

SOEPpapers

on Multidisciplinary Panel Data Research

359

Anika Rasner • Joachim R. Frick • Markus M. Grabka

**Extending the Empirical Basis for Wealth Inequality
Research Using Statistical Matching of
Administrative and Survey Data**

Berlin, January 2011

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
<http://www.diw.de/soeppapers>

Editors:

Georg **Meran** (Dean DIW Graduate Center)

Gert G. **Wagner** (Social Sciences)

Joachim R. **Frick** (Empirical Economics)

Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics)

Christoph **Breuer** (Sport Science, DIW Research Professor)

Anita I. **Drever** (Geography)

Elke **Holst** (Gender Studies)

Martin **Kroh** (Political Science and Survey Methodology)

Frieder R. **Lang** (Psychology, DIW Research Professor)

Jörg-Peter **Schräpler** (Survey Methodology)

C. Katharina **Spieß** (Educational Science)

Martin **Spieß** (Survey Methodology, DIW Research Professor)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Extending the Empirical Basis for Wealth Inequality Research Using Statistical Matching of Administrative and Survey Data*

by Anika RASNER¹, Joachim R. Frick² and Markus M. Grabka³

Abstract:

Social security entitlements are a substantial source of wealth that grows in importance over the individual's lifecycle. Despite its quantitative relevance, social security wealth has been thus far omitted from wealth inequality analyses. In Germany, it is the lack of adequate micro data that accounts for this shortcoming. The two main contributions of this paper are: First, to elaborate a statistical matching approach that complements information on net worth as surveyed in the German Socio-Economic Panel (SOEP), a population representative panel study, with information on social security wealth from the Sample of Active Pension Accounts (SAPA), a large-scale administrative dataset maintained by the German Statutory Pension Insurance. Second, we show to what extent the inclusion of social security wealth affects the level and the distribution of individual net worth as well as overall inequality. The present value of pension entitlements (including entitlements from the statutory pension system as well as from the separate system for civil servants) amounts to 5.6 trillion Euros, which corresponds to an average of 78,500 Euros per person – thus almost doubling the level of net worth. Compared to results based on net worth only, inequality of our amended wealth measure is about 25 percent less. Moreover, we present significant differences in pension entitlements across occupational groups with civil servants gaining most from the inclusion of public pension wealth in the extended wealth measure and self-employed benefiting the least. Overall, our results provide clear indication for the relevance of including the notional wealth held in pension entitlements providing a less biased picture of the level and the socio-economic structure of wealth in Germany. Above and beyond such within-country variation, our findings may also be most relevant for comparative analyses across welfare-regimes.

JEL-codes: C49, D31, D63, I39

Keywords: Wealth inequality, statistical matching, public pension entitlements, SOEP

* The authors gratefully acknowledge David Kantor for very helpful comments on Mahalanobis matching in STATA and Rainer Siegers for excellent programming assistance.

¹ Corresponding Author: DIW Berlin / SOEP, Mohrenstrasse 58, 10117 Berlin and MPI Rostock. Email: arasner@diw.de.

² DIW Berlin / SOEP, TU Berlin and IZA Bonn, jfrick@diw.de

³ DIW Berlin / SOEP and TU Berlin, mgrabka@diw.de

CONTENTS

1	INTRODUCTION	3
2	THE SYSTEM OF OLD AGE PROVISION IN GERMANY	5
3	DATA DESCRIPTION AND MATCHING VARIABLES	8
	3.1 Data Limitations and Availability	8
	3.2 Data Description	9
	3.3 Matching Variables, Slice Variables, Matching Population	10
	3.4 Statistical Matching: Alternative Approaches	12
4	STATISTICAL MATCHING OF RETIREES	16
	4.1 Evaluation Criteria and Sample Specification	16
	4.2 Results	16
	4.3 Discussion	19
	4.4 Taking Divorce into Consideration	22
5	WEALTH INEQUALITY	24
	5.1 Calculating the present value of pension entitlements	24
	5.2 Individual net worth without pension entitlements	25
	5.3 Individual net worth considering pension entitlements	27
	5.4 Wealth inequality and pension entitlements	29
	5.5 Net worth and pension entitlements by occupational status	32
6	CONCLUSION	36
7	REFERENCES	39
8	APPENDIX	41

1 Introduction

So far, public pension wealth¹ has not been adequately considered in research on the distribution of wealth even though essential in obtaining unbiased wealth estimates. This is surprising for two reasons: First, in most countries public pension wealth is a substantial component in the individual's wealth portfolio that grows in importance over the adult's working life. Second, with the payment of compulsory contributions shared equally by employers and employees, the majority of the working age population saves a significant amount of their gross earnings in the statutory pension scheme month by month. For countries like Germany with their matured pay-as-you-go pension (PAYG) systems, the latter is clearly more important in quantitative terms than it is in liberal welfare systems that put a greater emphasis on private old-age provision. In cross-country analyses, the omission of social security wealth from wealth estimates clearly raises issues of comparability (Frick and Headey 2009).

For each individual, the accumulation of social security wealth usually starts with the first job that is subject to social insurance contributions and ends with the transition into retirement. Therefore, social security wealth is a reflection of the individual's earnings history and retirement benefits are a proxy for a person's lifecycle labor market attachment. Whether a person accumulates social security wealth also depends on the employment status. While the majority of dependent workers are protected by Germany's statutory public pension scheme (*Gesetzliche Rentenversicherung*, henceforth *GRV*) and are thus obligated to contribute, civil servants and the majority of self-employed are not, and hence, are either covered by separate schemes or through their investment in private pension funds.

Apart from individual-level factors, the accumulation of social security wealth also depends on the respective legal, political and economic framework the individual operates in. In recent years, this institutional context has undergone a sea change - in parts because of numerous labor market and public pension reforms. These changes involve more flexibility, greater job insecurity and less social protection in the labor market. As a consequence, employment trajectories become more heterogeneous with likely adverse effects for the accumulation of social security wealth, at least for some groups of the population. At the same time, there is increasing pressure on the overall financial viability of the German pay-as-you-go pension system (Gruber and Wise 1999).

¹ The terms public pension wealth and social security wealth are defined as the total sum of individual pension rights accrued over the life-course. These pension rights translate into old-age public pension benefits as the individual retires. Both terms are used interchangeably throughout the paper and refer to entitlements in the old-age public pension scheme only.

This is – at least partly – due to far-reaching demographic changes resulting from an unprecedented rise in life-expectancy (Christensen et al. 2009) as well as decreases in fertility rates (Kreyenfeld 2009). These demographic pressures have prompted a sequence of reforms in the public pension system manifested in benefit cuts and changes in the public/private-mix in pension provision (Bonin 2009). When compared to today’s pensioners, future retirees with the same life-cycle labor supply will render less social security wealth.

All of the above developments will have significant repercussions on economic inequality among future retirees in Germany. Quantifying these distributional consequences is of innate importance, however difficult because of the lack of adequate micro data. Population surveys – at least in Germany – don’t ask for social security wealth, because respondents usually lack knowledge about the present value of their respective entitlements.² Administrative data keep detailed information about pension entitlements. However, information about other wealth categories and the individual’s socio-economic background including the household context is missing. Ideally, survey and administrative data could be directly linked with a unique identifier in order to bring together the *best of both worlds*. In Germany, record linkage is infeasible for data confidentiality reasons. Hence, *statistical matching* may be the second best solution but the one and only way to overcome the drawbacks of both data sources and make use of their respective merits.

In this paper, we present a *double* matching approach that provides an important methodological contribution. The matching links data from up to 24 waves of the population representative German Socio-Economic Panel with information from the Sample of Active Pension Accounts maintained by the German Social Security Administration. An additional link to data from the pension rights splitting divorce statistics (*Versorgungsausgleichsstatistik*) controls for otherwise unconsidered effects arising from a marital split. The matched data not only allow for the evaluation of effects reforms in the public pension scheme and the labor market have on social security wealth, it is also the most relevant prerequisite to reduce bias in wealth inequality research.

² In 2006, the SOEP group performed a pretest asking for the person’s social security wealth. The question generated more than 92% missings, the reliability of information given by the remainder of respondents are questionable.

The paper sets out some background information on the system of old age provision in Germany (Section 2). In Section 3, we describe the data and present the matching variables. Section 4 presents four potential imputation and matching strategies and compares the results of robustness tests for the group of retirees to understand which matching approach fits our data best. The technique that serves our purpose best is then applied to the total population providing the basis for a comprehensive wealth measure. Section 5 illustrates the effects of considering pension wealth in wealth inequality analyses.

2 The System of Old Age Provision in Germany

The statutory pension insurance is 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.³ In almost every job, employees pay contributions into the social security system, a certain fixed share of their earnings up to some maximum amount.⁴ The employer matches these payments.⁵ By paying contributions into the system, employees accumulate entitlements in the form of so-called *earning points*.⁶ Periods of non-employment can translate into *earning point equivalents* or pension rights as well. These periods are spells of education, unemployment, sickness, childcare or long-term care etc. The total sum of earning points multiplied by the actual pension value, and a factor reflecting the individual's retirement age then translates into the person's final pension benefit upon retirement. Therefore, public pensions are roughly proportional to the individual's labor income.⁷

For today's generation of retirees, they are still the predominant source of retirement income (Kortmann and Halbherr 2008c), with occupational and private pensions clearly playing a secondary role only. On average, the monthly public pension benefit of West German male white-collar workers amounts to 1,442 Euro in 2007 (Kortmann and Halbherr 2008a; Kortmann and Halbherr 2008b). To qualify for this average benefit, the employee and employer(s) have to pay a

³ Actively insured persons have at least one period of paid contributions (payment of compulsory or voluntary social insurance contributions, marginal employment) or creditable periods (*Anrechnungszeit*) stored in the individual's pension account.

⁴ In so-called *mini jobs*, one type of marginal part-time employment, individuals can earn up to €400, where only the employer pays 22 percent in contributions for health care and pension insurance. The sum of €400 is exempt from income tax and employee contributions.

⁵ For 2010, the contribution level is 19.9 percent paid in equal parts by employee and employer.

⁶ The individual earning points describe the earnings position of an individual relative to the average earnings of all the individuals that pay contributions into the public pension scheme. For any year t , the earning point equals 1 if the i^{th} individual earns as much as the average of all contributors in time period t .

⁷ The principle of equivalence, one of the guiding principles underlying the German public pension system, implies that public pension benefits drawn are roughly equivalent to contributions paid into the system.

total of 316,000 Euro (expressed in 2007 Euro) in pension insurance contributions over the course of the employee's working life. This amount stresses the relevance of social security entitlements in wealth inequality analyses, particularly for dependent workers, as the obligatory contributions cannot be invested in alternative forms of old-age provision.

Civil servants and the self-employed don't pay contributions into the public pension scheme. These occupational groups are either covered in separate schemes or invest in private pension arrangements.⁸ In 2008, the government-sponsored, non-contributory civil service pension scheme covers 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). With its generosity the German civil service pension scheme stands out in comparative perspective, mitigating the need for additional retirement income from other sources.⁹ The system of old-age provision is most heterogeneous for the 4.5 million self-employed in Germany. In fact, about 25 percent of them are covered by compulsory schemes, such as farmers or the liberal professions, but to very unequal conditions in terms of coverage and the provision of benefits (Loose and Frommert 2009).¹⁰ The rest lacks a formal coverage: some rely exclusively on private pension arrangements, whereas others accumulate entitlements in several different schemes.¹¹ This brief glance at the system of social security in Germany illustrates the quantitative relevance of the public pension scheme in the provision of retirement income for a large share of the active working age population. Furthermore, the composition of old-age income is highly correlated with the person's occupational status.

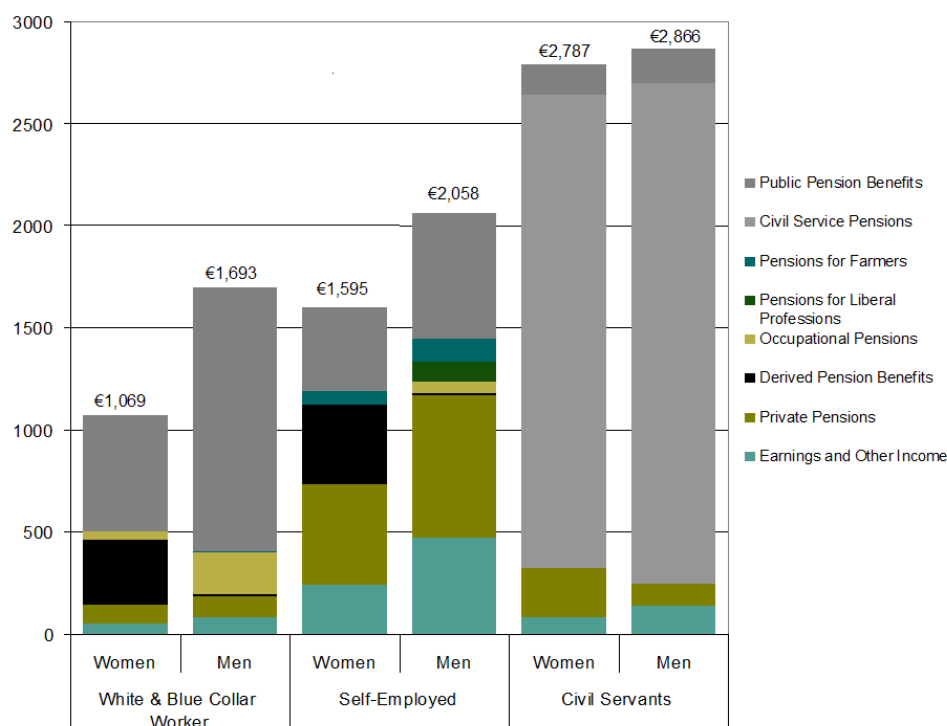
⁸ A certain share of persons is not in the civil service from the beginning of their career. Typically, they have already accumulated some entitlements in the public pension scheme prior to their transition into the civil service. At retirement, these entitlements are credited against their civil service pension.

⁹ 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. The higher tax burden decreases the replacement rate of former civil servants relative to that of dependent workers. On average, civil servants pay about €558 in taxes and transfers compared to €148 paid by dependent workers (Bundesministerium für Arbeit und Soziales 2008, p. 208 ff.)

¹⁰ While benefit levels for farmers are comparatively low, self-employed in the liberal professions enjoy replacement rates comparable to those of civil servants.

¹¹ The employment patterns of the growing group of solo-self-employed show alternating spells of self- and dependent employment explaining that they also receive benefits from the public pension scheme.

Figure 1 Composition and Level of Monthly Old-Age Income before Taxes for Men and Women Aged 65 and older



Source: ASID 2007, own calculations

These issues are captured in *Figure 1* illustrating the differences in level and composition of monthly old-age income for men and women aged 65 and older by occupational status.¹² For men, between-group differences in total income are very pronounced: on average, civil servants enjoy 70 percent 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. 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

¹² The categorization is based on information about the person's last occupational status prior to retirement.

indicating that many continue to work above age 65.¹³ Formerly self-employed women also rely on survivor's benefits amounting to 25 percent of their total income.

3 Data Description and Matching Variables

3.1 Data Limitations and Availability

The consideration of social security wealth in inequality research requires the use of complex micro data that have to satisfy certain criteria: First, these data have to be representative of the total population ideally providing a large number of observations to allow for subgroup analyses. Second, they have to include information on life-cycle earnings and wealth not only at the individual level but also for all members of the household and needs sharing unit. In Germany, neither administrative, nor survey data meet these criteria.

Although administrative data¹⁴ are often appreciated for the large number of observations and their presumably high precision, these data serve administrative purposes in the first place. However, these purposes don't necessarily meet researchers' interests and needs. For example, it is defined by program rules whether or not a person is part of the sample population.¹⁵ Therefore, administrative data might be selective for certain segments and hence, not always representative for the population as a whole. From a welfare economics perspective even more problematic, these data only include the individual's pension-relevant income. Information on income above the maximum contribution ceiling or wealth at the level of the sharing unit is not available. Furthermore, the data lacks other socio-economic information at the household level making distributional analyses practically impossible. In contrast, household panel surveys tend to be representative of the total population. But at the same time, they typically suffer from smaller numbers of observations and (selective) panel attrition. With incomplete lifecycle earnings and employment history information, it is infeasible to approximate social security wealth for all individuals who are not yet retired.

Augmenting survey information with administrative data over a unique identifier is one option to obtain more comprehensive data. So-called *record linkage* is a common practice in countries like the US or the UK, typically carried out using the individual's Social Security Number (*SSN*) or National Insurance Number (*NINO*). At least in Germany record linkage is not

¹³ For the group of self-employed, working above age 65 is not necessarily an economic exigency, but an intrinsic motivation.

¹⁴ The terms register data and administrative data are used interchangeably.

¹⁵ Due to legal changes, these program rules change over time (Stegmann 2008).

permitted without the explicit informed consent on behalf of survey respondents.¹⁶ Furthermore, it is technically infeasible due to the lack of unique identifiers.

Hence, *statistical matching* of administrative and survey data may provide a second best solution to overcome the drawbacks of both data sources (Rasner et al. 2007) and to reduce bias in wealth inequality research.¹⁷ Unlike record linkage, statistical matching does not aim at finding the exact same person, but links cases that are statistically similar – similar in terms of certain characteristics which are systematically related to the object under investigation and which are observed in a most similar way in both data sets.

3.2 Data Description

This paper presents a two-step matching procedure involving three data sets. The first match employs 24 waves of panel data from the population representative German Socio-Economic Panel Study (*SOEP*) and links them with the Sample of Active Pension Accounts (*Versichertenkontenstichprobe*, short *VSKT*) using statistical matching. The matched data is then complemented by information from the *Statistic for Pension Rights Splitting in Case of Divorce* (*Versorgungsausgleichsstatistik*, henceforth *Divorce Statistics*) another data set maintained by the German Social Security Administration that covers information on the splitting of pension rights following a divorce. This second matching can be accomplished by means of *record linkage*, as both datasets are available within the Statutory Pension Insurance.

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. Today, 25 waves of data for West and 19 waves for East Germany are available. The most recent accessible data was collected in 2008 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

¹⁶ Unlike record linkage statistical matching does not require any consent agreement on behalf of survey respondents. 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 et al. 2006).

¹⁷ According to data protection and confidentiality rules it is allowed to match two micro-data sets if both of them are considered as *de facto* anonymized. In this case also the matched file is considered to be *de facto* anonymized and thus no informed consent on behalf of the survey respondents for a record linkage is required.

topic modules are asked every few years. In 2007, a special wealth module collected detailed wealth data at the individual level (see Frick et al. 2007), but no information about social security entitlements.

The Sample of Active Pension Accounts (*VSKT*) is a one percent random sample of pension accounts, containing records for approximately 570,000 individuals, both actively insured and already retired. These records are representative of all individuals holding a pension account.¹⁸ The *VSKT* consists of several separate but linkable sub files. The so-called *fixed file* contains demographic and benefit information, including the individual's aggregated entitlements at the time the sample was drawn. The other files contain longitudinal information on monthly earnings, unemployment spells, and periods of child care as well as long-term care, etc. Unlike survey data, administrative records only provide information at the individual-level, but no household-based information, whatsoever.

The *Divorce Statistics*¹⁹ serves the purpose of administering the pension rights splitting. The data 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 1977.²⁰ Further, these statistics contain information about marriages and divorces that goes beyond information provided in the *VSKT*.

The most important outcome measure resulting from this *double* matching exercise is the present value of the individual's pension wealth as accumulated by 2007 in the statutory pension scheme. Along with data collected in the German SOEP in 2007 capturing net worth (from real estate, financial and business assets, tangible assets after deducting any outstanding mortgages and consumer debts) we expect that the consideration of our social security wealth estimate provides a well suited measure for a less biased analysis of inequality of an *extended wealth measure* in Germany.

3.3 Matching Variables, Slice Variables, Matching Populations

With record linkage being infeasible, we use statistical techniques to combine information from two or more data sets. For this data combination to be successful, data sets need to share a set of common variables defined and measured in similar ways. The data combination

¹⁸ A personal pension account is conditional on having at least one event over the life-course that constitutes rights in the statutory pension insurance. These entitlements can result from gainful employment, but also from other pension-relevant events such as periods of childcare, long-term care or entitlements resulting from a pension splitting in case of a divorce.

¹⁹ In the remainder of this paper, the Statistic for *Pension Rights Splitting in Case of Divorce* will be called *Divorce Statistic*.

²⁰ Data come in the form of a case-statistic with each row representing one divorce. Individuals with multiple divorces can show up multiple times in the data. Data only covers those divorces with a splitting of pension rights following a marital split.

distinguishes slice and matching variables. *Slice variables* partition the data to only match individuals within certain predefined strata. We slice the data sets by sex, region and immigrant status.²¹ The partitioning avoids matches of individuals that are sufficiently dissimilar, especially if these groups are believed to differ in how they accumulate pension entitlements.

Matching variables are continuous. They serve the purpose to identify pairs that are statistically close to each other; ideally, these paired cases are *statistical twins* with respect to the matching variables. A distance metric is used to assess how close both records are. Pensionable earnings and pension-relevant transfer payments (e.g. unemployment benefits, etc.) mainly determine the individual's social security wealth. An aggregate measure summarizes all income that qualifies for the accumulation of pension rights (earnings, unemployment benefits, sickness allowances, etc.).²² The income measure enters the equation as a three-year moving average to smooth individual income histories (average income for the years 1984-1986; 1985-1987, ..., 2003-2005). For all 2007 SOEP respondents with incomplete income profiles, we impute missing information starting in 2007 and going backwards to 1984 for West Germany and 1991 for East Germany. The imputation makes maximum use of all available longitudinal income information since the respondent's initial participation in the SOEP.²³ To make information available for all respondents and all survey years, we impute missing income information also for the years prior to the first SOEP interview. For reasons of comparability, earnings are cut at the effective maximum contribution ceiling for each year.²⁴

A woman's fertility history is an additional piece of information which ought to be considered in the statistical matching as it determines the number of childcare credits that are credited to the woman's pension account. For the statistical matching, we include the total number of childcare credits as of 2007. Women receive one childcare credit for all children born before

²¹ To assure a sufficiently high number of observations, we distinguish six groups: Men in East Germany; Men in West Germany; Women in East Germany; Women in West Germany; Male Migrants; Female Migrants.

²² In both data sets earnings and income information is available for the years 1983 to 2007 for West Germany and 1991 to 2007 for East Germany.

²³ 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.

²⁴ 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 Bund 2009a). For previous years' incomes thresholds need to be adjusted accordingly with the respective year- and region-specific values.

1992 and three years for all children born thereafter.²⁵ In addition, various duration variables enter the computation of the distance metric. These measures reflect the number of years spent in different activities such as employment, unemployment, education, compulsory military or community service (only for men), as well as long-term care giving. Finally, the statistical matching includes the age of the respondent as of 2007.

Given the population representativeness of the SOEP, we match VSKT information to SOEP data. Hence, the SOEP is the *recipient file*. This matching direction enables us to keep the population representative demographic structure of the SOEP population and eliminate potential issues of selectivity inherent to VSKT data. Non-validated pension accounts are a possible threat to the quality of matches. At some point during working life, the German Social Security Administration asks each person to validate the information stored in his or her pension account.²⁶ The account validation completes the individual record with all information missing but relevant for the calculation of benefit amounts and eligibility and corrects erroneous data.²⁷ Once validated, the individual pension records provide reliable data. We do not know, whether information is incomplete or flawed, but the likelihood of errors in non-validated accounts is higher than in validated accounts.²⁸ This paper is based on validated accounts only.²⁹ We also allow for sampling with replacement. This implies that one VSKT case can be the *statistical match* of several SOEP cases, whereas a certain share of VSKT cases is not matched at all.

3.4 Statistical Matching: Alternative Approaches

There are different strategies to complement SOEP data with VSKT information on social security wealth. To test which strategy fits the data at hand best, we compare the performance of four imputation and statistical matching approaches. Imputation is the generic term for a group of techniques typically used to handle item-nonresponse in survey data. The general idea is that

²⁵ Each credit is worth one earnings point, equivalent to the average earnings of all contributors in the respective year.

²⁶ This procedure is called account validation (*Kontenklärung*).

²⁷ Typically, periods spent in school or university, as well as periods of child-rearing are missing in the data, since those are not submitted electronically to the GSSA. In the course of the account validation, these periods are credited to the individual's earnings record.

²⁸ Exceptions are accounts that were validated several years ago. The period between the validation and the sampling lacks proper validation.

²⁹ Restricting analyses to validated accounts reduces the number of observations from 568,586 to 336,069 cases. Non-validated accounts are selective in that individuals are significantly younger, because they had no reason to validate their information yet. Migrants are largely overrepresented in non-validated accounts. Migrants who worked in Germany for some years and then left the country account for these overcounts. It is important to exclude these accounts, because they would be wrongfully matched. Also, most women with non-validated accounts have not yet reported the birth of their children. Childcare credits are a crucial variable in the statistical matching of women, which requires validated information (Rasner 2010).

incomplete records with missing data points are imputed applying different statistical models.

*Simple hotdeck imputation*³⁰ completes records with missing data points with values from statistically similar, but complete records. The imputation is carried out within predefined matching strata. For example, missing social security wealth for a West German woman in the SOEP is completed using observed information for a West German woman in the VSKT. Within those groups, the missing values are imputed by random assignment. This specification of a hotdeck imputation does not make effective use of the continuous matching variables identified above.

The *regression based imputation* estimates multivariate OLS regression models for each matching strata based on all VSKT observations. The individual's social security wealth is a linear function of the matching variables: pensionable income 1983-2007, age, number of children, years in employment, school, etc. Based on the estimated regression coefficients, we perform out-of-sample predictions of social security wealth, imputing the respective value for all SOEP observations. To mitigate the typical *regression to the mean effect* inherent in the prediction (in our case the social security wealth), residuals³¹ are randomly assigned and added to the respective predictions (Copas 1997). With this technique we are able to preserve the variance of the distribution, particularly improving the prediction of very small and very large values.

Univariate imputation sampling (UVIS) is a combination of parametric and non-parametric techniques to impute a single variable with missing values based on predictive mean matching. In a first step, *UVIS* makes use of a parametric model (OLS regression) that describes the individual's social security wealth 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 to each incomplete observation. Similar to the regression based imputation residuals are randomly assigned and added to the prediction to capture the variance of the distribution adequately.

Statistical matching provides alternative means to integrate information on social security wealth in SOEP data. This group of techniques aims at finding statistically similar observations in both data sets. Here we make use of *Mahalanobis distance matching*.³² This procedure is frequently

³⁰ Typically, both records - complete and incomplete - are part of the same data set which is the preferred data setup for the so-called *hotdeck imputation*. In our application social security wealth is missing for all SOEP records, but observed for all VSKT cases. This corresponds to the set-up for so-called *colddeck imputation* routines. We pretend that the underlying data coming from SOEP and VSKT belong to the same data set and therefore apply hotdeck imputation.

³¹ The residuals equal the difference between the actual social security wealth in the VSKT population and their respective prediction based on the regression results.

³² *Propensity score matching (PSM)* is a very popular statistical matching technique, commonly applied to the estimation

used in *cluster analysis*. For each observation x_i in the SOEP, a Mahalanobis distance d_{ij} to each observation x_j in the VSKT is calculated based on a vector of observed matching variables p . The *statistical donor* minimizes the distance between the SOEP respondent and the VSKT observation. Unlike the *Euclidean distance*, the Mahalanobis score incorporates both correlations between matching variables and differences in variances. First, this implies that highly correlated matching variables don't enter the computation of the *Mahalanobis distance* with the same weight. This property is particularly useful in our application, since the individual's annual income at time t is likely to be highly correlated with the annual income at time $t+1$.³³ Further, the Mahalanobis distance controls for differences in variances of the considered matching variables.

A systematic bias in our results requires an additional match that accounts for the splitting of pension rights between ex-spouses that is carried out in the statutory pension insurance. For each divorce, pension rights accrued by husband and wife during their marriage are summed up and then split in half right upon divorce.³⁴ The splitting cannot be adequately considered in the matching of SOEP and VSKT. This inadequacy results out of the fact that at the time SOEP respondents are asked to report their monthly pension benefit, it is impossible to tell entitlements from employment (or other individual pension relevant circumstances) and those resulting from the pension splitting apart. Information available in the *divorce statistics* correct for this bias resulting from the statistical matching of SOEP and VSKT that does not consider the *divorce effect*.³⁵ To estimate the divorce correction, we combine VSKT and divorce statistics using record linkage. The unique identifier is a combination of exact amount of the premium (*bonus*) or deduction (*malus*) from pension splitting, gender, region, and age. Based on the linked data, we estimate group-wise linear regression models that provide us with the divorce correction for the SOEP data. *Figure 2* illustrates the statistical matching process.

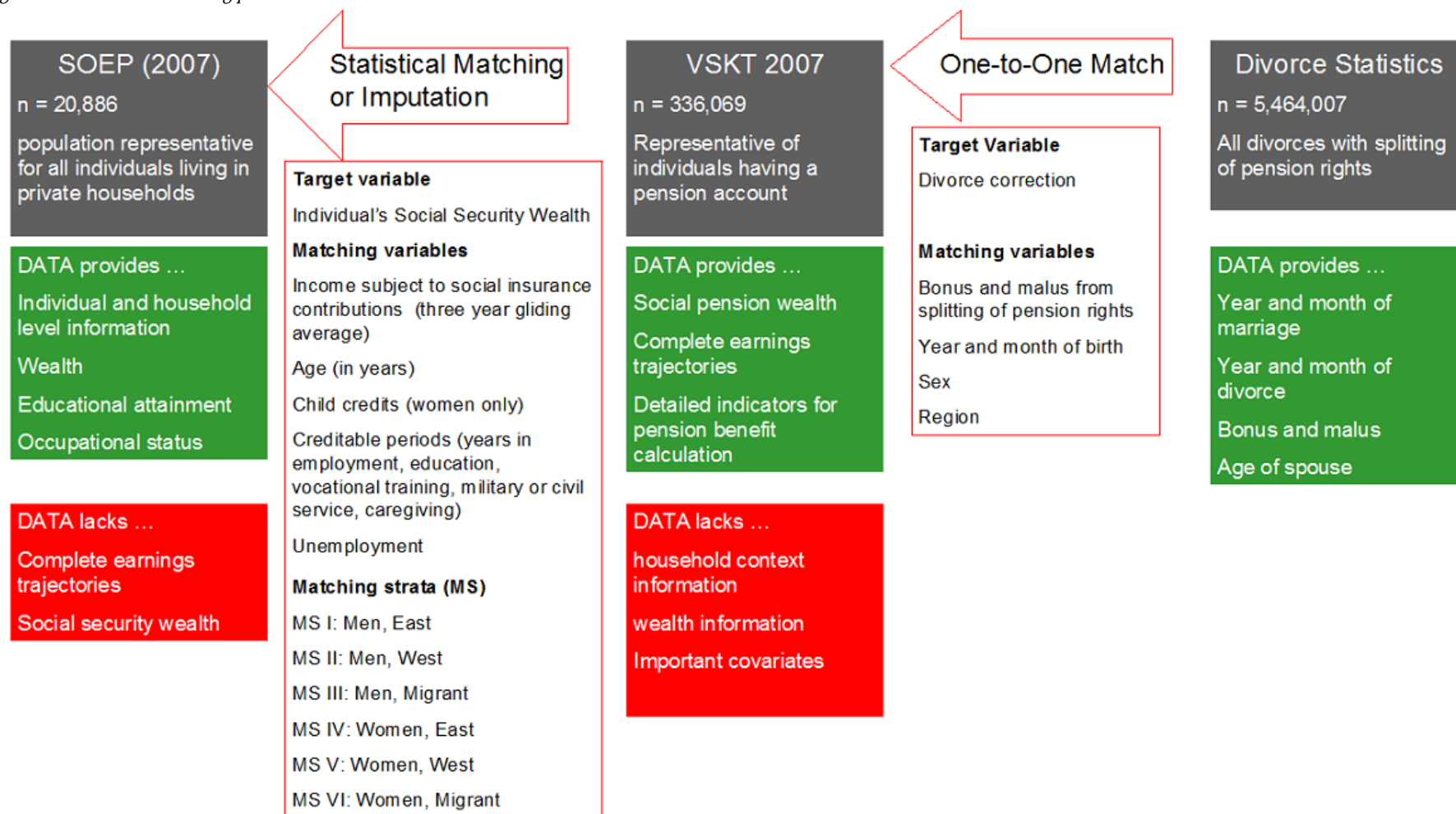
of treatment effects in evaluation studies. PSM serves as a strategy to correct for selection bias that results out of systematic differences between two samples. The statistical matching is carried out using an estimated propensity score that is the conditional probability of assignment to a treatment given a vector of covariates. The propensity score is the predicted value of a (*logit* or *probit*) regression that indicates the probability of receiving a treatment. Applied to SOEP and VSKT data, there is no reason to expect systematic differences in the two sample populations. Cases cannot be assigned to treatment and control group in a meaningful way. Therefore, *PSM* is not considered in the robustness tests performed in the following section. For an application of *PSM* using SOEP and VSKT see Geyer and Steiner (2010).

³³ If the Mahalanobis distance measure did not control for the correlation between the various matching variables, the matching process in our application would be completely driven by income variables as those provide the majority of matching variables.

³⁴ The partner who earned higher pension rights transfers half of the difference in entitlements to his/her former spouse. 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 place right upon divorce, bonuses and maluses remain unaffected by remarriage of either ex-partner.

³⁵ The first matching fails to take divorce into consideration.

Figure 2 Statistical matching process - SOEP, VSKT and Divorce statistics



Source: Own Illustration

4 Statistical Matching of Retirees

4.1 Evaluation Criteria and Sample Specification

Since there is no theoretical reason to pick one imputation or statistical matching technique over its alternatives, the decision process is solely based on empirical reasoning. To validate which approach fits the data best, we run robustness tests for the group of retirees in both data sets. Only for this segment of the population, there is an effective control for the quality of the match. Survey respondents provide the presumably *true* pension benefit (henceforth *observed benefit*) that allows for the comparison with the *simulated* benefit resulting out of the four matching and imputation techniques.³⁹ We assess the validity of each approach using three criteria: 1) correlation coefficient of the *observed* and *matched* public pension benefit; 2) absolute average difference between *observed* and *matched* benefit; and 3) for the overall fit graphical representation (*kernel density plots*) of the difference between *observed* and *matched* benefit. Each criterion is evaluated for the total population and within each of the six matching strata.⁴⁰

For the robustness tests, both datasets are restricted to retired individuals aged 60 to 67 who actually report an old-age pension benefit from the statutory pension scheme. Individuals receiving disability benefits are omitted from the analysis because of significant differences in eligibility rules and pension benefit calculation. 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. Potential SOEP matching partners come from Samples A, B, and C. Ideally, these cases are observed from the very beginning of the panel study in 1984 in West Germany and 1990 in East Germany, respectively, therefore providing the maximum information for the backward imputation of missing income data. The specification of the matching populations reduces the sample size for this validation exercise in both datasets to 34,353 VSKT and 659 SOEP observations.

4.2 Results

For the total population of retirees (all six matching strata) pairwise correlations between *observed* and *matched* public pension benefits are best for Mahalanobis matching. The correlation

³⁹ 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 response error

⁴⁰ We conducted additional robustness tests to assess the stability of results by drawing five random samples with replacement and five disjoint random samples without replacement. Results are available upon request.

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. Hotdeck imputation clearly lags behind ($r_{HOT}=0.22$). The lack of association between observed and matched benefit for hotdeck 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 within the six strata, without taking 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.

The absolute mean distance d indicates how far off the matched benefit is from the observed public pension benefit as reported in the survey. A small average distance and standard deviation are indicators for a good match. Considering the distance criterion for the total population, hotdeck 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 approaches. 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 1 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: Own calculations

In *Table 1*, 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). Mahalanobis matching works best for West German men ($d_{MAHA}=-34.7$) and male migrants ($d_{MAHA}=-90.7$). Hotdeck 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.

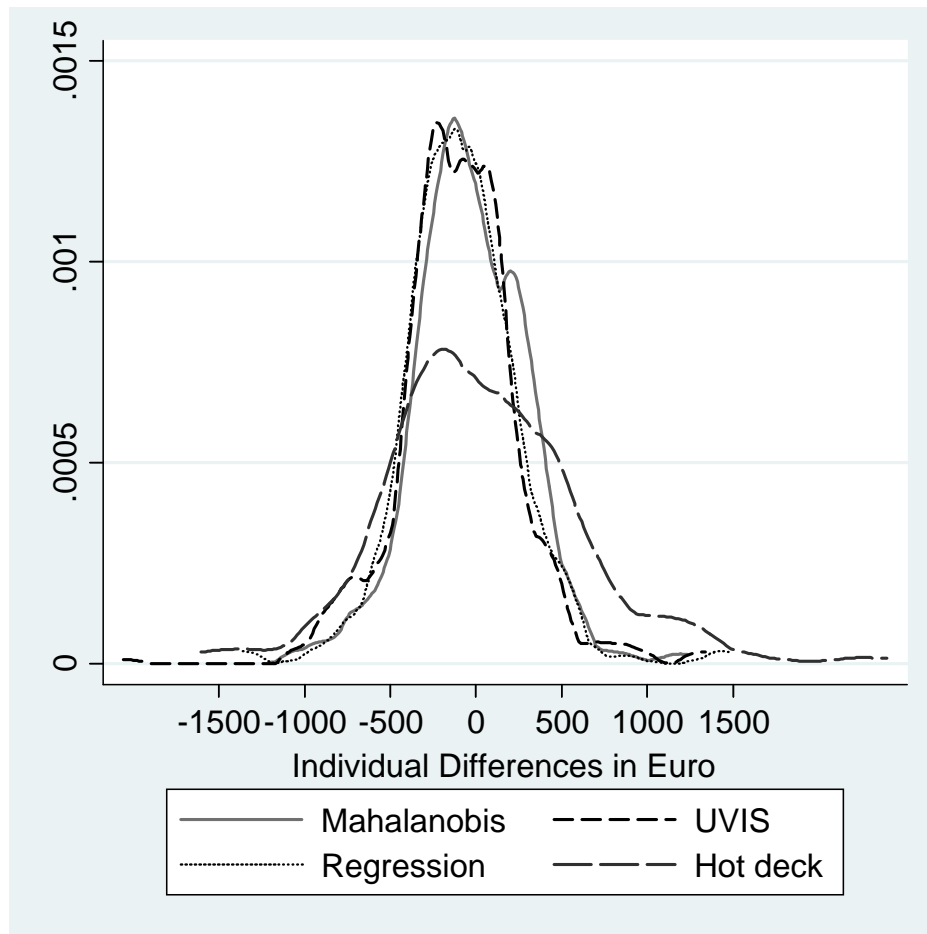
Despite comparatively small average distances for groups 4 to 6 under hotdeck, the standard deviation is by far the highest. With respect to the standard deviation, UVIS performs best for East German men as well as East and West German women, whereas Mahalanobis is better for West German men and male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

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 3* presents the kernel density plots for the total population.

The graphic representation underlines that hotdeck is not the appropriate imputation technique. The distribution of differences has a substantial standard deviation with very long tails to both sides. Despite its better performance compared to hotdeck, the UVIS distribution has no unambiguous peak. The distribution (red curve) appears to be much wider at the top with several smaller peaks. The kernel density plots for the regression approach and Mahalanobis come closest to the ideal. The distribution for Mahalanobis is centered around zero but shows a small bump at +250. The kernel density curve for the regression-based technique has no such bump, but the peak of the distribution is more spread out.⁴¹

⁴¹ For the group specific kernel density plots see Appendix A1.

Figure 3 Kernel Density Plots for Differences between Observed and Matched Benefit Information - Total Population



Source: VSKT 2007 and SOEP 2007; own illustration

4.3 Discussion

Mahalanobis distance matching is the best performing technique if we factor in all three criteria. Nonetheless, certain patterns require further explanation. First, there is a systematic negative bias in the absolute average difference between observed and matched benefit information. Second, it is obvious that there are significant between-group differences indicating that on average, the quality of matches is better for some groups than for others. According to *Appendix A2* the absolute average distances between reported and matched benefit information are negative for the total population as well as for most of the individual groups. This bias indicates that matched information is systematically higher than observed public pensions. The payment of insurance contributions for health and long-term care is one possible explanation for this bias. On the one hand, in VSKT data, we observe the *gross* public pension benefit (*Rentenbetrag*). From this gross benefit, the German Social Security Administration pays

contributions into the health and long-term care insurance and then transfers the remaining (net) public pension benefit (*Rentenzahlbetrag*) to the beneficiary (Deutsche Rentenversicherung Bund 2009b).⁴² On the other hand, it is most likely that retirees in the SOEP report their public pension benefit *net* of these contributions.

A special administrative feature of VSKT data might also contribute to the bias. For each observation in the VSKT a fictitious disability pension benefit is calculated. This step serves the purpose of calculating a present value of social security wealth for not (yet) retired persons under consideration of current pension law. Unfortunately, a fictitious disability pension is also calculated for individuals who already draw benefits (instead of providing the actual public pension benefit). Differences in eligibility criteria and benefit calculation for disability and old-age retirees possibly cause the overestimation of approximated pension benefits in the VSKT, because it fails to account for actuarial adjustments in case of early retirement.

Concerning the obvious between-group differences (see *Table 1*) robustness tests yield some explanations: First, 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 positive income information. Second, the quality of the match depends on how good a predictor the observed information is for the individual's final public pension benefit.

Matches are particularly good for West German men. For this group, we observe the annual income for a maximum number of 23 years (1983 through 2005). Comprehensive income information for more than half of West German men's working life is a good predictor for their final public pension benefit.⁴³ Contrasting results for persons having 40 and more years of employment with individuals having less than 40 years of employment reinforces this line of argument. Observed and matched benefit information fit significantly better for individuals with continuous labor market attachment.

Matches are less good for East German men. Several reasons account for larger differences between observed and matched benefits: Data collection in East Germany started only after the fall of the wall. Therefore, the matching algorithm includes income for the years 1991 through 2005 - dismissing almost two thirds of elderly East German men's working life. These two thirds are crucial to determine East German men's final public pension benefit: First, because there are claims in special and additional pension schemes (*Sonder- und Zusatzversorgungssysteme*) for GDR

⁴² The beneficiary and the Statutory Pension Insurance equally share contributions to the health and long-term care insurance for retirees. The retiree's contribution amounts to 10.15 percent of her final public pension benefit (Deutsche Rentenversicherung Bund 2009a).

⁴³ We also tested a matching algorithm that was 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. Results are made available upon request.

elites that increase the final benefit significantly but cannot be controlled for in the data at hand.⁴⁴ Second, the labor markets East German men worked in before and after reunification had not that much in common. In the centrally planned economy, everybody had a job and unemployment was no issue, whatsoever. It was therefore favorable in terms of pension entitlements that West German pension law was extended to the time individuals lived in the former GDR (Börsch-Supan and Schnabel 1999). However, post-reunification experiences were quite different with age-earnings profiles being flat and returns to tenure and experience significantly lower when compared to West German men (Orlowski and Riphahn 2009).⁴⁵ Unemployment, for a significant share even long-term unemployment largely limited the ability to accumulate pension entitlements.⁴⁶ Taken together, these reasons explain why the observation of only the most recent years might not be the very best predictor for the final pension benefit of East German workers.

The quality of matches are better for East German women than for East German men. At first sight, this evidence is surprising as both men and women are observed for the very same period. However, East German women benefit to a lesser extent from the transfer of entitlements from special and additional pension schemes than men (Seitz 2003). The adverse effects these benefits have on the matching quality are therefore less strong for women.⁴⁷ Further, despite the supposedly existing gender equality in the former GDR, older cohorts of East German women never achieved gender parity (Trappe 2007). Women in our sample were disadvantaged with respect to earnings and occupations before, but also lacked proper employment opportunities after reunification.⁴⁸ The period of East German women's lives actually observed in the SOEP data is to a greater extent *representative* for the unobserved part, which improves the matching quality for this group. Additionally, East German women have by

⁴⁴ The compulsory social insurance pension scheme (*Sozialpflichtversicherung*) and the voluntary pension scheme (*freiwillige Zusatzrentenversicherung*) were the two building blocks of the pension system in the former GDR. A certain privileged segment of the population had access to special and additional pension schemes. Persons covered by the special pension scheme (*Sonderversorgungssystem*) - app. 300,000 individuals - accrued entitlements exclusively in this scheme. Persons covered by the additional pension scheme received benefits as a supplement to their regular old-age pension from the compulsory social insurance pension scheme (Seitz 2003). Benefits from both schemes can be considered rather generous.

⁴⁵ Orlowski and Riphahn suggest that for many East German men job-specific human capital was outdated and did not match requirements of job market in unified Germany (2009).

⁴⁶ Given that the *production* of VSKT 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 than is the case in West Germany.

⁴⁷ 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).

⁴⁸ Both, men and women in these intermediate birth cohorts were described as a lost generation that "was too young to retire and too old for investing in occupational retraining or expecting their employers to make such investments" (Trappe 2007, p. 124).

and large more homogeneous working careers. For these reasons, differences between reported and matched benefit information are smaller for East German women than for men.

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 following the years of childbearing and childrearing. This inactivity is particularly prevalent among older birth cohorts of women as a consequence of the strong *female caretaker/male breadwinner notion* the West German postwar welfare state promoted.⁴⁹ 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.⁵⁰

Following this line of argument it comes at 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 (*Vertragsrenten*) also account for large differences between observed and matched benefit information among migrants.⁵¹ For the matching, it is infeasible to separate benefits earned in Germany from benefits earned in other countries.

4.4 Taking Divorce into Consideration

The second match serves the purpose of taking the pension rights splitting into account, which could not be approximated in the SOEP. If the *divorce effect* was not considered, matches would be systematically biased for all divorcees understating the compensating effect arising from the pension splitting.⁵² The divorce correction by and large improved the quality of matches. Table 2 presents the absolute average distances (using *Mahalanobis distance*) with and without the divorce correction for the divorced population only.⁵³

⁴⁹ 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.

⁵⁰ The average difference between observed and predicted benefit for women with 40 and more years of employment is -8.98 Euro compared to -87.06 Euro for women with less than 40 years of employment.

⁵¹ Persons who worked and accrued pension rights in Germany and another country receive a *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*).

⁵² Considering divorce as an additional slice variable is infeasible due to the small numbers of observations.

⁵³ Because of the small number of divorced migrants in our sample, those results are not further discussed.

Table 2 The Effect of the Divorce Correction – Only Divorcees

Difference between true and simulated pension benefit (Euro)		With divorce Correction	Without divorce Correction	Effect of the divorce correction
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

Source: VSKT 2007 and SOEP 2007, own calculations

Migrants are omitted because of the small number of divorced migrants in the SOEP population

Because the initial match failed to take divorce appropriately into account, the matched benefits were either too large for most of the divorced men or too small for most of the divorced women. The divorce correction shifts results in the expected direction. Consequently, the absolute average distance changed from $d_{W/O \text{ Correction}} = -124.0$ to $d_{W. \text{ Correction}} = -50.9$ for East German men and from $d_{W/O \text{ Correction}} = -190.9$ to $d_{W. \text{ Correction}} = 3.4$ for West German men, respectively.⁵⁴ In turn, differences for women shifted in the other direction. The absolute average distance for East German women is $d_{W/O \text{ Correction}} = -8.5$ without and $d_{W. \text{ Correction}} = -60.7$ with the correction. For West German women, the difference shifts from $d_{W/O \text{ Correction}} = 117.5$ to $d_{W. \text{ Correction}} = -58.4$. For almost all groups, the standard deviation of differences between observed and matched benefit can be reduced applying the divorce correction and consequentially, the quality of matches for divorcees converges to the quality of non-divorced individuals.

The results outlined in Sections 4.2 and 4.4 justify the application of the double matching strategy to the working age population in order to obtain an augmented database for the analysis of economic well-being for the total population in Germany. The matching excludes retirees assuming that for this population of retirees the amount stated in the questionnaire is more reliable than any simulated benefit could possibly be. The matching samples for the working age population are specified in line with the criteria outlined in Section 4.1. We assume the matching

⁵⁴ 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.

quality to be better for the active population, i.e. those not yet retired, because the income data which is only available in the survey from 1984 onwards, covers a larger share of respondents' working lives, and thus, *ceteris paribus*, feeds more reliable estimates into the statistical matching process. Overall, the matching links 288,655 VSKT observations with 14,247 SOEP observations.

5 Wealth Inequality

In this section, we perform an analysis of wealth inequality. The measure of interest is total net worth, an aggregate measure obtained from the SOEP 2007 wealth module.⁵⁵ We extend the measure of net worth by taking the matched present value of pension entitlements accrued from statutory, company and private old-age schemes into account and analyze the impact on wealth inequality.

The calculation of total pension wealth differs by employment status (retired vs. non-retired) and occupational group (e.g. dependent employment, civil servant, etc.). For already retired individuals, we use benefit information respondents provide in the SOEP income section. For non-retired persons, we use accrued pension entitlements in the statutory pension scheme derived from the statistical matching (Section 4). This method of approximation is infeasible for special occupational schemes for farmers or the liberal professions (e.g. lawyers or doctors) and company pensions. Like dependent employees, survey respondents engaged in these professions are unable to report their current amount of entitlements in these schemes. Second, there is no data available to replicate the statistical matching to gather similar information on pension entitlements in these special occupational schemes. For non-retired individuals in professions not covered by the statutory pension scheme, entitlements are therefore largely underestimated. This underestimation does not apply to the civil servants, because their entitlements can be reliably approximated: 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.⁵⁶

5.1 Calculating the present value of pension entitlements

The wealth inequality analysis considers the present value of recurring future pension payments. The estimation of this present value requires the consideration of information on

⁵⁵ Total net worth is the sum of owner-occupied and other real estate holdings, financial assets, assets from life assurance policies and private pension schemes, building loan contracts, business assets, valuables, net of all outstanding consumer and mortgage debts.

⁵⁶ The three-year average of earnings are multiplied by factor 1.79375 and the number of service years. Therefore, the maximum replacement level after forty years of employment amounts to 71.75%. The assumption to use the current pay level leads to an underestimation of the future pension entitlement, given that subsequent increases in the pay level cannot be considered.

remaining life expectancy, the individual's retirement age, and the taxation of retirement income. The 2005/2007 life tables of the German Federal Statistical Office provide information on remaining life expectancy by sex and region (East and West Germany). In line with the 2007 Pension Reform Act (*Altersgrenzenanhebungsgesetz*), we take the gradually increasing retirement age into consideration. Starting with the 1949 birth cohort, the retirement age for full benefit rises by one month per calendar year (today 65, rising to 67). Differential taxation of retirement income by occupational group is another important factor that calls for inclusion. The information resulting from our statistical matching approach for members of the statutory pension scheme as well as from the approximation exercise for the civil servants based on their most recent earnings is the *gross* pension entitlement. However, these occupational groups also differ with respect to the tax-treatment of their retirement income. For example, life annuities and other benefits such as public pensions, agricultural old-age funds, or pension schemes organized by professional associations are not fully taxed yet. The taxable share of pension income depends on the year in which a person draws a pension for the first time and is gradually increasing (§22 of the German Income Tax Act [*Einkommensteuergesetz*]). For example, the taxable share for first-time pensioners in 2007 is as high as 54 percent. With each additional year, the share increases to 80 percent by the year 2020 and to 100 percent by 2040 for pensions paid for the first time. On the other hand civil servants' annuities are already fully taxed as of today. For each occupational group, the respective 2007 tax rate applies.⁵⁷

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.⁵⁸ For discounting purposes, we assume an interest rate of two percent.⁵⁹ Entitlements from private pension schemes require no present value calculation, as this wealth component is already covered by the SOEP questionnaire.

5.2 Individual net worth without pension entitlements

In a first step, we look at individual net worth alone, completely omitting the present value of pension entitlements. Aggregated net worth for individuals in private households in Germany amounts to about 5.9 trillion Euros in 2007 (*Table 3*). If this amount was evenly split, each adult

⁵⁷ By applying the current individual tax rate, a relatively high tax burden is assumed for active insured persons. The actual tax burden would, however, have to be simulated separately for each age cohort at the time of entry into retirement. Because of the recently introduced deferred taxation, however, such a simulation requires major assumptions about the future income situation of the persons in question, which clearly goes beyond the scope of this paper.

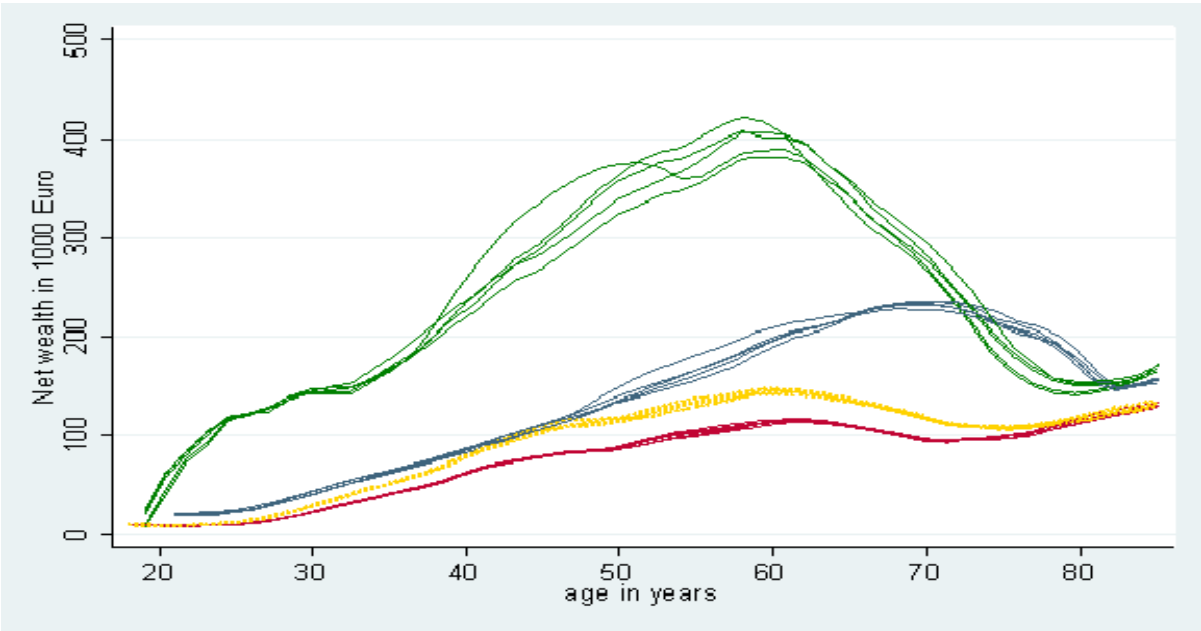
⁵⁸ Because of the pension reform of 2004 and the sustainability factor introduced with it, future pension entitlements will decline from birth cohort to birth cohort for the same number of payment points. For reasons of simplicity, however, this aspect is ignored here.

⁵⁹ In alternative specifications we vary this interest rate between one and three percent; the choice of interest rate influences, by definition, the amount of the present value but changes little in the basic relationships according to occupational groups as described here.

person would have about 83,000 Euro at his or her disposal. Comparing mean and median of the distribution gives a first indication for the degree of overall inequality. In fact, median wealth amounts to only 15,000 Euro. Thus, the mean exceeds the median by factor 5.5.

The standard life cycle hypothesis states that individuals save by cutting down on consumption during working life to finance their retirement (Modigliani 1988). Hence, net worth and age are strongly related and savings serve the central purpose of consumption smoothing over the life cycle. *Figure 4* confirms this age-pattern for Germany, but also shows pronounced differences in levels of life-cycle wealth accumulation by occupational status. The self-employed accumulate the highest levels of net worth. At age 60, self-employed have a total net worth of almost 400,000 Euros, which is twice the amount of civil servants and 3.5 times more than net worth of dependent employees insured in the statutory pension scheme. These marked differences also hold for dependent employees with continuous occupational careers with no or few spells of unemployment ending up with an average net worth equal to 140,000 Euros at the age of 60 years, i.e., somewhat before retirement.

Figure 4 Individual net worth by status group, Germany 2007



— Self-employed, — civil servants, — Members of the statutory public pension scheme (GRV), — Persons with continuous GRV insurance

Note: for each group the five lines represent the individual net worth derived from multiply imputed information. Source: SOEP 2007

5.3 Individual net worth considering pension entitlements

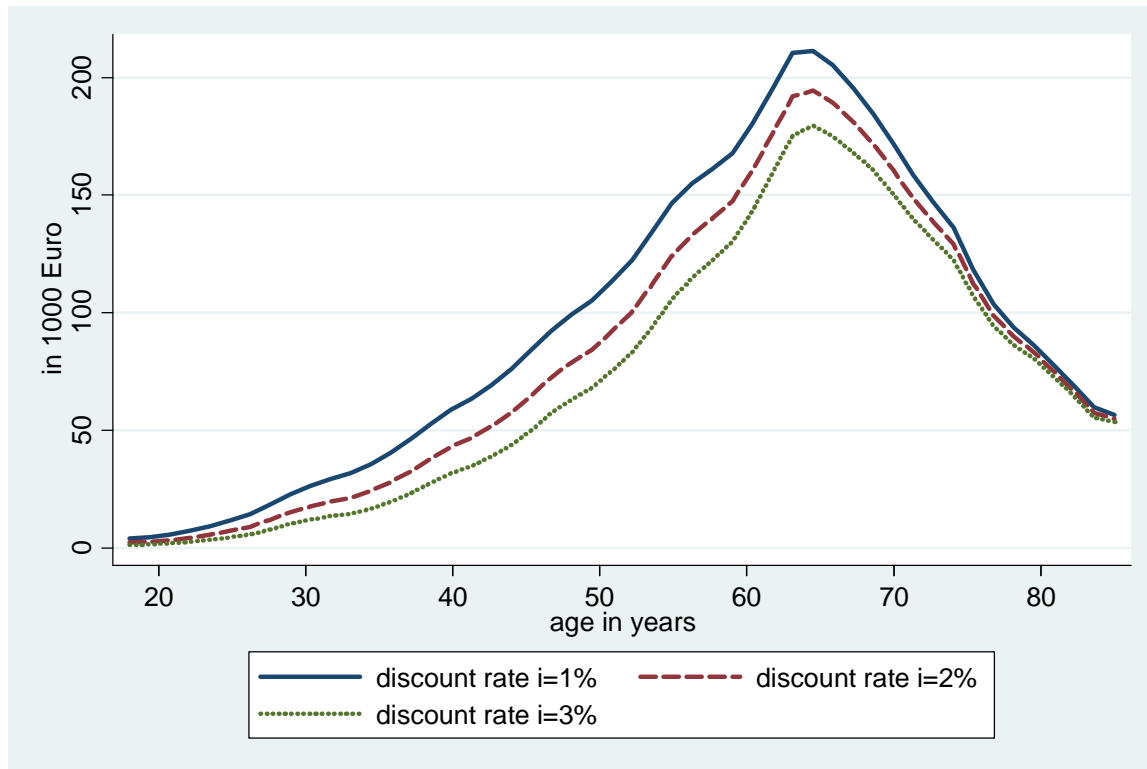
In the following analyses we add the present value of pension wealth to the net worth measure used above. A first look at the shape of the age-profile of pension entitlements meets our expectation: Entitlements grow with increasing age up to the time around retirement, with the slope in the second phase of working life being somewhat steeper (*Figure 5*). The shape of the present value of pension entitlements is similar to age-earning profiles, because paid contributions are a fixed share of earnings up to some maximum threshold (*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.⁶⁰

Figure 5 also illustrates that the level of the discount rate matters: With a discount rate of two percent the net present value of all pension entitlements 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 like a somewhat normative decision, 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 5*, 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.

⁶⁰ With the *statistical death* of an individual, the present value of entitlements equals zero.

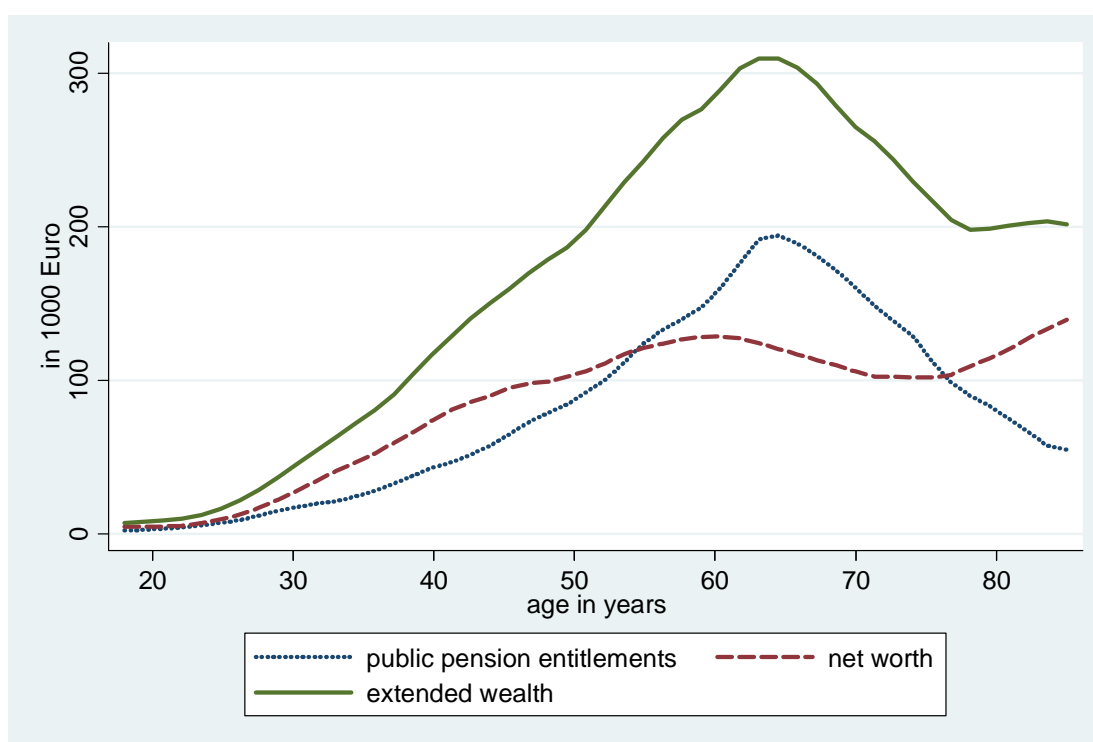
Figure 5 Present value of pension wealth entitlements by age for different discount rates, Germany 2007



Source: SOEP 2007

What happens to the distribution of total net worth if we incorporate the present value of pension entitlements? *Figure 6* 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 might reflect 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 6 Net worth, present value of pension wealth entitlements and extended wealth by age, Germany 2007



Source: SOEP 2007; Public pension wealth using a discount rate of 2%

5.4 Wealth inequality and pension entitlements

For 2007, *Table 3* indicates that the Gini-coefficient for net worth is 0.80. The top 20 percent of the adult population hold almost 80% of total net worth, whereas the three bottom quintiles own less than 3% of total net worth. About one twelfth of the adult population has negative net worth (not explicitly shown in *Table 3*), indicating that liabilities exceed gross wealth. On the other hand, the Gini-coefficient for public pension wealth is only 0.566, thus significantly lower than the one we observe for net worth. The extended measure of net worth including pension wealth brings the Gini-coefficient down to around 0.6, decreasing inequality by one quarter when moving from the distribution of net worth to the one using our amended measure.

Results for the *Half-squared coefficient of variation* (HSCV) point in the same direction as those using the Gini-coefficient, but they are even more pronounced. For net worth, the HSCV is exceptionally high with 6.5, whereas the corresponding value for public pension entitlements is only 0.7. Accordingly, the HSCV for the extended net worth measure is “reduced” to 2.0, decreasing inequality by more than two thirds. Compared to the Gini coefficient, the reduction is more pronounced for the HSCV because of the top-sensitivity of this indicator and the shape of the upper tail of the wealth distribution being much wider. According to the P90:P50 percentile ratio, the lowest amount of wealth found in the top decile is 14 times higher than the median of

the net worth distribution. For the extended wealth measure, the respective ratio is only 3.8, again indicating as significant reduction decrease in inequality (by 73%).

Table 3 Net worth and public pension wealth in Germany¹, 2007

	Net Worth (€)	SSW² (€)	Extended Wealth (€)	Change (%)
Basic Statistics	(1)	(2)	(3)	[(1) / (3)]
- Mean	83,077	78,479	161,556	94.5
- Median	14,751	46,680	94,675	541.8
- Sum in trillion Euro	5.908	5.581	11.489	94.5
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
Wealth shares (%)				
- lowest Quintile	-1.5	0.9	0.4	126.7
- 2 nd Q.	0.4	5.2	4.5	1025.0
- 3 rd Q.	3.9	12	11.8	202.6
- 4 th Q.	17.3	24.1	22.4	29.5
- highest Quintile	79.9	57.7	60.9	-23.8
Population with zero or negative wealth (%)	28.1	4.5	3.3	-88.3

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 dependants.

³ Lowest value for the top ten percent in the wealth distribution in relation to the median (50 percent).

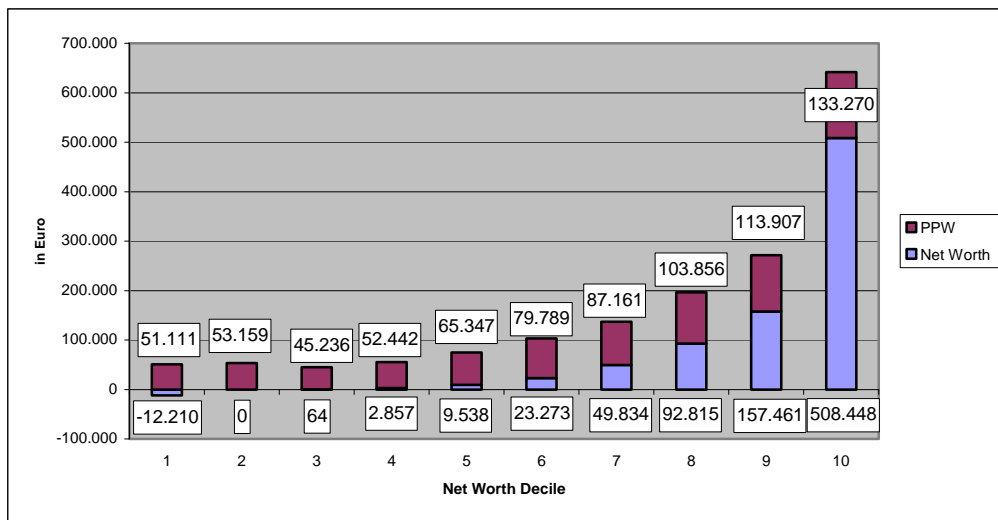
Source: SOEP 2007.

All in all, our empirical evidence clearly suggests that pension entitlements are less concentrated than net worth with various factors contributing to this difference. First, almost every person living in Germany accumulates entitlements in at least one pension scheme, whereas not every person holds positive net worth. Second, pension entitlements can only take on positive values – at least it cannot become negative according to our measurement. Therefore, even the bottom quintile of the adult population has a positive present value of pension wealth, which is not true for the bottom quintile of net worth. According to *Table 3*, the share of individuals holding zero or negative net worth is 28%, whereas less than 5% - mostly very young individuals – have not yet accrued any public pension entitlements. In addition, pension benefits are capped at some maximum amount, which by definition limits inequality.

Figure 7 restates the equalizing effect pension wealth has on the extended wealth measure, comparing the present value of pension wealth by deciles of net worth. In the graphical representation, the lower part of each bar depicts the average net worth ranging from minus 12,000 Euros in the bottom decile to more than 500,000 Euros in the top decile. In this illustration, we hold the deciles of net worth fixed and add the corresponding present value of pension wealth, which yields the average extended wealth by net worth deciles. For the lower half of the distribution, variation in the mean of total pension wealth is small (between 45,000 and 65,000 Euros). Nevertheless, for this bottom half of the distribution, pension wealth is obviously

far more important than total net worth.

Figure 7 Mean net worth and mean present value of pension wealth entitlements by net worth decile, Germany 2007

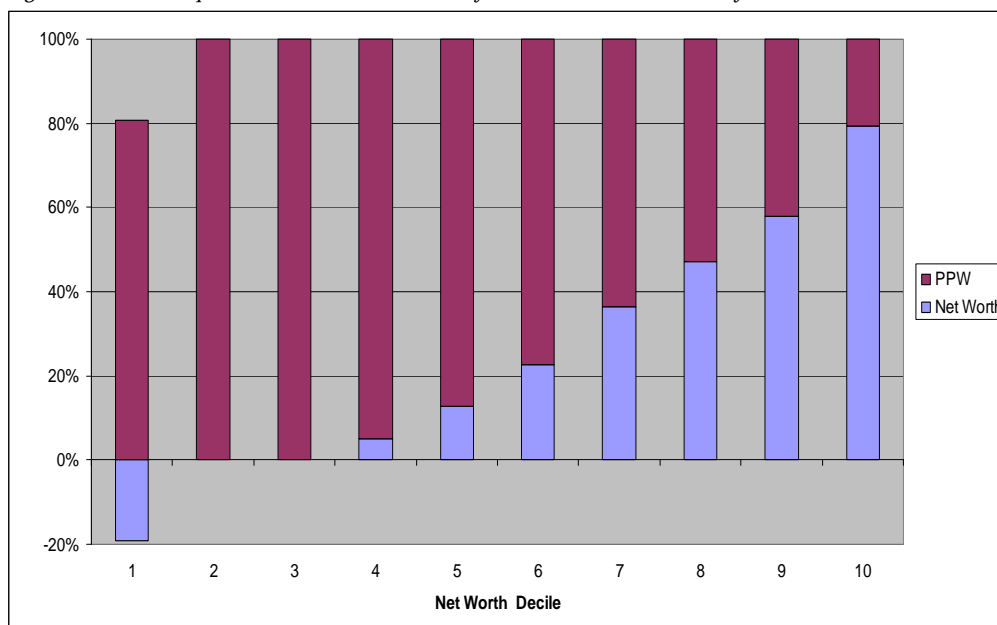


PPW: Public pension wealth using a discount rate of 2%
Source: SOEP 2007

Pension entitlements gradually increase for the upper half of the net worth distribution. This increase can be seen as a function of the principle of equivalence according to which high income earners also pay higher contributions (in absolute terms) yielding higher pension entitlements. At the same time, we can assume civil servants with rather favourable pension entitlements, especially from the executive class, to be overrepresented in the top wealth deciles. But this increase in PPW is not at the same pace as the one for net worth. For the top decile of net worth, pension wealth amounts to approximately 130,000 Euros compared to the corresponding net worth of more than 500,000 Euros. First, this finding is influenced by a change in the composition of the population as we find an increasing share of self-employed with rather little PPW when moving up the wealth hierarchy. Second, the upper contribution ceiling in the statutory pension insurance top-codes the amount of contributions to be paid, and accordingly, the maximum pension benefit received.

Figure 8 depicts the relative composition of extended wealth by deciles of net worth. For the bottom of the distribution, pension wealth almost accounts for the total extended net worth. For the upper half of the distribution, the relative importance of pension entitlements in the wealth portfolio gradually decreases with increasing extended net worth. Pension wealth accounts for 80% of extended net worth in the 6th decile, 50% in the 8th decile and for only 20% in the top decile.

Figure 8 The composition of extended wealth by net worth decile, Germany 2007



PPW: Public pension wealth using a discount rate of 2%
 Source: SOEP 2007

5.5 Net worth and pension entitlements by occupational status

The individual's employment career is an important proxy for the accumulation of wealth. Among other factors, the occupational status determines a person's income level and also his or her ability to save.⁶¹ Table 4 provides evidence on how occupational status relates to wealth holdings in Germany. In 2007, unskilled, semi-skilled workers and salaried employees (without vocational training) held roughly 34,000 Euros in financial and material assets. In comparison, skilled workers such as foremen or masters came close to 70,000 Euros in assets. Net worth for employees with management responsibilities amounts to more than 120,000 Euros.

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 Euros. Civil servants in the sub-clerical or clerical service accumulate substantially less (67,000 Euros), but still more than skilled workers and salaried employees. The self-employed have the highest level of financial and material wealth. Business assets largely account for this accumulation. The more employees work in the business of a self-employed SOEP respondent, the higher total net worth: For those with no employees individual net worth equals nearly 170,000 Euros, self-employed with one to nine employees more than double this amount (about

⁶¹ In what follows, persons are differentiated by occupational position as measured in the SOEP survey year 2007. One person may nevertheless have been active earlier in another occupational position, which may affect both the amount of financial and material wealth and old-age pension entitlements.

350,000 Euros). Finally, self-employed with ten or more employees have an individual net worth of more than 1.1 million Euros, clearly leading the wealth hierarchy.

Trainees and apprentices are at the early stages of their occupational career. At a median age of 21 years, they had little time to accrue much net worth or pension entitlements. Wealth holdings of the unemployed are also below average with about 15,000 Euros. In line with the standard life-cycle model of saving, the elderly have above average net worth. This is especially true for retired civil servants with a measure of net worth of nearly 190,000 Euros. In comparison, pensioners in the statutory public pension scheme have net worth of less than 100,000 Euros at their command. Civil servants are at an advantage in the accumulation of wealth not only because of the higher educational attainment of this group, but also because they do not have to pay a similarly high share of contributions into their pension scheme as dependent employees. Civil servants only pay an implicit contribution to their tax-financed old-age provision, which *ceteris paribus* allows for greater savings.

As pointed out before, there are marked differences in old-age pension schemes in terms of organization and the provision of benefits by occupational group. *Table 4* illustrates these differences in the present value of pension entitlements. Among those individuals still in the workforce, the relative position of civil servants benefits most from the consideration of pension entitlements. For low and medium level civil servants, pension entitlements amount to 92,000 Euros. For high level civil servants (executive and administrative class) these entitlements are even higher (almost 150,000 Euros). In fact, their accumulated pension wealth nearly doubles their net worth. In contrast, dependent employees do not benefit to the same extent from the inclusion of the present value of pension entitlements. For the various groups of dependent employees, total pension wealth makes up rather similarly for 54,000 Euros to 75,000 Euros. Even the pension wealth of currently unemployed individuals is quite similar (52,000 Euros). 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. However, with the most recent labor market reforms in 2005 in Germany, pension entitlements for long-term unemployment have been drastically reduced. Thus, old-age poverty might be on the rise for segments of future retirees. The likelihood of becoming poor is even higher for this group, because they do not have much command over any relevant net worth (in financial and material assets).

Table 4 Net worth and public pension wealth by occupational status in Germany¹, 2007

	Occupational Group	Standard Net Worth	Present Value of SSW ³	Extended Wealth	Relative Change	Age in years	Share Female
		in Euro			in %	Median	in %
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 ²	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.

2 Including family members helping out

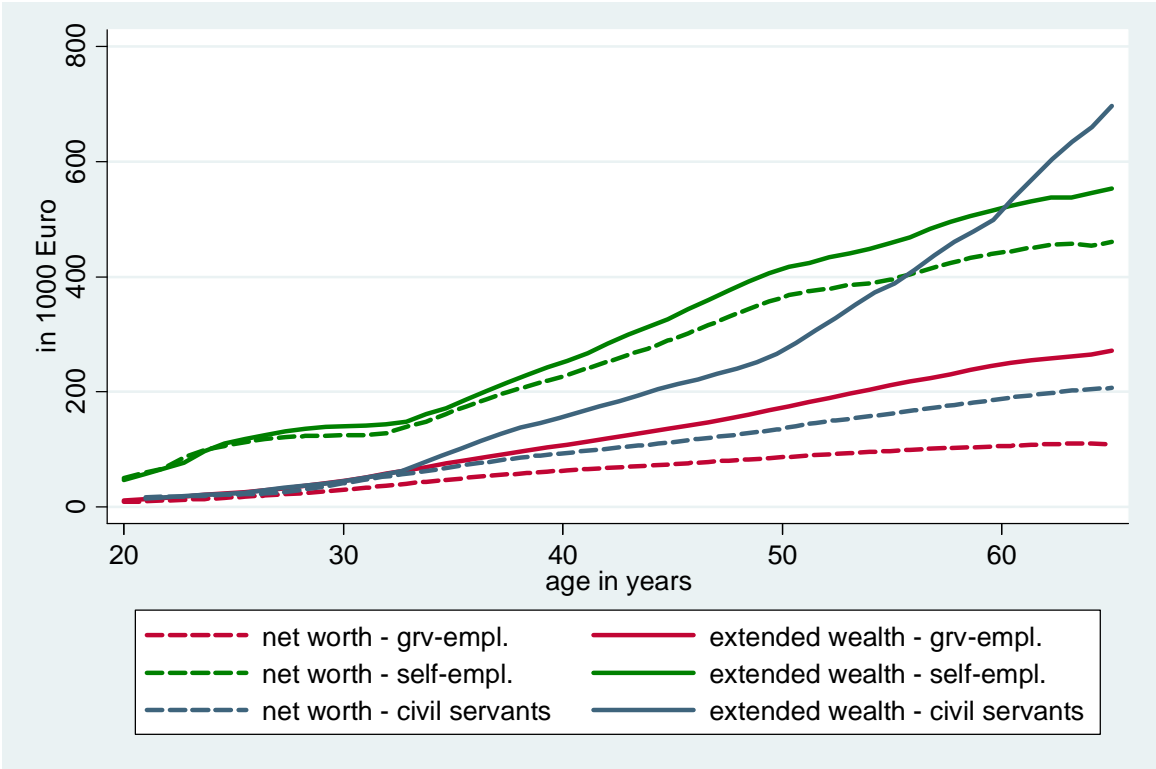
3 With a discount rate of two percent, without provision for dependants.

Source: SOEP 2007.

For the self-employed the respective figures vary on a somewhat lower level compared to dependent employees (between 35,000 Euros and 56,000 Euros). 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 consideration of the present value of pension wealth in the extended wealth measure clearly improves the position of civil servants relative to the self-employed. Nevertheless, the self-employed by and large remain on top of the wealth distribution.

Focussing only on working-age individuals, the age profiles of net worth and extended wealth for self-employed, civil servants, and dependent employees in the statutory pension scheme support these findings (Figure 9). The self-employed are on top of the net worth distribution throughout their working life (20 to 65 years). However, the profitable old-age provision for civil servants gives them a lead in the extended wealth distribution after age 60.

Figure 9 The relevance of public pension wealth by occupational status in Germany¹, 2007



¹ Population of employable age (20-65 years). Source: SOEP 2007.

Returning to the results on the basis of the entire population as presented in Table 4 above, the elderly benefit substantially from the consideration of pension entitlements in the extended wealth measure. Each additional year reduces the present value of pension wealth, because we take their remaining life expectancy into account. Keeping this in mind, we compare the pension wealth of individuals who collect benefits from the statutory pension scheme

(including potential occupational pensions), with retired civil servants and self-employed. Retirees in the statutory scheme are entitled to a pension wealth of almost 130,000 Euros. The favorable position of civil servants improves further with total pension entitlements of more than 310,000 Euros. Their total net worth including pension entitlements sum up to more than 500,000 Euros in 2007. Thus, retired civil servants have more *extended* wealth at their disposal than self-employed with up to nine employees (roughly 390,000 Euros).

Various factors contribute to this result: Apparently, men are over represented in the group of (retired) civil servants, an indicator for continuous full-time working careers. In turn, women are over represented among retirees in the statutory pension insurance. This finding reflects both the higher female life expectancy and the female surplus in the oldest cohorts as induced by World War II. Above and beyond such demographic factors, there are two additional institutional characteristics, which explain the pronounced difference between pensioners in the statutory pension scheme and retired civil servants: First, most civil servants enjoyed a continuous employment history without any unemployment spells. Other things being equal, such interruptions reduce the amount of contributions paid and accordingly, the level of benefits from the public pension insurance. Second, the pension scheme for civil servants is a final salary scheme and their earnings typically peak during the final years of service, whereas benefits in the statutory pension scheme take the entire wage history into account.⁶² Taken together, these aspects emphasize that the civil servants pension scheme is by far more generous than the statutory pension scheme.

6 Conclusion

Entitlements from old-age pension schemes – statutory, company, and private – represent a considerable source of wealth. However, wealth inequality analyses have so far failed to adequately take these entitlements into account because of data limitations. This paper takes on this challenge by using a statistical matching technique. We successfully link individual record survey data from the German Socio-Economic Panel (SOEP) – containing information on individual net worth – to detailed administrative data on pension entitlements from the statutory pension insurance in Germany as of 2007. We calculate the present value of the future flow of pension payments arising from the entitlements accrued so far: Assuming a discount rate of 2% yields an overall amount of individual pension entitlements (not including entitlements for survivors) of roughly 5.6 trillion Euros or - on average - 78,500 Euros per adult. When this is combined with individuals' net worth held in financial and material assets, which amount to an

⁶² In addition, any company pensions that might accrue to employees subject to social insurance generally represent a voluntary benefit from the employer.

average of 83,000 Euros, the result is a more comprehensive measure of extended wealth of more than 160,000 Euros. This extended measure of wealth shows considerably less inequality (a reduction in the Gini coefficient by about one quarter) than one sees from traditional analyses, which refer only to financial and material assets. This reduction in inequality is mainly the result of the widespread existence of entitlements in various old-age pension schemes for nearly all adults in Germany.

We also find marked differences in levels of pension entitlements across occupational groups. With respect to their position in the wealth hierarchy civil servants appear to profit most from the additional consideration of pension wealth, whereas the wealth position of self-employed becomes somewhat less favourable as they tend to typically invest in financial and material assets for old age and thus hold rather low public pension entitlements.

Future research in this area may discuss the appropriate way to determine the present value of pension entitlements including the proper definition of the discount rate, the projection of the development of future pension adjustments and the consideration of selective mortality, thus taking into account that high earners typically live longer than low-income groups. Furthermore, it is necessary to discuss whether and how to take liabilities to the old-age pension schemes into account. In a pay-as-you-go pension scheme, there is an implicit liability that starts with birth and persists approximately up to the middle of the working life. Adequately dealing with this phenomenon would also bring in the consideration of liabilities outside the private household sector, as employers and the state both have to contribute considerable degree to overall pension expenditures.⁶³

Summing up, this paper gives clear evidence that neglecting public pension wealth in wealth analyses yields massive bias with respect to level, inequality and socio-economic variation. These findings, however, may not be specific to the German case but most likely hold—to a varying degree—across most industrialized countries. Differences in welfare regimes and the emphasis these regimes put on the magnitude and generosity of publicly provided old-age insurance systems call for the consideration of pension entitlements in comparative wealth studies, for example the analyses based on the Luxembourg Wealth Study (LWS, see Sierminska et al. 2006). While net worth can be successfully collected by means of survey data, the knowledge of pension entitlements on behalf of respondents is often limited. This shortcoming can be solved ideally by the collection of the Social Security Identifier (or equivalent information) and informed consent as to allow for record linkage of the microdata at hand with the relevant

⁶³ A precise quantification of the liabilities to old-age pension systems is complex. It requires, among other things, simulating cohorts as yet unborn, all of whom would already have negative wealth by the time they are born. Here we refrain from any liabilities, i.e. future contributions, to old-age pension schemes.

administrative data. As this, however, may pose a significant threat to the respondents' willingness for (further) participation in (panel) surveys, the statistical matching approach presented in this paper – considering all relevant information in the choice of the matching variables – can surely be interpreted as the best available solution to develop a comprehensive wealth measure from survey and administrative sources.

7 References

- Bonin, H. (2009). "15 Years of Pension Reform in Germany: Old Success and New Threats." The Geneva Papers 34: 548-560.
- Börsch-Supan, A. and R. Schnabel (1999). Social Security and Retirement in Germany. Social Security and Retirement Around the World. J. Gruber and D. A. Wise. Chicago, University of Chicago Press: 135-180.
- Bundesministerium für Arbeit und Soziales (2008). Ergänztender Bericht der Bundesregierung zum Rentenversicherungsbericht 2008 gemäß §154 Abs. 2 SGB VI: 239.
- Christensen, K., G. Doblhammer, R. Rau and J. 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.
- Deutsche Rentenversicherung Bund, Ed. (2009a). Rentenversicherung in Zeitreihen. Berlin.
- Deutsche Rentenversicherung Bund (2009b). Rentenzugang 2008. Berlin, Deutsche Rentenversicherung Bund.
- Frick, J. R., M. M. Grabka and J. Marcus (2007). Editing and Multiple Imputation of Item-Non-Response in the 2002 Wealth Module of the German Socio-Economic Panel (SOEP). SOEPPapers 18. Berlin, Deutsches Institut für Wirtschaftsforschung (DIW).
- Frick, J. R. and B. Headey (2009). "Living Standards in Retirement: Accepted International Comparisons are Misleading." Schmollers Jahrbuch 129(2): 309-319.
- Geyer, J. and V. Steiner (2010). Public Pensions, Changing Employment Patterns, and the Impact of Pension Reforms Across Birth Cohorts - A Microsimulation Analysis for Germany. SOEPPapers 276. Berlin, Deutsches Institut für Wirtschaftsforschung.
- Gruber, J. and D. A. Wise, Eds. (1999). Social Security and Retirement Around the World. Chicago, Chicago University Press.
- Himmelreicher, R. (2005). Analysepotenzial des Scientific Use Files Versichertenrentenzugang. Forschungsrelevante Daten der Rentenversicherung: Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg. D. R. Bund. Berlin: 38-92.
- Jenkins, S. P., L. Cappellari, P. Lynn, A. Jäckle and E. 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.
- Kortmann, K. and V. Halbherr (2008a). Alterssicherung in Deutschland 2007 (ASID '07). Tabellenband 1: Alte Länder. München, TNS Infratest: 507.
- Kortmann, K. and V. Halbherr (2008b). Alterssicherung in Deutschland 2007 (ASID '07). Tabellenband 2: Neue Länder. München, TNS Infratest: 507.
- Kortmann, K. and V. Halbherr (2008c). Alterssicherung in Deutschland 2007 (ASID '07): Zusammenfassung wichtiger Ergebnisse. München, TNS Infratest: 147.

Kreyenfeld, M. (2009). "Uncertainties in Female Employment Careers and the Postponement of Parenthood in Germany." European Sociological Review Advance Access Published Online May 4, 2009.

Loose, B. and D. Frommert (2009). "Verstetigung in der Alterssicherung durch Integration ungesicherter Selbständigkeit in die gesetzliche Rentenversicherung." Deutsche Rentenversicherung 64(6): 501-517.

Modigliani, F. (1988). "The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth." The Journal of Economic Perspectives 2(2): 15-40.

Orlowski, R. and R. Riphahn (2009). "The East German wage structure after reunification." Economics of Transition 17(4): 629-659.

Rasner, A. (2010). Analysis within Boundaries? The Power of Administrative Data in the Study of the Material Well-Being of the Elderly, Forthcoming.

Rasner, A., R. K. Himmelreicher, M. M. Grabka and J. 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. SOEPpapers 70. Berlin, Deutsches Institut für Wirtschaftsforschung: 210.

Seitz, H. (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.

Sierminska, E., A. Brandolini and T. Smeeding (2006). "The Luxembourg Wealth Study - A cross-country comparable database for household wealth research." Journal of Economic Inequality 4(3): 375-383.

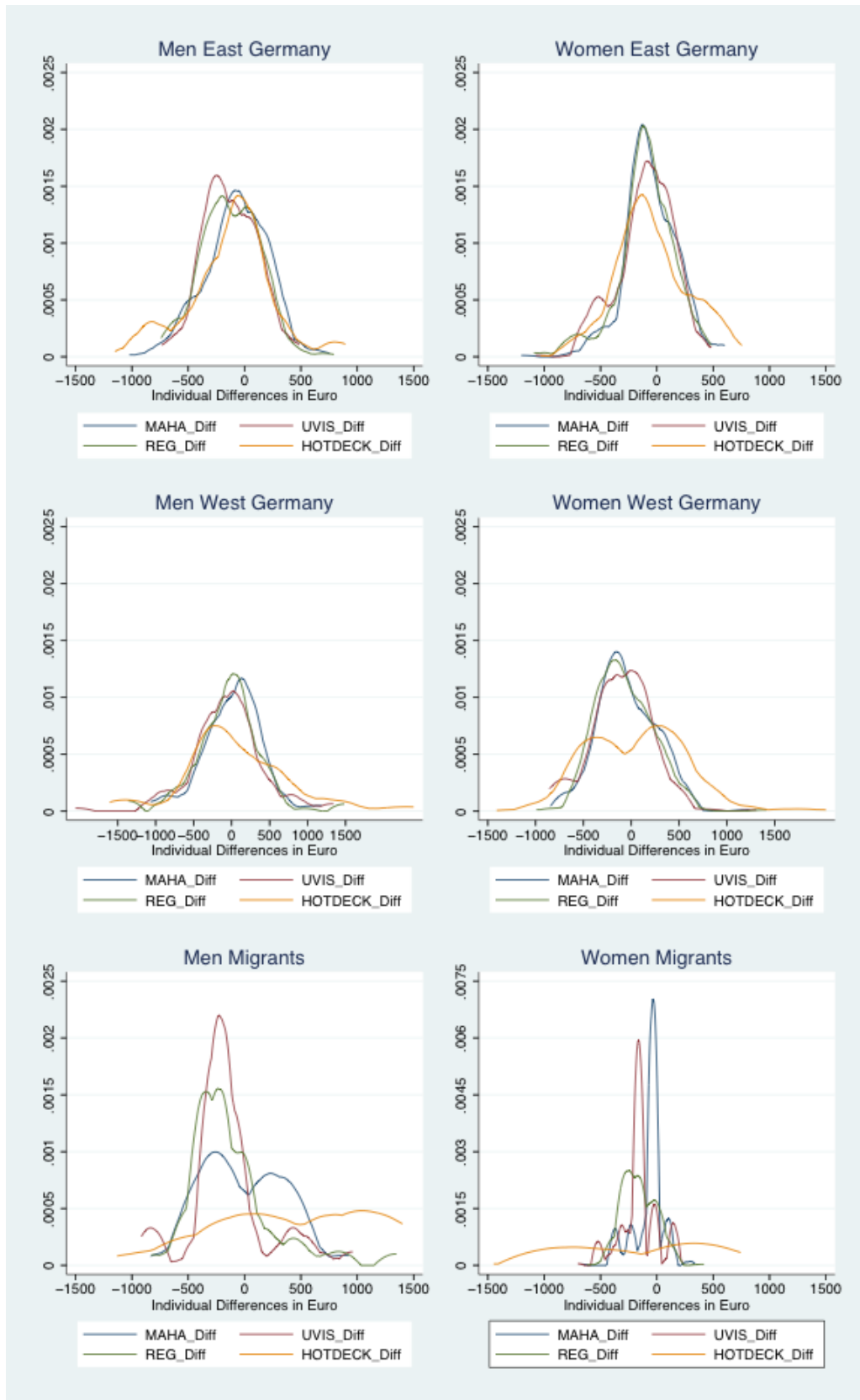
Stegmann, M. (2008). "Aufbereitung der Sondererhebung "Versichertenkontenstichprobe (VSKT)" als Scientific Use File für das FDZ-RV." DRV-Schriften Band 79.

Trappe, H. (2007). Lost in Transformation? Disparities of Age and Gender. After the Fall of the Wall: Life Courses in the Transformation of East Germany. M. Diewald, A. Goedicke and K. U. Mayer. Stanford, Stanford University Press: 116-139.

Wagner, G. G., J. R. Frick and J. Schupp (2007). "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements." Schmollers Jahrbuch - Journal of Applied Social Science Studies 127(1): 139-169.

8 Appendix

Figure A1: Group-Specific Kernel Density Plots for Differences between Observed and Matched Benefit Information



Source: SOEP 2007 and VSKT 2007, own calculations

Table A2 Individual differences between observed and matched monthly public Pension Benefit

		Overall (n=634)	Men East Germany (n=126)	Men West Germany (n=138)	Men Migrants (n=47)	Women East Germany (n=141)	Women West Germany (n=154)	Women Migrants (n=28)
Hot deck	Median	-14.30	-67.08	28.32	193.95	-82.26	43.86	178.70
	Mean	16.38	-124.24	112.23	202.97	-53.56	39.08	90.88
	Std. Dev.	542.81	399.72	701.35	681.73	338.59	566.00	486.18
Regression	Median	-103.81	-97.84	-34.74	-280.27	-101.25	-123.05	-137.09
	Mean	-105.52	-114.72	-42.81	-163.02	-125.80	-111.41	-142.05
	Std. Dev.	328.78	290.75	386.51	408.82	298.02	314.92	218.07
UVIS	Median	-112.58	-107.17	-67.71	-174.92	-81.51	-138.43	-165.30
	Mean	-101.58	-105.78	-70.02	-131.47	-90.56	-119.46	-145.24
	Std. Dev.	327.47	256.29	447.69	388.51	250.39	312.04	218.96
Mahalanobis	Median	-78.16	-100.66	-19.84	-170.33	-81.91	-88.88	-55.54
	Mean	-74.75	-106.86	-34.77	-90.68	-70.46	-76.91	-110.36
	Std. Dev.	320.59	315.89	375.54	381.78	259.96	317.35	218.50

Source: SOEP 2007 and VSKT 2007, own calculations