The evolution of tangibles, financial and social security wealth over the lifecycle: estimates for Germany

Beckers, Benjamin; Himmelreicher, Ralf K.; Schröder, Carsten

The Evolution of Tangibles, Financial and Social Security Wealth over the Lifecycle: Estimates for Germany

Benjamin Beckers, Ralf K. Himmelreicher & Carsten Schröder[*]

Abstract: *»Die Entwicklung von Geld-, Sach- und Sozialversicherungsvermögen über den Lebenszyklus: Eine Analyse für Deutschland«*. Using survey and administrative micro data, we describe the wealth distribution in Germany between 1978 and 2003, focusing on the birth cohort 1939 to 1953 resident in West Germany. Estimates are provided for three types of wealth, financial wealth, real wealth and social security wealth, i.e. the number of accumulated earning points in Germany's public pension insurance. While financial and real wealth are rather unequally distributed, inequality in accumulated earning points is substantially lower, indicating that Germany's pay-as-you-go pension system plays a prominent role in mitigating the inequality in overall wealth after retirement.

Keywords: wealth, wealth accumulation, lifecycle, savings, distribution, inequality, pensions, statistical matching, data fusion.

1. Motivation

This paper studies the distribution of wealth and its evolution over the lifecycle for (West) German households between 1978 and 2003. The objective is to shed some light on the wealth situation of the cohorts born between 1939 and 1953 with respect to three sources of wealth, namely financial assets, real assets, and social security wealth. By including all three sources of wealth, we are able to draw a more comprehensive picture of the wealth situation than other studies, which usually disregard social security wealth, as measured by accumulated earning points in Germany's pay-as-you-go (PAYG) pension system, a close proxy of future PAYG pension entitlements. Particularly, we empirically

[*] Address all communications to: Benjamin Beckers, University of Kiel, Department of Economics, Wilhelm-Seelig-Platz 1, 24118 Kiel, Germany;
e-mail: beckers@economics.uni-kiel.de.
Ralf K. Himmelreicher, Free University of Berlin, Institute of Sociology; Konstanzer Str. 42, 10709 Berlin, Germany; e-mail: ralf@himmelreicher.de.
Carsten Schröder, University of Kiel, Department of Economics; Wilhelm-Seelig-Platz 1, 24118 Kiel, Germany; e-mail: carsten.schroeder@economics.uni-kiel.de.
We owe special thanks to participants of the Workshop "Statistical Matching", November 26th 2010, in Jena.

describe the inter-temporal evolution of wealth levels, wealth accumulation over the lifecycle, and inequality in wealth levels.

Wealth is an important economic measure for several reasons: First, it provides the foundation to finance expenditures after retirement and acts as a buffer against unexpected shocks in the employment biography, i.e. consumption smoothing (Cagetti 2003, 339, Davies and Shorrocks 2000, 1). Second, it also depicts a household's ability to afford certain expenditures, such as a private home or financing the children's education (Cagetti 2003, 340). Additionally, wealth generates additional income through interest (Hauser and Stein 2001, 23-4), can be passed on to future generations and it provides the wealth-holding individual with a certain social status and power (Davies and Shorrocks 2000, 1). Savings and the stock of wealth are therefore decisive measures for the economic situation of households.

Regarding the importance of wealth, relatively few studies on its distribution and development have been carried out for Germany (e.g., Fachinger 1998; Fuchs-Schündeln et al. 2009; Hauser and Stein 2001 and 2003). In the present article we provide, for all three types of wealth, respective levels and their distribution. In order to do so, we combine information from two valuable data sets, the German "Sample Survey of Income and Expenditure" (IES) and the "Insurance Account Sample" (VSKT). To our knowledge, so far only Frick and Grabka (2010), Rasner et al. (2011) and Westermeier et al. (2012) have provided wealth data including social security wealth. These authors statistically matched data from the German Socio-Economic Panel and social security data (Sample of Active Pension Accounts).

In our study, data on financial and real assets stems from the "Sample Survey of Income and Expenditure" (IES), which is a cross-sectional survey on private households that has been carried out by the German Federal Statistical Office and the Statistical Offices of the Laender since 1964. Every five years, about 50,000 households provide not only socio-demographic information (age, sex, education, place of residence) but also information on their income and financial status, wealth formation, their tax burden and other charges as well as on household expenditures (for details, see Federal Statistical Office, 2003). Thus, the IES offers a one-of-a-kind compendium of variables on the socio-economic and socio-demographic situation of German households. Information is assembled at the household level and for many characteristics also on the level of individuals. However, inter-temporal analysis is made difficult by the cross-sectional character of the study. In particular, the database does not contain sufficient information on social security wealth, and, due to its cross-sectional nature, panel-methods cannot be applied directly.

Data on social security wealth in form of accumulated earning points for retirement entitlements from the statutory pension scheme is derived from the "Insurance Account Sample" (VSKT). Earning points measure the individual's contribution to the statutory pension scheme. An annual earning points value of

1.0 reveals that the insurant has earned a remuneration equal to the average remuneration of all insurants subject to the statutory pension scheme during the course of the year. The higher (lower) the remuneration, the higher (lower) is the value of the annual personal earning points. Earning points can also be credited during specific periods, e.g. after the birth of a child or in periods of unemployment. Retirement entitlements are then derived from the accumulated earning points over the employment period of the insurant. We interpret the number of accumulated earning points as our indicator for social security wealth. The VSKT is a register-based longitudinal study with monthly data on the employment biographies of state-insured individuals. Its longitudinal character, especially the long survey period from the first contribution until retirement, makes it interesting for panel studies. However, the VSKT is limited with regard to the range of variables as well as the space of included individuals, since only process-produced information from the accounts of individuals under the public pensions scheme are collected. Thus the VSKT lacks information on the household composition, on the private wealth status or on savings of civil servants and most self-employed.

Although the two data sets offer important complementary insights on the distribution and accumulation of financial, real and social security wealth, a contemporaneous analysis of all three sources of wealth on the household level is not possible, given the restrictions of the two individual data sets and their specific characteristics (cf. Fachinger 1998, 21). First, VSKT permits an assessment of *individual* social security wealth only as wealth information on other household members is not provided. Second, VSKT is not representative for the entire population. Instead, only persons at least once in contact with the public pension system are included. As a result, public servants' public pension entitlements and the private pensions of most of the self-employed persons are not contained in the data.

A successful statistical combination of both datasets could however compensate for their weaknesses and generate a valuable complementary database next to the two individual samples, at least for the sub-population of publicly ensured individuals. Therefore, we also present some preliminary considerations regarding the statistical matching of the cross-sectional IES data and the longitudinal VSKT data. In the course of a subsequent research project, both datasets shall be matched statistically to answer research questions in the field of decided inter-temporal analyses on wealth formation (including social security wealth) with a synthetic panel. A statistically matched dataset would provide a for Germany unique pool of information for answering research questions concerning the issue of wealth – its accumulation over the lifecycle, its levels, its composition and its distribution. Especially in light of demographic change, financial crises, the increasing importance of private and occupational retirement provision given decreasing public retirement entitlements, and

changing employment biographies, information of this kind is of central relevance.

The article is structured as follows. Chapter 2 presents the data sets of the SUFs-IES and the SUF-VSKT. Successively, methodological considerations and the data processing will be discussed in Chapter 3. Afterwards, the sample results from both the IES and VSKT will be presented in Chapter 4. In Chapter 5, we shortly illustrate the idea of statistical matching, and the options it provides for a deeper analysis of the wealth accumulation. In Chapter 6, we compare the informational contents of several variables, which can be considered for the identification of statistical twins. Furthermore, the IES- and VSKT-subpopulations suitable for the matching process are specified and necessary steps for data processing are explained. Also, aspects of data security will be discussed. Conclusion and outlook follow in Chapter 7.

2. Brief Descriptions of the Datasets

Both datasets are so called Scientific Use Files (SUFs), thus provided in anonymous micro-data formats, that are provided to the scientific community for non-commercial research purposes by the Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Laender[1] (in case of the SUF-IES) or the Research Data Centre of the German Pension Insurance (FDZ-RV)[2] (in case of the SUF-VSKT)[3].

Since 1983, the VSKT is a yearly drawn random sample from the insurance accounts of the statutory pension funds. Hence, it is a so called process-produced dataset. In order to carry out substantively coherent analyses, the sample is drawn according to several stratification criteria. These include sex, nationality, type of insurance and year of birth. The VSKT is a disproportionate panel, which offers representative results on the population of insurants by the use of expansion factors or frequency weights. The VSKT contains longitudinal information on all individuals covered by state insurance, especially on the income and employment biography. The primary statistical unit in the VSKT are individuals.

The Scientific Use File of the VSKT, e.g. SUF-VSKT-2003, is a 25 percent sub sample from the VSKT, thus containing about 60,000 individuals restricted to be between 30 and 67 years of age. Only Germans with residence in Ger-

[1] For the Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Laender refer to <http://www.forschungsdatenzentrum.de/en/index.asp>.

[2] For the Research Data Centre of the German Pension Insurance refer to <http://forschung.deutsche-rentenversicherung.de/FdzPortalWeb/dispcontent.do?id=main_fdz_english>.

[3] Both Research Data Centres were established and further developed with support of the Federal Ministry of Education and Research and the German Data Forum <http://www.ratswd.de/eng/index.html>.

many and whose account has been cleared at the reporting date are considered. Thus, some population groups – especially self-employed, assisting family workers and civil servants – are underrepresented (Richter and Himmelreicher 2008, 35ff.).

The VSKT is a dataset of flexible length that depicts the life of the insurant completely until the recording date. It consists of a "fix" part with invariable personal and specific insurance information and several "variable" parts. The information contained in the variable parts document multiple insurance-relevant characteristics of the insurant on a monthly basis over the entire active insurance biography, such as the adduced earning points, the employment situation, etc. The fix part is unambiguously linked to the variable part via the indicator key CASE.

Within the framework of the Sample Survey of Income and Expenditure (IES), private households are regularly interviewed on their income and expenditures, their wealth formation, their endowment with consumer goods and their living situation. Aim of the survey is to realistically collect and depict the socio-economic situation of private households in Germany from the point of view of the distribution and use of income. Households are hence the primary statistical unit in the IES, but some information is also available on the individual level. Thus it is possible draw conclusions from the socio-demographic characteristics of the household and its individuals.

Table 1: Central Characteristics of the Datasets SUF-VSKT and SUF-IES

Central Characteristics	Dataset	
	SUF-VSKT	SUF-IES
Releases	2003, 2005, 2006, 2007, 2008, 2009	1978, 1983, 1988, 1993, 1998, 2003, 2008 (since recently)
Population	Individual pension insurants from birth cohorts 1938 to 1975, residence in Germany	Households with residence in Germany
Survey method	Longitudinal	Cross-sectional
Sample size	about 60,000	about 45,000 per cross section
Kind of anonymisation	Factual	
Statistical unit	Individual level	Household and individual level
Dimension of the survey	Monthly panel data	Quarterly or annually

Source: Authors' design.

The IES is carried out every five years including all private households in Germany, whereas the participation is voluntary. Around 0.2 percent (ca. 60,000) of all German households is interviewed. All households are selected and interviewed by a given ratio schedule, where the household population is subdivided into groups according to certain quotation characteristics. For every group, the quota defining the number of the households to be interviewed is

predetermined. Scientific Use Files of the IES are made available by the Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Laender for scientific purposes. Individuals living in shared accommodations or institutions are not included in the SUF-IES. Furthermore, these datasets do not provide information on households above a certain net monthly income (IES 2003: 18,000 Euro) since it is assumed that these households rarely participate in the survey.

Central features of both datasets (SUF-IES and SUF-VSKT) are presented in Table 1, to provide an overview on their similarities and differences.

3. Methodological Considerations and Data Processing

3.1. IES Working Sample

In order to evaluate the formation of financial and real wealth of German households, the study is conducted using the SUFs from 1978 to 2003. Our working sample includes all IES households with household heads from the birth cohorts 1939-43 (Cohort I), 1944-48 (Cohort II) and 1949-53 (Cohort III). Since East-German households are not in the scope of the IES before the German reunification, the sample regards West-German households only. Altogether, our working sample of West-German households with heads born between 1939 and 1953 contains 75,149 households. The variables of interest are the aggregates net financial wealth, net real wealth and of course net non-pension wealth, the sum of net financial and net real wealth. Our procedure closely follows Fuchs-Schündeln et al. (2009). Gross financial wealth is made up by the IES variables saving accounts, saving certificates and term deposits, building society saving contracts and several components of securities. Consumer credits are eventually deducted to obtain net financial wealth. Net real wealth is made up by the assessed tax value of property less outstanding mortgages. The computations are based on the pooled IES dataset described in Bönke et al., 2010.

3.2. VSKT Working Sample

Our VSKT 2003 working sample includes all insured persons living in West Germany who were born between 1939 and 1953. We measure the total social security wealth over the lifecycle in Germany's PAYG pension system with the indicator accumulated total earning points. Earning points are calculated as follows: the individual gross monthly labour earnings which are due for social

security contributions are divided by gross earnings per average employee.[4] Total earning points include labour based earnings and welfare based points like child and care benefits. Therefore earning points are dimension free as the division annuls the influence of all factors with the same effect on denominator and nominator – like the real economic activity – and deflation is unnecessary. Earning points are censored both from below ("Geringfügigkeitsgrenze") and from above by the earnings ceiling ("Beitragsbemessungsgrenze"), and censoring points differ over time, potentially biasing the inequality estimates (Fachinger and Himmelreicher 2010).

3.3. Wealth Adjustments for Changes in Consumer prices and Differences in Household Size

Our indicator of social security wealth, the number of accumulated earning points over the lifecycle over the period t to $t+x$ requires no price adjustments as earnings point are given as the ratio of the insurant's social security income relative to mean social security income (see Section 3.2 for details). Moreover, as the information is provided on the level of individuals only, an adjustment for differences in household size is not required (and not feasible). So, if one insurant has accumulated a total of 30 earning points and another 20 until period $t+x$ and both retire in $t+y \geq t+x$, the formers' actual pension entitlement is 1.5 times higher.

As real and financial wealth are measured on the household level and are expressed in monetary units (DM or EUR), we have adjusted all values for changes in consumer prices (CPIs) with data provided by the German Federal Statistical Office[5]. All data are expressed in 2003 EUR prices. CPI adjustments and the DM-Euro conversion factor can be taken from Table 2.

Table 2: Consumer Prices and DM-Euro Conversion

Year	CPI	DM-Euro conversion factor
1978	54.3	
1983	68.8	
1988	72.9	1/1.95583
1993	85.9	
1998	93.4	
2003	100.0	--

Source: Consumer price indices are available online at German Federal Statistical Office.

Furthermore, household wealth levels are adjusted for differences in household size using the square root equivalence scale. Equivalence scales account for the

[4] § 70 Book VI of the German Social Welfare Code (SGB VI) and Annex 1 Book V of the German Social Welfare Code (SGB V).

[5] See <http://www-genesis.destatis.de/genesis/online>.

increasing needs in consumption of larger households, and adjust for the disproportional relationship which is due to economies of scale in consumption. The OECD square root equivalence scale "divides household income by the square root of household size" (OECD Statistics Portal, 2010). Here, wealth levels are adjusted and the outcome is equivalent wealth.

We report means, standard deviations, Gini coefficients and equivalent wealth distributions for all SUFs from 1978 to 2003 and the three aforementioned birth cohorts. All data is weighted by frequency weights as provided in the two datasets. The Gini coefficient is an inequality measure expressed as twice the area between the 45°-line and the so-called Lorenz curve. The Lorenz curve is derived by ordering the households according to their equivalent wealth and then accumulating the proportion of equivalent wealth across households in relation to aggregate total equivalent wealth. The Lorenz curve then depicts which proportion of the population holds which proportion of total equivalent wealth. Hence can the Gini be defined as the "…average difference between all possible pairs of income [here: equivalent wealth] in the population, expressed as a proportion of total income [here: total equivalent wealth] …" (Cowell 2011, 26) and can be calculated for a discrete distribution as:

$$\text{Gini}=\frac{1}{2n^2\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i-y_j\right|$$

,where y_i is a measure of household/individual i's resources.

As earning points are provided on the individual level, a needs-adjustment by means of the OECD equivalence scale is neither necessary nor possible. Accordingly, we deal with equivalent real and financial wealth distributions but with distributions of individual social security wealth.

Of course, financial and real wealth levels can sometimes be negative. Particularly, in our case this is true for net real wealth. As the Gini coefficient is defined for non-negative resource levels only, simply erasing observations with negative wealth from the database would yield biased results, particularly in case of inter-temporal analysis if the share of the population with negative wealth is not constant. For this reason, in the inequality analysis, households' financial and real assets were set to be at least equal to zero. Accordingly, inequality indices can be interpreted as a lower bound of the "true" inequality level. The descriptive statistics rely on the actual wealth levels, containing also zero and negative wealth levels. For our measure of social security wealth, the number of accumulated earning points, we have followed a different procedure. Here, a zero in a particular year means that, so far, the individual has not been in contact with the pension insurance. In this sense, in the particular year, it cannot be assigned to the group of insured persons. Accordingly, we refrain from considering the person in the particular year in our calculations of social security wealth.

4. Sample Results

Figure 1 presents our findings regarding the distributions of net financial, real and net non-pension wealth (excluding social security wealth) across all represented cohorts for the six SUFs of the IES. As explained in Section 3.3, all IES data are adjusted for price changes and also for differences in household size (using the square-root equivalence scale).

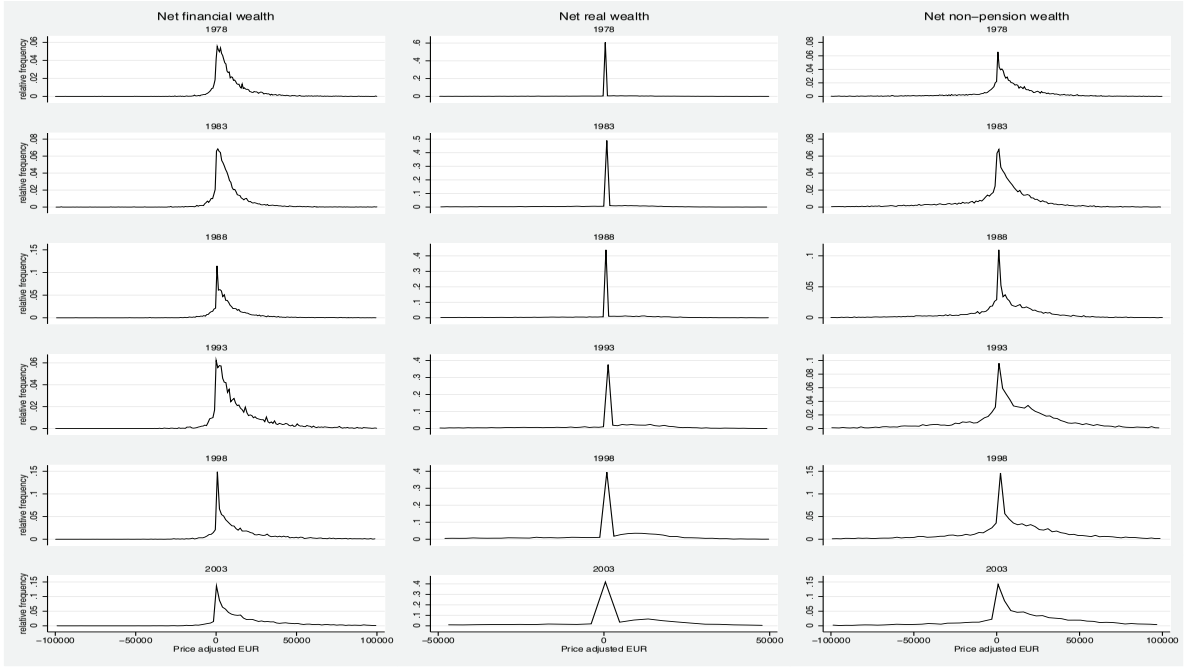
Each column relates to a particular type of wealth, and comprises six graphs, one for each observation year. In each graph, the abscissa gives the level of wealth, and relative frequencies are depicted on the vertical axis. Net financial wealth is the most widespread type of wealth, although absolute levels are usually small. Net real wealth is most strongly concentrated around zero, hence do only relatively few households hold real assets compared to financial assets. Furthermore, there exist a number of households with negative wealth in all periods which cannot be disregarded.

In Table 3, we present descriptive statistics of non-pension wealth, distinguished by cohorts and years. We comment on financial wealth first. Means of both net and gross financial wealth generally increase over time for all cohorts and usually more than 90 percent of a cohort in a particular year has a non-zero financial wealth.[6] For birth cohort 1 (born between 1939-1943), both gross and net financial wealth almost triple over the observation period, reaching levels of slightly below 30,000 EUR in year 2003. For the two later born cohorts, wealth levels grow at a lower rate and are always lower compared to cohort 1. For example, gross financial wealth in 2003 is 29,263 EUR for cohort 1, 27,177 EUR for cohort 2 (born between 1944 and 1948), and only 20,497 EUR for cohort 3 (born between 1949 and 1953).

While financial wealth is held by a vast majority of households, the fraction of our sample holding real wealth is substantially lower. For example, in year 1978 only 22 percent of the third birth cohort owns gross real wealth. Of course, the share rises over time. Due to the smaller share of the sample actually possessing real wealth, also the means are lower compared to financial wealth. It is also interesting to note that net real wealth, on average, is negative. This is due to the fact that, due to data restrictions, our computations rely on assessed tax value of property, and these are typically lower than the actual market values.

[6] For the years 1998 and 2003 the number is slightly lower. This might be a statistical artifact coming from changes in the survey design. Particularly, between 1993 and 1998 the surveying period has been reduced from twelve to three month.

Figure 1: Distributions of Equivalent Non-Pension Wealth



Database: IES 1978-2003.

Table 3: Means and Standard Deviations for Equivalent Financial and Real Wealth in EUR, IES

		Gross financial wealth				Net financial wealth			Gross real wealth			Net real wealth			Net non-pension wealth		
Year	Co-hort	N	Mean	SD	Share with posi-tive wealth	Mean	SD	Share with posi-tive wealth	Mean	SD	Share with posi-tive wealth	Mean	SD	Share with posi-tive wealth	Mean	SD	Share with posi-tive wealth
1978	1	6,145	10,740	18,454	97.22	9,805	18,811	97.90	10,939	17,994	51.10	-3,905	26,527	51.06	5,900	32,647	98,50
	2	4,153	10,159	12,901	96.81	8,790	14,067	97.70	8,281	17,950	37.35	-5,372	25,748	37.30	3,418	28,244	98.27
	3	2,915	9,101	11,046	96.63	7,824	11,702	97.77	4,841	16,035	22.07	-2,956	20,228	21.98	4,868	22,913	97.89
1983	1	5,865	9,818	16,412	95.02	8,700	17,285	96.88	11,898	17,749	60.71	-6.812	35,608	60.66	1,888	39,907	97.55
	2	4,874	9,330	14,910	95.45	7,778	15,246	97.33	10,650	17,328	54.31	-11,648	36,896	54.18	-3,870	38,814	97.95
	3	4,887	8,597	11,768	95.17	6,850	15,479	97.00	7,718	18,971	39.91	-8,217	31,693	39.84	-1,367	34,011	97.45
1988	1	5,609	12,332	22,988	91.97	10,892	23,732	94.84	12,359	16,138	60.52	-4,223	31,571	60.42	6,669	40,567	95.94
	2	4,766	11,078	20,196	93.06	9,407	20,548	94.82	11,510	15,564	58.92	-10,753	43,610	58.82	-1,346	47,152	95.86
	3	5,356	9,083	18,375	92.00	7,286	17,929	94.69	8,903	13,151	51.49	-10,214	31,111	51.44	-2,928	34,121	95.57
1993	1	3,173	19,655	29,304	94.27	18,674	29,811	95.29	13,934	17,649	66.86	-4,609	39,279	66.74	14,065	49,843	96.39
	2	2,982	18,573	38,894	95.59	17,219	39,401	96.63	12,432	16,889	63.72	-10,461	51,239	63.66	6,758	65,358	97.14
	3	3,772	13,768	21,776	95.64	12,630	22,329	96.76	10,268	13,861	61.14	-11,705	37,193	61.06	925	43,037	97.51
1998	1	3,549	24,031	42,474	89.80	23,076	42,869	92.03	13,396	22,584	64.05	-4,993	53,148	63.82	18,083	67,638	94.63
	2	3,264	19,765	39,547	87.67	18,538	39,669	90.27	11,854	20,169	63.76	-10,771	48,231	63.59	7,766	58,372	92.58
	3	4,428	16,205	31,581	88.32	14,371	33,856	92.20	9,689	15,149	58.83	-13,997	43,082	58.78	374	53,680	94.19
2003	1	3,126	29,263	50,578	87.74	28,766	50,734	89.31	14,977	27,276	57.27	-1,115	44,899	58.26	27,651	69,869	92.29
	2	2,772	27,177	53,491	88.58	26,320	53,929	90.60	15,701	27,294	59.79	-8,740	56,456	62.07	17,580	76,595	92.86
	3	3,513	20,497	37,322	89.24	19,383	38,007	92.03	13,818	44,834	57.80	-10,150	68,050	60.28	9,234	72,960	93.65

Database : IES 1978-2003

Note: Cohort 1: born between 1939-1943, Cohort 2: born between 1944-1948, Cohort 3: born between 1949-1953. Own calculations. SD denotes standard deviation. N is the non-weighted number of observations. Wealth levels are weighted using the frequency weights as provided in the underlying database.

Finally, concerning net non-pension wealth, the aggregate of financial and real wealth, we find a remarkable picture. While it is positive for all cohorts at the very beginning of the observation period (year 1978), and also for the period from 1993 to 2003, it becomes negative in 1983 and 1988 for cohorts 2 and 3. What is underlying this particular pattern is that the fraction of households actually holding real wealth rises rapidly over time in combination with net real wealth being negative for many households.

Results on wealth inequality, as indicated by Gini coefficients, are provided in Table 4. Again, results are distinguished by type of wealth. We would like to remind the reader that Gini coefficients are derived by replacing all negative and zero wealth levels by zeros (see Section 3.2). Overall, Gini coefficients indicate that wealth inequality in Germany is high. For example, in year 2003 Gini coefficients range between around 67 percent (financial wealth) and more than 80 percent (net real wealth). Furthermore, inequality tends to be higher for the later born. Over time, financial wealth becomes more unequally distributed, while real wealth and net non-pension wealth inequality exhibits no distinct inter-temporal pattern.

Table 4: Gini Coefficients by Types of Non-Pension Wealth

Year	Cohort	Gross financial wealth	Net financial-wealth	Gross real-wealth	Net real wealth	Net non-pension wealth
1978	1	57.558	60.830	69.952	86.465	67.237
	2	57.242	61.401	78.249	91.946	68.345
	3	56.442	60.947	88.019	95.241	66.839
1983	1	58.821	61.918	63.020	84.156	70.618
	2	60.237	63.344	67.279	90.211	74.957
	3	58.021	62.855	76.605	93.740	73.776
1988	1	62.760	65.755	62.250	80.416	69.526
	2	64.461	68.138	62.917	85.161	74.836
	3	63.487	67.763	67.645	88.612	76.077
1993	1	59.669	61.737	59.183	76.446	64.761
	2	63.879	66.376	61.432	81.819	71.777
	3	60.929	63.895	61.993	84.188	71.763
1998	1	67.415	68.588	62.166	75.209	68.214
	2	67.868	69.723	60.912	81.534	73.384
	3	67.806	70.191	62.989	84.914	76.209
2003	1	66.483	66.988	68.617	77.974	66.604
	2	67.152	68.201	67.700	81.881	70.423
	3	67.534	68.937	70.257	86.445	74.722

Database: IES 1978-2003; Note: Cohort 1: born between 1939-1943, Cohort 2: born between 1944-1948, Cohort 3: born between 1949-1953.

In sum, wealth is substantially more unequally distributed compared to (disposable) income, which usually shows Gini coefficients around 30 percent in Germany. Since financial wealth and tangibles will become more important for individuals after retirement in the next years due to the previous pension re-

forms, the high inequality in wealth levels will map in higher inequality in living standards among elderly in the next decades.

The more-even distribution of income (and disposable incomes in particular) is also reflected in the distribution of accumulated earnings points, our indicator of social security wealth. Table 5 presents our findings regarding the distributions of social security wealth (excluding financial, real and net non-pension wealth) across all represented cohorts based on VSKT 2003. Means of social security wealth increase over time and are higher at a given point in time for earlier born cohorts. The latter finding, of course, corresponds with the differences in underlying length of working histories, reflecting different time spans, career effects, labour force participation etc. For example, the oldest members of cohort 1 are born in 1939. If they started working at the age of 14 years in 1953, working life in the year 2003 is 51 years maximum. By contrast, for the youngest members of cohort 3 (born in 1953) the maximum length of working life in 1978 is 12 years (again assuming that they start to work at age 14).

Table 5: Means, Standard Deviations, and Gini Coefficients of Aaccumulated Total Earning Points

Pension wealth					
Year	Cohort	N	Mean	SD	Gini
1978	1	34,462	12.16	7.82	36.901
	2	29,955	8.67	5.42	35.785
	3	31,057	5.84	3.39	33.197
1983	1	34,462	15.19	10.25	38.648
	2	29,955	11.63	7.70	37.968
	3	31,057	8.62	5.31	35.351
1988	1	34,462	18.22	12.90	40.428
	2	29,955	14.53	10.26	40.389
	3	31,057	11.42	7.60	38.142
1993	1	34,462	21.09	15.56	42.007
	2	29,955	17.33	12.92	42.478
	3	31,057	14.08	10.14	41.144
1998	1	34,462	23.73	18.05	43.211
	2	29,955	20.15	15.55	43.831
	3	31,057	16.93	12.72	42.770
2003	1	34,462	25.37	19.74	44.041
	2	29,955	22.73	18.06	44.979
	3	31,057	19.70	15.33	44.158

Database: FDZ-RV-VSKT2003; Note: Cohort 1: born between 1939-1943, Cohort 2: born between 1944-1948, Cohort 3: born between 1949-1953. N is unweighted, results are weighted. Own calculations. SD denotes standard deviation.

Not only social security wealth levels are rising, but so are inequality levels, as indicated by rising standard deviations and Gini coefficients. Comparing the Gini coefficients for financial, real and social security wealth, it is transparent that social security wealth, by far, is the most equally-distributed type of wealth. It is, however, not ruled out that this finding is due to the fact that cen-

soring points differ over time (see Section 3.2). Accordingly, it potentially serves as a device for smoothing differences in living standards in the period after retirement.

5. Statistical Matching

In the following two sections, we present some preliminary considerations regarding the statistical matching of cross-sectional survey data from the IES and longitudinal data from the VSKT.[7] Specifically, a synthetic panel could be created by combining several SUFs of IES-cross-sections via the longitudinal information of the SUF-VSKT. With such a synthetic panel, decided intertemporal analyses on wealth formation (including pension entitlements) could be carried out, which would not be possible with both single datasets used separately.

5.1. Basic Concept

The method of statistical matching enables a statistical consolidation of several datasets, such that the resulting dataset combines all relevant information. In the data fusion process one distinguishes between the primary dataset and the independent secondary file at first. Here, the primary file forms the "receiving" dataset and the secondary file forms the "sending" file. Primary and secondary file can contain common as well as dataset specific information:

Let
(1) A = **(X,Y)**
be the primary dataset and
(2) B = **(X,Z)**

the from A independent secondary file, then the aim is to generate a merged dataset *F* that combines the information of both datasets. Of course, an exact matching of the datasets (*record linkage*) is possible only if a unique and common primary key, such as e.g. the identification card number, is contained in both datasets; this is not the case in SUF-IES and SUF-VSKT.[8] The idea of statistical matching, instead, is to identify pairs of observations from both datasets which are statistically close according to particular characteristics (as de-

[7] Statistical matching does not only provide interesting options for data fusion (Rässler 2002), but also for the treatment of missing values (cf. as fundamental papers Rubin (1987) as well as Rubin and Little (1987)) or the statistical assessment of underlying effects (cf. as fundamental papers Heckman et al. (1997, 1998), Rosenbaum and Rubin (1983, 1984), Rubin (1974), as well as Holland (1986)).

[8] The IES and VSKT do not come with a common primary key. The resulting factual anonymised Scientific Use Files, generated by the respective research data centres do not contain direct identifiers due to reasons of data security.

fined by the matching variables). To guarantee that sufficiently dissimilar observations will not be matched, the population can first be partitioned in predefined strata, e.g. female and male persons. Statistical matching techniques include, for example, nearest neighbours, kernel, local linear regression, spline or Mahalanobis matching. For such and other methods see Rasner et al. (2011).

So far, we have not conducted a statistical matching of the two datasets. However, in the next sections we will outline preliminary steps which are necessary before the actual implementation. This includes a preferably identical classification of the populations under consideration, the identification of suitable matching variables as well as the establishment of their comparability.

5.2. Population Under Investigation

The populations in both datasets differ according to several criteria. The IES is supposed to draw a representative picture of all private households as long as they are not living in a shared accommodation or institution. Thus, pensioners, unemployed, employees, civil servants, self-employed as well as students are all included in the data base. In contrast, the VSKT is a sample of all German residents subject to the statutory pension scheme, where individuals between 15 and 67 years of age are included only if they have accumulated personal earning points for at least one month between 1953 and 2006. As a consequence, some groups such as civil servants and self-employed (who are not obliged to follow the statutory pension scheme) are underrepresented. Furthermore, for example in the SUF-VSKT 2006, only individuals born between 1939 and 1976 are included. Therefore certain SUF-IES birth cohorts cannot be linked directly with their SUF-VSKT equivalents. Hence, the possibility of linking IES-statistical units via the VSKT is limited: The matching can only be carried out for German pension insurants of certain birth cohorts.

5.3. Matching – Variables and their Comparability in SUF-VSKT and SUF-IES

Only the variables listed in both datasets, or those which can be broken down from the available data to the individual level, can be considered for the statistical matching. Furthermore, it has to be assured that these variables are comparable with regards to content. If this is not the case, it has to be checked if a recodification of the existing variables or a utilisation of further variables or statistics, can generate variables which are suitable for the matching process.

For some matching variables, such as sex, place of residence or age, the informational content is identical in SUF-IES and SUF-VSKT. For further central matching variables, especially income, employment situation or family background, additional work, a re-codification is necessary first and foremost. This is due to the fact that the informational content of SUF-VSKT-variables is formed according to the pension law regulations by which they have been

collected. Thus children are only listed in the SUFs-VSKT if their existence is relevant to the individuals under consideration, e.g. because periods of parental leaves are credited to the pension accounts. Therefore, male pension insurants list children only in rare cases. In contrast, the IES accounts for all children living in the household.

A linkage between SUF-VSKT and SUF-IES via income from employment also requires adjustments to ensure comparability. The SUF-VSKT provide information on the incomes of the insurants only indirectly, namely through their individual gross remunerations in form of so called personal earning points. An annual earning points value of 1.0 reveals that the insurant under consideration has earned a remuneration equal to the average remuneration of all insurants subject to the statutory pension scheme during the course of the year. The higher the remuneration, the higher is the value of the annual personal earning points. A lower remuneration equivalently leads to a lower value of earning points. But also periods of parental leave or unemployment can result in earning points. Furthermore, it has to be noted that the annual earning points cannot exceed a certain maximum value, which is defined according to the contribution assessment ceiling. This hence implies that also the gross remunerations[9] defined by the earning points are right-censored. Additionally, the left-censoring following from the minimum account limit (so called 400 EUR-Jobs) and special rules for insurants from the New Laender have to be taken into consideration.

In the SUFs of the IES it is possible in principle to calculate individual gross remunerations, since decided information on the different types of income (such as income from employment/self-employment (personal level), public transfer payments, income from rent and lease, etc. (for a detailed overview on the included income variables compare Bönke et al. 2010)) is available on the individual level. However, the SUF-IES are truncated: Information on households with a household net income above a certain threshold is not included.

6. Aspects of Data Security

In general, both datasets are factual anonymised micro-datasets, which are examined on data security aspects by either the Research Data Centres of the Federal Statistical Office and the Laender or the FDZ-RV. The factual ano-

[9] The (annual) remuneration subject to contributions to the pension scheme is calculated (in 2009) in case of ALG I: 80% of the former gross income, in case of unemployment benefits: amount of the benefits, in case of ALG II: 205 EUR per month, in case of ATZ: at least (mostly) 90% of the former gross income, in case of a monthly income between 400-800 EUR: successively lower than the gross income according to the sliding social security contribution scale (Gleitzonenregelung) and in case of high incomes according to the contribution assessment ceiling, etc.

nymisation was monitored by the respective departments for data security regarding the relevant legal background (Federal Statistics Act (BStatG[10]) versus Social Code (SGB[11]). From the perspective of the data-related implementation, the legal differences are less important, since the factual anonymisation of SUF-VSKT and SUF-IES has already been carried out by the elimination of direct identifiers and sensitive characteristics (name, insurance account number, etc.), the sampling process itself, and the oversimplification and classification of properties. To achieve this, criteria for the concrete organisation of research data centres in Germany have been compiled in a joint effort between the German Data Forum (RatSWD) and the publicly funded research data centres in 2007 (Bender et al. 2010, 217ff.).

From the perspective of the FDZ-RV, several cooperative projects are currently conducted with the aim of enlarging the informational content of the FDZ-RV cross-sectional data by statistical matching procedures. Here, a project on the consolidation of data from the socio-economic panel (SOEP) (with its information on the household situation and further old-age incomes) with data from the SUF Vollendete Versichertenleben, can be cited exemplarily, since it has been certified as compliant with data protection regulations (Rasner et al. 2007).

The basic prerequisite for matching projects is that the scientific user has obtained the anonymous, individual datasets in line with legal and valid license agreements. In case of the SUF-VSKT and SUF-IES, this implies that for both datasets, data license agreements have been concluded with the respective research data centre. Following, an application for a matching project has to be filed at the research data centres. The validation of data security aspects is a legal prerequisite for the matching.

Since factual anonymous datasets are exclusively matched statistically, an explicit re-identification of individuals is not possible even with a disproportionate amount of time, costs and workforce. Aim of the matching is not the identifications of individuals, but the matching of statistical twins, i.e. individuals that share a maximum amount of common characteristics. From a technical perspective, a direct linkage is not possible, since the underlying anonymous survey- and register data neither contain bijective identifiers (e.g. insurance account number) nor names or exact dates of birth which could provide with a possibility to match identical individuals.

[10] The provision of micro-data to the scientific community by the statistical offices necessitated the insertion of § 16 Abs. 6 in the BStatG, to pass on micro-data with sufficient potential for analysis (Bauer 2008, 159).

[11] §§ 67-85a SGB X defines the regulations for the protection of social data.

7. Conclusion

This paper has shown the potentials of both datasets to investigate the distribution of wealth in all three dimensions – financial assets, tangibles and social security wealth. However, to draw a complete picture of total wealth on household or individual level, information from a statistically matching of both datasets is necessary. The construction of a synthetic panel by linking several Scientific Use Files of the IES with their VSKT-counterparts is a challenging task. This results from differences in the survey designs of both datasets and in the underlying populations as well as from differences in the informational content of the respective variables. While the SUFs of the IES contain manifold socio-demographic and – economic information on household and individual level from voluntary participation in interviews, the SUFs of the VSKT provide information relevant to the German pension law (which are thus valid and enforceable) on insured individuals.

Due to differing aims of the surveys, the included variables differ also systematically in their specifications. A successful statistical matching hence requires that the variables are re-codified in a manner that guarantees their comparability. Moreover, both datasets differ in the represented population groups: The IES is supposed to draw a representative picture of all private households in Germany as long as they are not living in a shared accommodation or institution. The VSKT however is a sample of all German residents between 15 and 67 years of age who are subject to the German statutory pension scheme and have been insured for at least one month.

References

Bauer, P. 2008. Das Forschungsdatenzentrum der Statistischen Landesämter. Ein Beitrag zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. In *Fortschritte in der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag*, ed. G. Rolf, M. Zwick and G. G. Wagner, 159-83. Baden-Baden: Nomos.

Bender, S., R. K. Himmelreicher, S. Zühlke, and M. Zwick. 2010. Access to Microdata from official statistics. In *Building on Progress. Expanding the Research Infrastructure for the Social, Economic, and Behavioral Science*, ed. RatSWD – German Data Forum, volume 1, 215-30. Budrich UniPress Ltd.: Opladen & Farmington Hills MI, S.

Bönke, T., C. Schröder, and C. Werdt. 2010. *Compiling a Harmonized Database from Germany's 1978 to 2003 Sample Surveys of Income and Expenditure*. Discussion Papers of the Free University Berlin, School of Business & Economics, 4.

Cagetti, M. 2003. Wealth Accumulation over the Life Cycle and Precautionary Savings. *Journal of Business and Economic Statistics* 21 (3): 339-53.

Cowell, F. A. 2011. *Measuring Inequality*, 3rd edition. Oxford: Oxford University Press.

Davies, J. B., and A. F. Shorrocks. 2000. Chapter 11, The distribution of wealth. In *Handbook of Income Distribution*, *Elsevier*, ed. Anthony B. Atkinson and Francois Bourguignon, Volume 1. 605-75.

Deaton, A. 1994. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore and London: The John Hopkins University Press.

Fachinger, U. 1998. Die Verteilung der Vermögen privater Haushalte: Einige konzeptionelle Anmerkungen sowie empirische Befunde für die Bundesrepublik Deutschland. *ZeS Working Paper*, 13.

Fachinger, U., and R. K. Himmelreicher. 2010. Income Mobility. Some Remarks on Security and Incentive Aspects of Income Mobility – Illustrated by an Empirical Analysis for Germany. *ZAG Working Paper*, 7.

Federal Statistical Office. 2003. *Wirtschaftsrechnung; Einkommens- und Verbrauchsstichprobe – Aufgabe; Methode und Durchführung der EVS 2003*. Fachserie 15, 7.

Frick, J. R., and M. M. Grabka. 2010. Alterssicherungsvermögen dämpft Ungleichheit – aber große Vermögenskonzentration bleibt bestehen. *DIW Wochenbericht*, 77: 2-12.

Fuchs-Schündeln, N., D. Krueger and M. Sommer. 2009: Inequality Trends for Germany in the Last Two Decades: A Tale of Two Countries. *Review of Economic Dynamics* 13: 103-32.

Hauser, R., and H. Stein. 2001. Die Vermögensverteilung im vereinigten Deutschland. *Stiftung Der Private Haushalt* 37.

Hauser, R., and H. Stein. 2003. Inequality of the Distribution of Personal Wealth in Germany, 1973-1998. The Levy Economics Institute of Bard College Working Paper, 398.

Heckman, J., H. Ichimura, and P. Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies* 64: 605-54.

Heckman, J., H. Ichimura, and P. Todd. 1998. Matching as an Econometric Evaluation Estimator: Theory and Methods. *Review of Economic Studies* 65: 261-94.

Holland, P. W. 1986: Statistics and Causal Inference. *Journal of the American Statistical Association* 81 (396): 945-60.

OECD Statistics Portal. 2010. What are Equivalence Scales, Long Abstract, <http://www.oecd.org/LongAbstract/0,2546,en_2825_497118_35411112_1_1_1_1,00.html>.

Rässler, S. 2002. *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches. Lecture Notes in Statistics, 168*. New York: Springer.

Rasner, A., J. R. Frick, and M. G. Grabka, M.G. 2011. Extending the Empirical Basis for Wealth Inequality Research Using Statistical Matching of Administrative and Survey Data. *SOEPpapers of DIW Berlin* 359.

Rasner, A., R. K. Himmelreicher, M. G. 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 of DIW Berlin* 70.

Richter, M., and R. K. Himmelreicher. 2008. Die Versicherungskontenstichprobe als Datengrundlage für Analysen von Versicherungsbiografien unterschiedlicher Altersjahrgänge. *DRV-Schriften* 79: 34-61.

Rosenbaum, P. R., and Rubin, D. B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70 (1): 41-55.

Rosenbaum, P. R., and D. B. Rubin. 1984. Reducing Bias in Observational Studies using Subclassification on the Propensity Score. *Journal of the American Statistical Association* 79: 516-24.

Rubin, D. B. 1974. Estimating Causal Effects of Treatments in Randomised and Non-Randomised Studies. *Journal of Educational Psychology* 66: 688-701.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Rubin, D. B., and R. J. A Little. 1987. *Statistical Analysis with Missing Data*. Wiley, New York, XIV.

Westermeier, C, A. Rasner, and M. G. Grabka. 2012. The Prospects of the Baby Boomers: Methodological Challenges in Projecting the Lives of an Aging Cohort. *SOEPpapers of DIW Berlin* 440.