trajectories in Germany and the U.S., Retiree Population



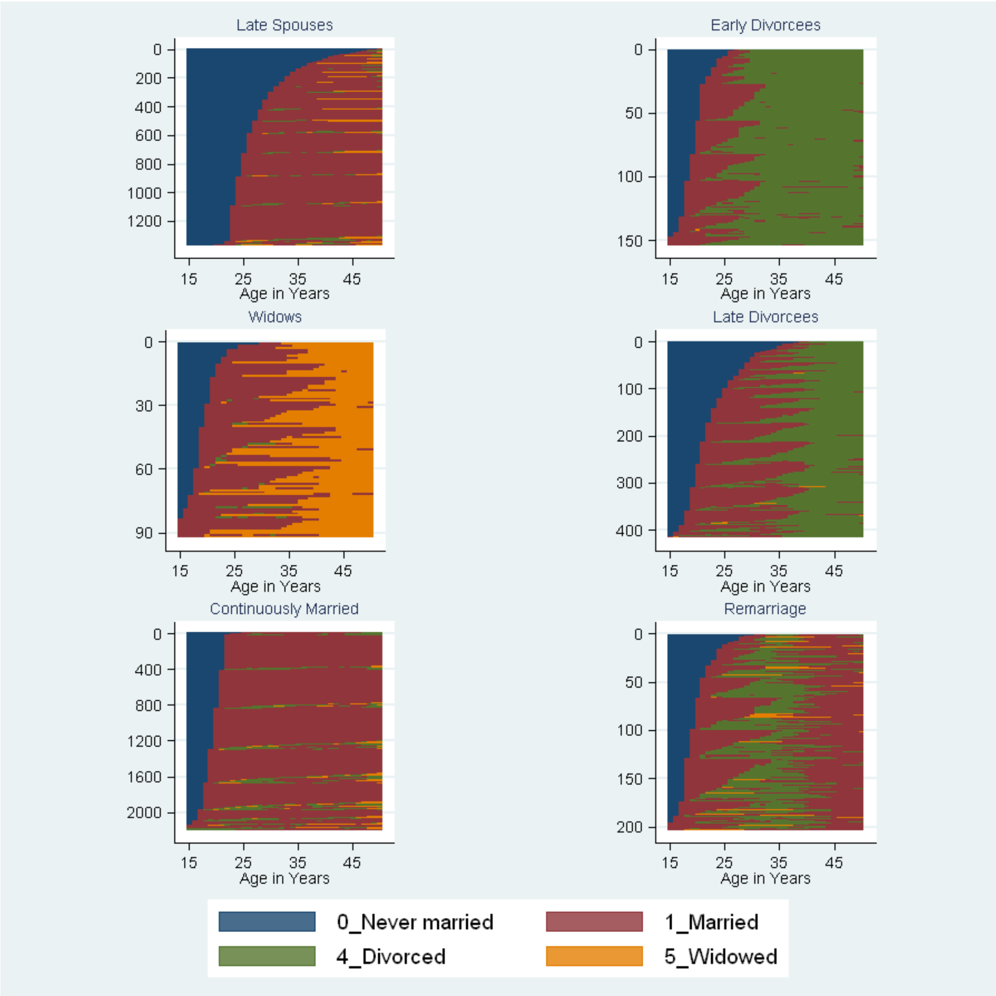
Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

Table A48 Determinants of Monthly Public Pension Benefits of Retired Women in East and West Germany

	East Germany		West Germany	
	b	se	b	se
Marital Clusters (Ref.: Never Married)				
Late Spouses	-145.395	(81.38)	-58.654	(130.77)
Early Divorcees	28.238	(62.37)	-68.856	(66.66)
Mid/Late-Life Widows	-205.995***	(54.47)	-107.617	(64.91)
Late Divorcees	13.593	(72.01)	-92.909	(86.39)
Continuously Married	-288.727***	(49.01)	-169.340**	(56.54)
Experienced Widowhood	56.659	(37.77)	-25.451	(38.48)
Experienced Divorce	213.629***	(33.85)	-14.161	(38.53)
Widowed after Age 65	-9.911	(29.66)	-12.634	(32.58)
Number of Children (Ref.: No Children)				
One Child	-90.655**	(33.97)	47.297	(38.93)
Two Children	-118.428***	(32.29)	67.497	(38.55)
Three + Children	-88.821**	(33.23)	100.455*	(39.66)
Educational Attainment				
No Degree	85.351***	(22.69)	119.459***	(28.15)
Degree Unknown	145.634***	(43.30)	68.069	(64.59)
Intermediate Secondary School	-74.485*	(34.04)	386.966***	(40.85)
A-Levels	49.473	(48.66)	11.015	(83.50)
College Degree	-1.808	(59.47)	10.052	(78.68)
Received Vocational Training	70.475***	(20.01)	58.406*	(24.96)
Number of Years Worked	10.678***	(0.64)	10.901***	(1.22)
Working Past Age 65	-192.631***	(39.13)	-130.076	(78.08)
Migrant	-63.479	(43.63)	.	.
Constant	506.541***	(51.02)	416.111***	(75.19)
R-Squared	0.342		0.29	
N	1529		569	

Notes: [1] The reference category for educational attainment is lower secondary school (Hauptschule). Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Selected parameter estimates only. Abbreviations: n/a = not applicable. Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

Figure A17 Clusters of Marital Trajectories in Germany & the U.S., Pre-Retirement Cohorts



Notes: *Figure A16* omits the never married cluster because the sequence plot is the same as in *Figure A15*. Only the number of individuals that fall in the *never married cluster* is higher in the pre-retirement cohorts with 114 individuals in Germany and 117 in the U.S., respectively. Source: HRS Permissions 2004 and SOEP 2007; Author's calculations

Table A49 Clusters of Marital Trajectories and their Prevalence in Germany and the U.S., Pre-Retirement Population

	Cluster 1 Never Married	Cluster 2 Late Spouses	Cluster 3 Early-Life Divorcees	Cluster 4 Mid-/ Latelife Widows	Cluster 5 Late-Life Divorcees	Cluster 6 Continuously Married	Cluster 7 Remarriage
Average Duration being ...							
Never Married	36.0	13.0	4.9	4.2	7.7	4.8	5.1
Married	0	22.4	8.1	16.7	14.8	30.4	19.6
Divorced	0	0.3	23.0	0.5	13.4	0.8	10.5
Widowed	0	0.3	0	14.7	0	0.1	0.7
Prevalence of Cluster ...							
Germany Share in Percent/(n)	4.3 (114)	34.6 (927)	2.8 (74)	1.8 (48)	7.7 (207)	46.1 (1,235)	2.8 (74)
U.S. Share in Percent/(n)	5.9 (117)	22.4 (446)	4.0 (80)	2.2 (44)	10.6 (210)	48.4 (963)	6.5 (129)

Source: HRS Permissions 2004 and SOEP 2007; Author's calculations