

"Representative wealth data for Germany: The impact of imputation and the choice of the aggregation unit"

Joachim R. Frick *, Markus M. Grabka & Eva M. Sierminska

Abstract: The definition and operationalization of wealth information in population surveys requires more or less normative assumptions. Such decisions made in both the pre- and post-data-collection stage may interfere considerably with the substantive research question. Looking at 2002 wealth data from the German SOEP, this paper focuses on the impact of collecting information at the individual rather than household level, and on "multiple imputation and editing" as a means of dealing with measurement error and non-response behavior.

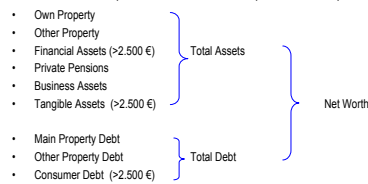
First, when measured in "per capita household" terms, wealth is clearly less unequally distributed than at the individual level. This is the result of significant redistribution within households, and also provides evidence of a significant persisting gender wealth gap. Secondly, multiple imputation appears to be an effective means of coping with selective non-response. There is a significant impact of imputation on the share of wealth holders and on aggregate wealth whereas the results are ambiguous with respect to inequality: The Gini coefficient for *net* worth decreases when considering imputed values, whereas a top-sensitive measure (HSCV) doubles. The non-random selectivity built into the missing process and the consideration of this selectivity in the imputation process clearly contribute to this finding.

Keywords: Wealth, Item non-response, multiple imputation, SOEP

JEL classification: D31, C81, I32

SOEP Wealth module in 2002

Individual level (all HH members >16): n=23.900 (12.500 HH)



Not included: cars, public pension entitlements, durables

Surveying wealth information at individual level

(A) Are you personally the owner of the house or apartment in which you live?

Yes ☐ No ☐

Value: If you were to sell today, how much would you receive for your house/apartment including land? EURO

Burden: If you still have a loan taken out on your house/apartment, how high is the remaining debt (excluding interest)? EURO

Personal share of property: Are you the sole owner (100%) or co-owner (e.g. with your spouse)? Sole Owner ☐ Share in %

The principles of imputation

Alternative Strategies to deal with INR

- Case-wise deletion (only valid observations)
- Weighting
- Single & multiple imputation (Rubin 1976, 1987)

Non-Response treatment in SOEP data

- Unit-NR → weighting
- Partial-NR ~5% → imputation
- Item-NR ~15-33% → imputation
- "Inconsistencies" <10% → editing

Principles of Editing & Imputation

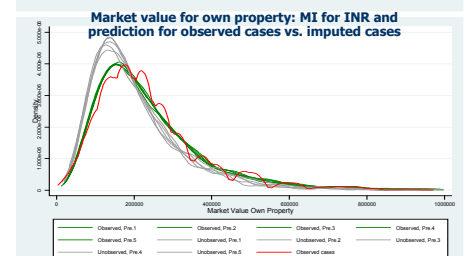
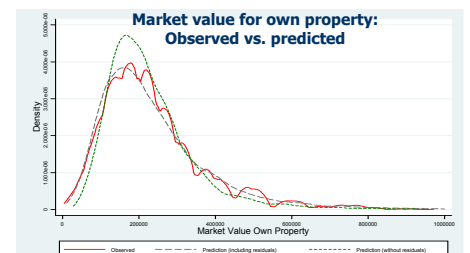
- Editing
 - Extreme outliers (e.g., missing 1,000 digits)
 - Co-owning couples (esp. owner-occ. property)
- Logical Imputation
 - Derived from non-missing partner information
- Imputation of missing information (INR/ PUNR)
 - Logit: Filter, Share (50% vs. full ownership)
 - Regression: Market Value, Debt
 - Heckman sample selection correction
 - Controlling for regional clustering effects
 - Maintaining variance by adding random residuals
 - Incorporating uncertainty of imputation process: Multiple imputation (k=5)

Item non-response, editing and imputation (population share affected, %)

Basis: Total Population (aged 17 and over)												
	Own Prop.	Oth. Prop.	Fin. Assets	Priv. Pens.	Bus. Assets	Tang. Assets	Total Assets	Main Prop. Debt	Oth. Prop. Debt	Oth. Debt	Total Debt	Net Worth
Observed	86,9	97,7	86,5	80,6	93,4	91,8	67,6	88,2	93,8	93,3	83,6	63,4
Inconsistency	4,0	0,4	0,0	0,0	-	-	5,4	2,3	0,2	-	3,6	5,5
INR	9,1	1,9	13,5	19,4	6,6	8,2	27,0	9,5	6,0	6,7	12,8	31,1
Total	100	100	100	100	100	100	100	100	100	100	100	100
N	23.892											
Basis: Adult Population holding wealth/debt component												
	Own Prop.	Oth. Prop.	Fin. Assets	Priv. Pens.	Bus. Assets	Tang. Assets	Total Assets	Main Prop. Debt	Oth. Prop. Debt	Oth. Debt	Total Debt	Net Worth
Observed	73,5	81,5	77,6	67,5	66,8	71,5	57,4	70,3	79,3	84,3	69,2	54,0
Inconsistency	9,8	2,9	0,1	0,0	-	-	7,1	9,7	2,9	-	9,0	7,4
INR	16,8	15,6	22,3	32,5	33,2	28,5	35,5	20,0	17,8	15,7	21,9	38,7
Total	100	100	100	100	100	100	100	100	100	100	100	100
Basic N	9.597	2.929	10.774	12.307	1.247	2.263	18.185	5.326	1.484	2.663	7.886	17.393

Correlates of INR on "Total Assets" (Heckman selection model)

- (1) Selection model → Prob(TA = 1)
 - + male, higher age, high educated, self-employed, rural
 - unemployed, pensioners
- (2) Probability model → Prob{(INR = 1) | (TA = 1)}
 - + Low education, self-employed, self-administered interview
 - male, civil servants, number of interviews



The impact of imputation on wealth inequality (individual level)

Population share holding wealth components before and after editing & imputation

	Obs. 1	Final 2	% chg. 3	% chg. aggr. wealth
Owner occ. property	31,3	36,2	+15,7	+29,6
Other Property	8,2	10,0	+22,0	+25,7
Financial Assets	36,2	43,0	+18,8	+25,3
Private Pensions	35,1	47,3	+25,8	+30,7
Business Assets	2,7	4,2	+55,6	+49,5
Tangible Assets	5,8	8,4	+44,8	+32,7
Total Assets	63,2	73,5	+16,3	+30,7
Main Property Debt	15,1	18,2	+20,5	+31,3
Other Property Debt	3,7	4,6	+24,3	+24,9
Other Debt	9,1	10,7	+17,6	+17,7
Total Debt	40,0	46,2	+15,5	+27,3

Wealth inequality before and after editing & imputation

Total Population				Population with component			
Selected comp.	Obs. 1	Final 2	% chg. 3	Obs. 1	Final 2	% chg. 3	
Owner occ. property	0,814	0,761	-6,5**	0,345	0,341	-1,3	
Gini	2,314	1,688	-27,0**	0,298	0,293	-1,6	
HSCV							
Financial Assets	0,871	0,833	-4,3**	0,637	0,612	-4,0	
Gini	12,167	8,861	-27,2**	4,019	3,527	-12,2	
HSCV							
Business Assets	0,994	0,993	-0,1**	0,783	0,825	5,5	
Gini	938,845	627,704	-33,1	24,884	25,245	1,4	
HSCV							
Other Debt	0,969	0,965	-0,4**	0,683	0,674	-1,4	
Gini	100,788	77,855	-22,8**	9,25	7,845	-15,2	
HSCV							
Net Worth	0,837	0,787	-6,0**				
Gini	6,791	14,681	116,2**				
HSCV							

Mean wealth before and after editing & imputation

Total Population				Population with component			
Selected components	Obs. 1	Final 2	% chg. 3	Obs. 1	Final 2	% chg. 3	
Owner occ. Property	38.008	57.660	+51,7**	152.360	143.546	-5,8**	
Se	296	311		734	579		
Financial Assets	8.264	11.620	+40,6**	28.066	25.768	-8,2**	
Se	159	156		456	335		
Business Assets	8.325	18.223	+118,9**	301.674	363.117	+20,4	
Se	1.001	1.609		31.939	31.907		
Other Debt	2.096	2.981	+42,2**	26.545	26.744	+0,8	
Se	171	179		1.224	1.050		

(**) indicates significant differences. Standard errors are bootstrapped (100 reps).
(1) Only those with observed personal share and value are included.
(2) After editing and imputation.
(3) (final-obs)/obs.

The choice of the aggregation unit and the household internal redistribution process

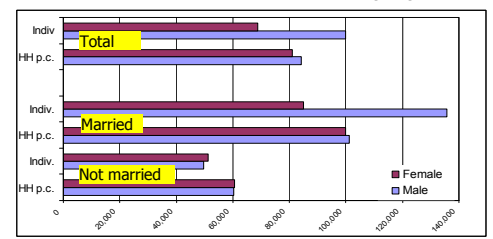
Aggregation unit and the distribution of net worth

Unit of analysis: Individual 1				% change	
	Individual wealth	HH per capita wealth	(b)-(a)/(a)	(b)-(a)/(a)	
Mean	81.713	81.797	0,1		
min	-3.692.144	-1.152.392	-68,8		
P10	0	0	0,0		
P25	0	1.897	0,0		
Median	15.000	27.500	+83,3		
P75	96.588	99.000	2,5		
P90	208.000	197.081	-5,2		
P95	313.942	290.981	-7,3		
P99	729.711	612.148	-16,1		
max	99.221.992	51.763.632	-47,8		
sum (bill €)	5.512	5.512	0,0		
FGT(0)	0,427	0,413	-3,1		
FGT(1)	0,506	0,396	-21,8**		
FGT(2)	9,216	1,065	-88,4**		

Aggregation unit and subgroup inequality (net worth)

Unit of Analysis: Individual 1				% change	
	Indiv.	HH per capita		(b)-(a)/(a)	
Gini (total)	0,756	0,697	-7,8		
by AGE					
<=24	0,968	0,746	-23,0		
25-34	0,946	0,849	-10,3		
35-44	0,754	0,708	-6,0		
45-54	0,684	0,654	-4,5		
55-64	0,645	0,617	-4,4		
65-74	0,656	0,628	-4,2		
75+	0,700	0,666	-4,9		
by SEX					
Female	0,766	0,704	-8,1		
Male	0,743	0,689	-7,3		

Net worth in Individual- and in Household-perspective



Concluding Remarks

Pre-Data Collection: Survey design *wrt* aggregation: Individual vs Household

- * significant redistribution effect within households → "gender wealth gap"
- * but: missing wealth held by children!

Post-Data Collection: Item-non-response (on wealth) is selective (no MCAR),

- but multiple imputation is an effective means to cope with selective NR
- * significant impact on share of wealth holders, mean, aggregate, inequality

Data producers must document imputation/editing & flag imputes

Outlook: Cross-national harmonization of wealth data (e.g. LWS)

- * Complete wealth measures by simulating pension entitlements
- * Need for harmonization of imputation strategy across surveys?
- * Sacrifice the "superior" information at individual level for the sake of comparability?

Full paper download: <http://www.diw.de/deutsch/produkte/publikationen/diskussionspapiere/docs/papers/dp672.pdf>

Contact: Joachim R. Frick, German Socio-Economic Panel Study (SOEP), DIW Berlin, Germany, jfrick@diw.de