

“Representative wealth data for Germany: The impact of methodological decisions around imputation and the choice of the aggregation unit”

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

*Official Statistics System (OSS) Seminar Series, Statistics New Zealand,
8 November 2007, Wellington*

Contents

1. Motivation
2. SOEP Wealth Data 2002
3. Empirical Results
 - ⇒ The impact of imputation
 - ⇒ The aggregation unit
4. Concluding Remarks


Motivation


1. Measurement error in economic outcome measures (income, wealth)
2. Survey specific decisions on ...
 - ⇒ **Pre**-data collection: How and whom to survey ?
 - Individuals versus Households
 - ⇒ **Post**-data collection: How to deal with measurement error
 - Inconsistencies → Editing
 - Item-non-response → Imputation
3. Impact of methodological and surveying decisions on substantive results (aggregates, inequality, mobility)
4. Consequences for cross-national datasets (CNEF, LWS) ?

PRE-DATA COLLECTION STAGE:

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	<input type="text"/>
Burden: If you still have a loan taken out on your house/apartment, how high is the remaining debt (excluding interest)?	EURO	<input type="text"/>
Personal share of property: Are you the sole owner (100%) or co-owner (e.g. with your spouse)?	Sole Owner	<input type="checkbox"/>
If the latter, how high is your own share?	Share in %	<input type="text"/> <input type="text"/>

POST-DATA COLLECTION STAGE: Item-Non-Response (INR) in (panel) data

- *Schräpler* (2003): complexity of surveyed construct
- *Hill & Willis* (2001): formulation of questions matters
- *Schräpler & Wagner* (2001): interviewer-respondent matching
- *Rendtel* (1995), *Riphahn & Serfling* (2003): interviewer change
- *Jarvis & Jenkins* 1998; *Biewen* 2001; *Frick & Grabka* 2005; *Riphahn & Serfling* 2005, *Hawkes & Plewis* 2006; *Wooden & Watson* 2006:
INR strongly related to Income Inequality and mobility (e.g. higher refusals in tails of income distribution)
- *Lee et al* (2004): INR and UNR not independent (to be modelled together)
- *Loosfeldt et al* (1999): INR in t = predictor of UNR in $t+1$
- *Burton et al* (1999): *Cooperation continuum*:
complete answers \rightarrow incomplete answers [INR] \rightarrow no answer at all [UNR]

POST-DATA COLLECTION STAGE: How to deal with INR ?

Rubin (1976, 1987):

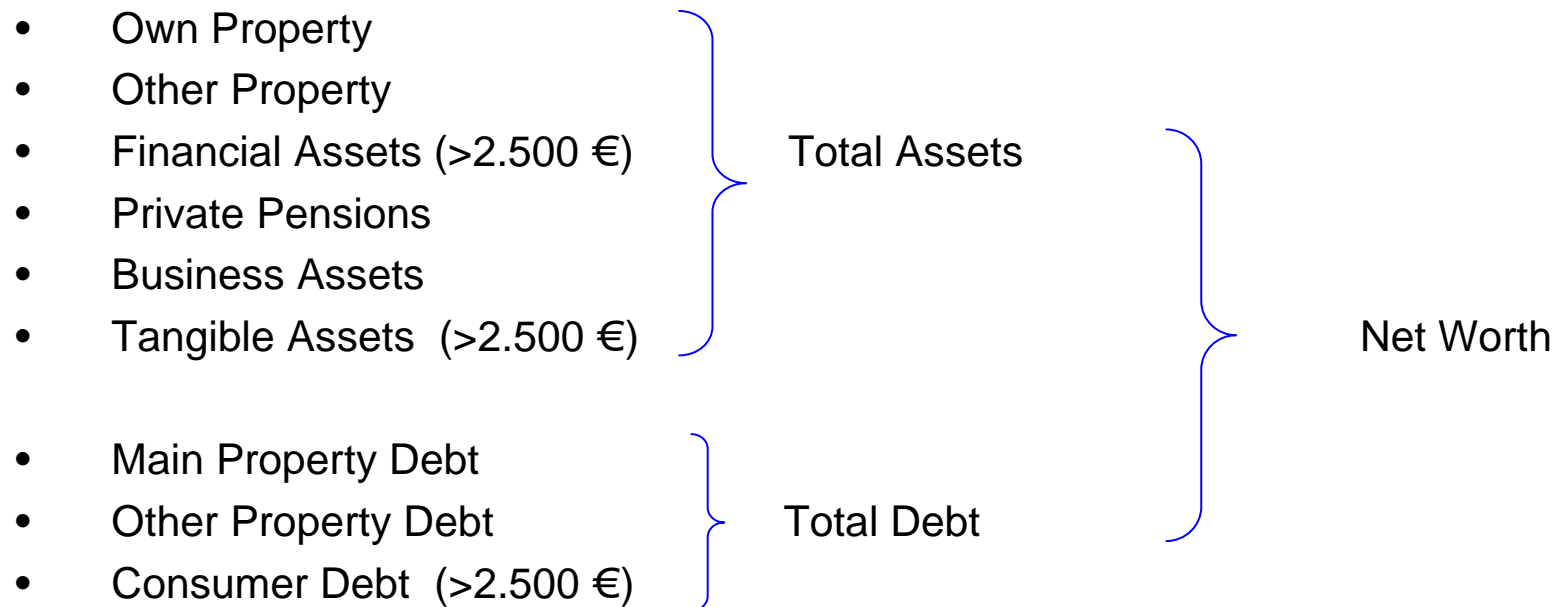
Missing mechanisms (MCAR, MAR, MNAR) and Imputation

- case-wise deletion (only valid observations)
- weighting
- Imputation (*single & multiple* imputation techniques)

Data: SOEP

SOEP Wealth module in 2002

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



Not included: cars, public pension entitlements, durables

Data: SOEP

Non-Response

- Unit-NR → weighting
- Partial-NR ~5% → imputation
- Item-NR ~15-33% → imputation

Other Measurement error

- “Inconsistencies” <10% → editing

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

<i>Basis: Total Population (aged 17 and over)</i>												
	Own Prop.	Other Prop.	Fin. Assets	Priv. Pens.	Bus. Assets	Tang. Assets	Total Assets	Main Prop. Debt	Other Prop. Debt	Other 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											
<i>Basis: Adult Population holding wealth/debt component</i>												
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 TA” (Heckman selection correction)

(1) Selection model \rightarrow Prob(TA = 1)

- + male, higher age, high educated, self-employed, rural
- unemployed, pensioners

(2) Probability model \rightarrow Prob{(INR = 1) | (TA = 1)}

- + Low education, self-employed,
self-administered interview
- male, civil servants, *number of interviews*

Principles of the Editing & Imputation (1)

Editing

- Extreme outliers (e.g., missing 1,000 digits)
- Co-owning couples (most relevant for owner-occupied property)

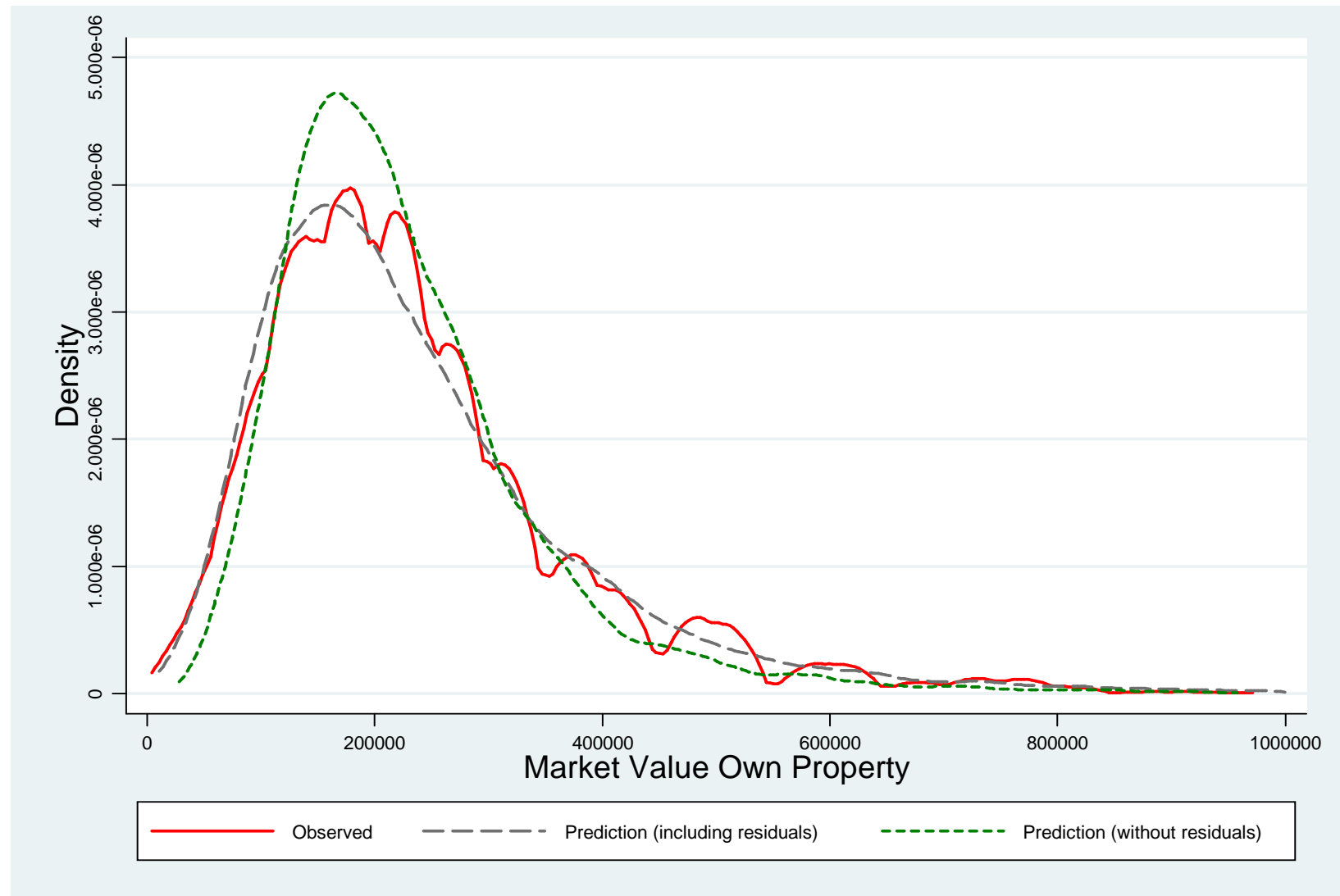
Logical Imputation

- Filter, market value, debt, personal share for co-owners
→ may be derived from non-missing information provided by one partner

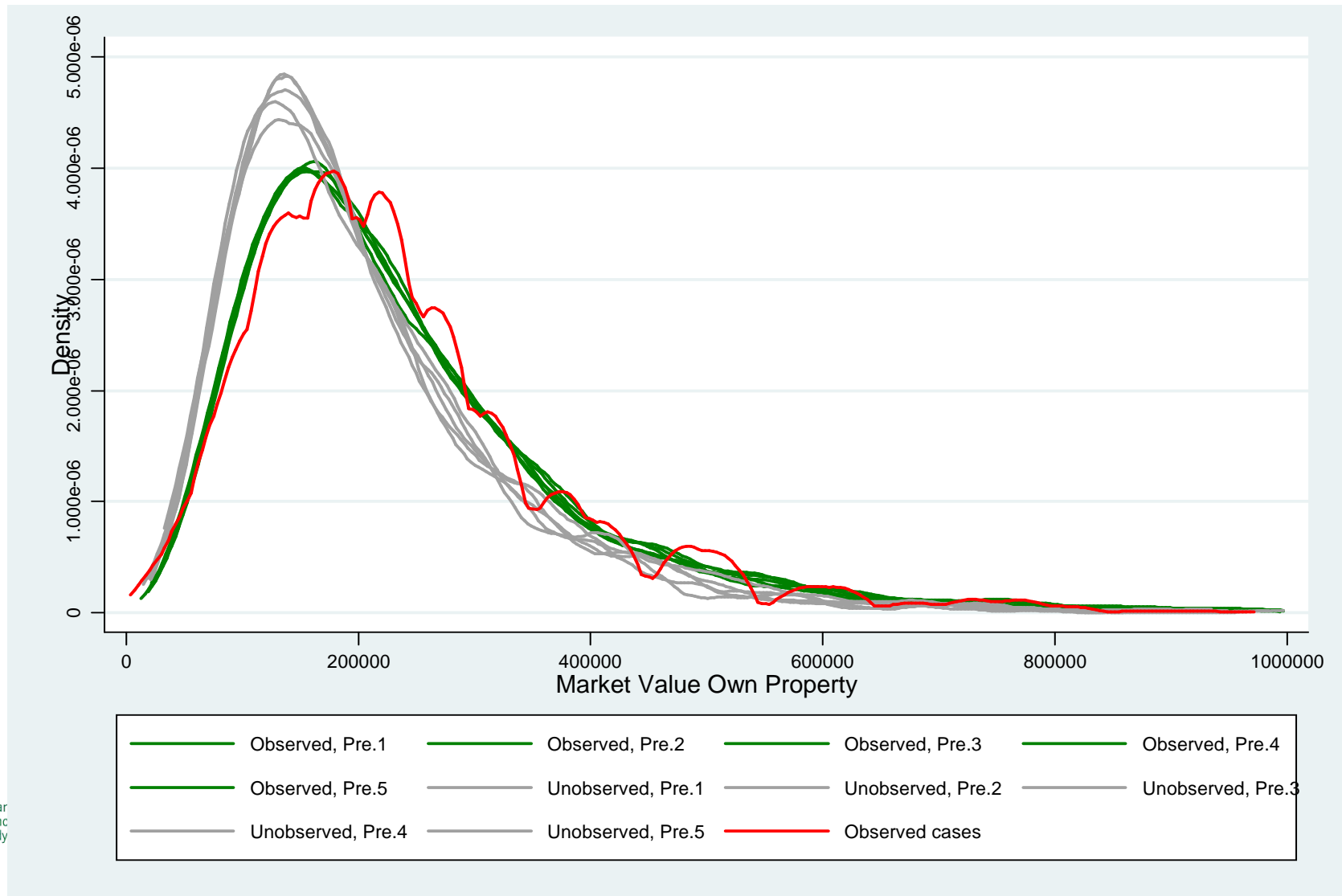
Principles of Imputation (2)

- Imputation of missing information (INR/PUNR)
 - Logit: Filter, Share (either 50% or full ownership)
 - Regression: Market Value, Debt
- Heckman sample selection model
- Controlling for regional clustering effects
(market value of private property)
- Maintaining variance by adding random residuals
(taken from the true distribution)
- Incorporating uncertainty of imputation process
→ Multiple imputation ($k=5$)

Market value for own property: Observed vs. predicted and the relevance of considering residuals



Market value for own property (PR): MI for INR and prediction for observed cases vs. imputed cases



Empirical Results

a) The impact of imputation as a means of post-survey treatment on ...

⇒ Population share holding wealth

⇒ Wealth Aggregate

⇒ Mean Wealth

⇒ Wealth Inequality

⇒ Asset Poverty

⇒ Incidence & Relevance of imputation across the wealth distribution

Population share holding wealth components before and after editing & imputation

	Observed ¹	Final ²	% change ³	% change in aggr. wealth
Owner occupied 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

Source: SOEP 2002; Population: Adult population (17 years and over) with interview

1: Only those with observed value are included.

2: After editing and imputation

3 (final-obs)/obs

Mean wealth before and after editing & imputation (individual level, weighted)

	Total Population			Population with component		
	Mean		% change ³	Mean		% change ³
	Obs. ¹	Final ²		Obs. ¹	Final ²	
Owner occupied 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	111	119		1.224	1.050	

Source: SOEP 2002: (**) 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

Wealth inequality before and after editing & imputation (individual level, weighted)

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

Source: SOEP 2002: (**) 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

Net worth and asset poverty before and after editing & imputation (individual level, weighted)

	Obs. ¹	Final ²	%change ³
Mean	60.235	81.713	35,7
sd	819	1.302	59.9
Mean if NW>0	104.466	117.812	12,8
sd	1.039	1.539	48.1
1	-20.000	-20.000	0,0
5	-3.118	-1.540	-50,6
10	0	0	---
25	0	0	---
50 (median)	5.000	15.000	200,0
75	60.000	96.588	61,0
90	174.760	208.000	19,0
95	275.000	313.942	14,2
99	600.000	729.711	21,6
Gini	0,837	0,787	-6,0
HSCV	6,791	14,681	116,2
Headcount ratio (FGT0)	0,451	0,427	-5,4
Normalized pov. gap (FGT1)	0,828	0,506	-38,9

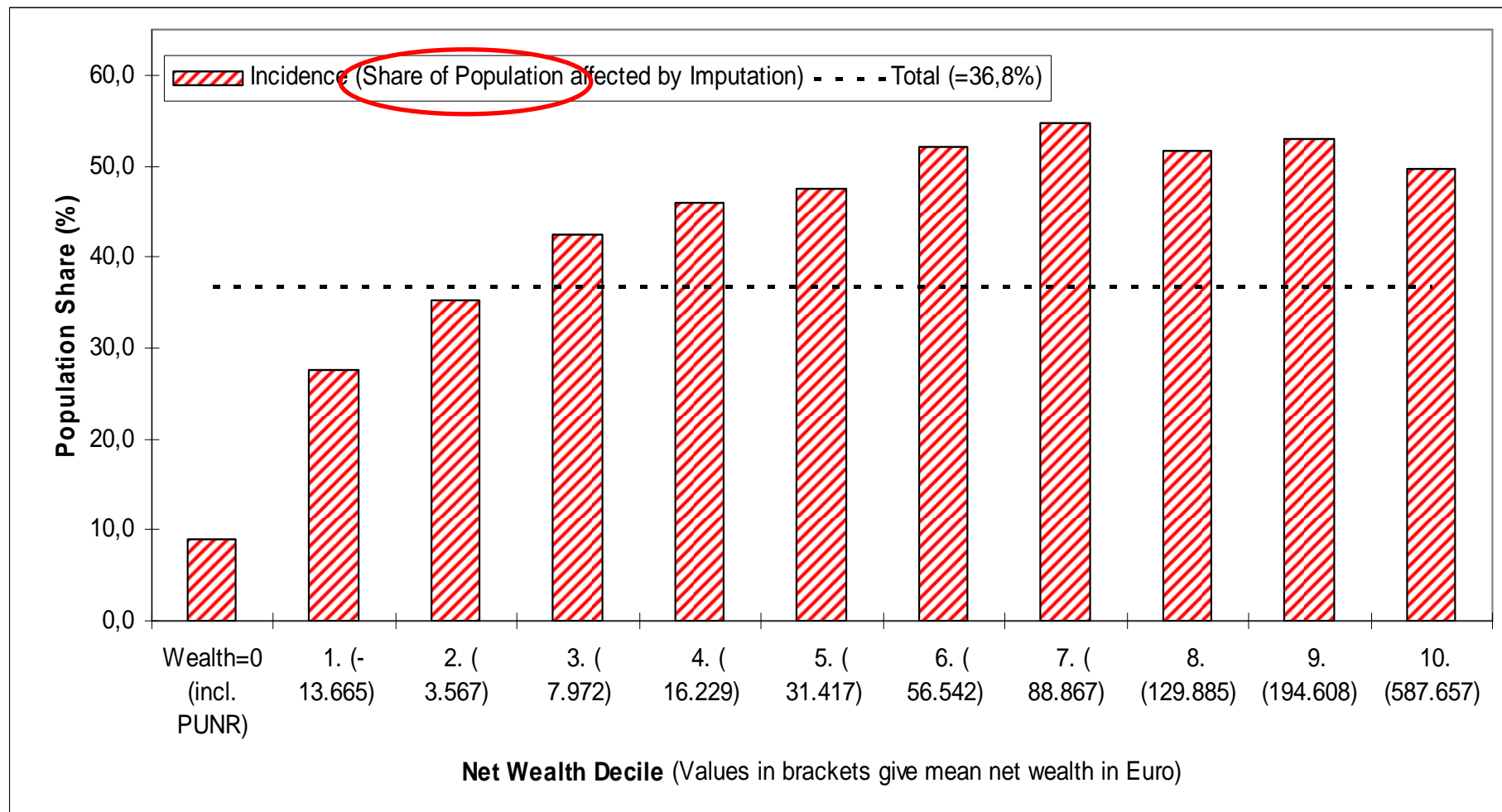
Source: SOEP 2002:

1 Only those with observed personal share and value are included.

2 After editing and imputation

3 (final-obs)/obs

Incidence of Editing & Imputation by Net Wealth Decile (conditional on imputation)



Relevance of Editing & Imputation by Net Wealth Decile (conditional on imputation)



Empirical Results

- b) The choice of the aggregation unit
 - the household internal redistribution process

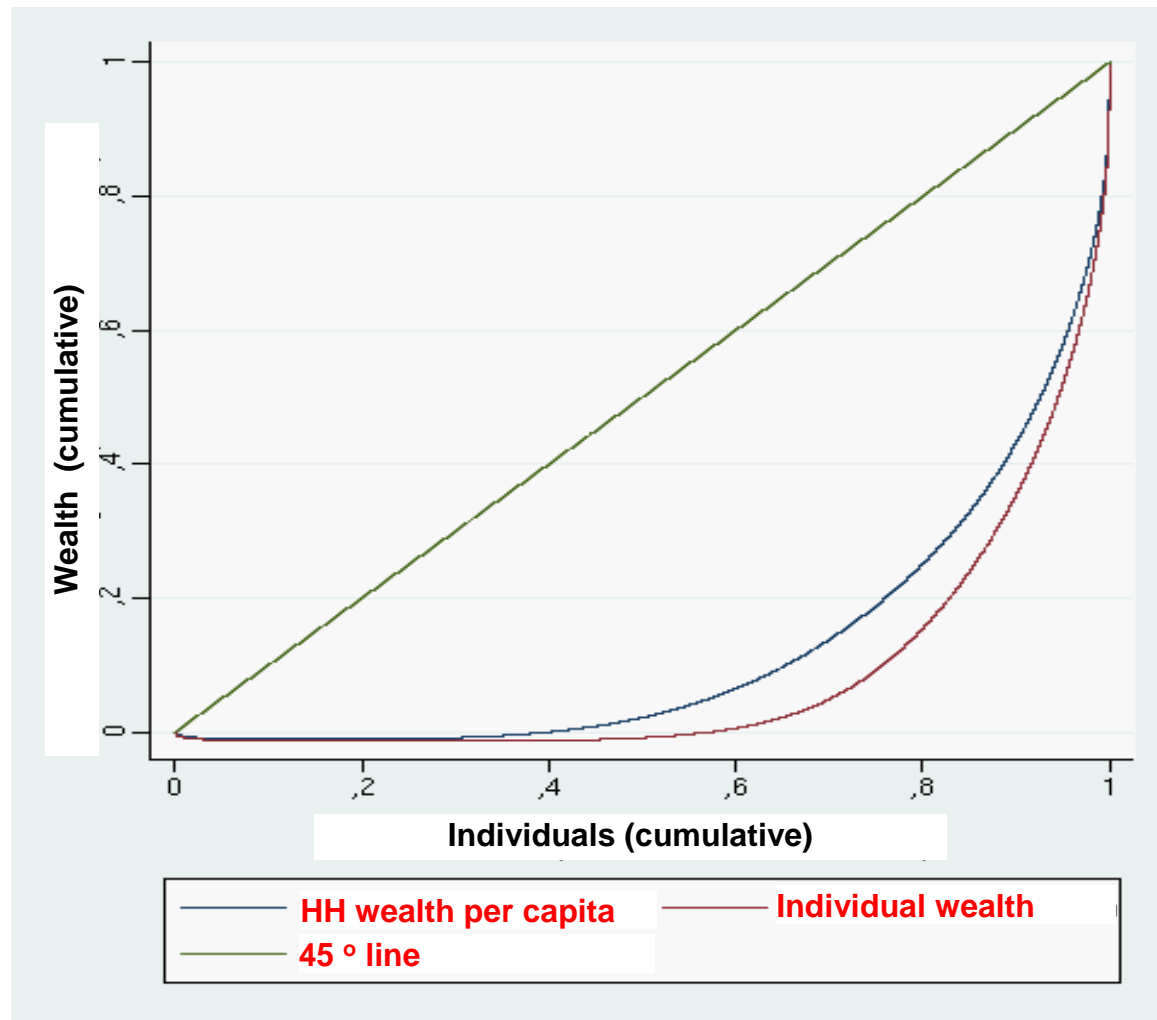
The individual as analysis unit and as aggregation unit

	Unit of analysis: Individual ¹	
	Individual wealth	
	(a)	
Mean	81.713	
min	-3.692.144	
P10	0	
P25	0	
Median	15.000	
P75	96.588	
P90	208.000	
P95	313.942	
P99	729.711	
max	99.221.992	
sum (bill €)	5,512	
FGT(0)	0,427	
FGT(1)	0,506	
FGT(2)	9,216	

Aggregation unit and the distribution of net worth → HH internal redistribution !

	Unit of analysis: Individual ¹		% change
	Individual wealth	HH per capita wealth	
	(a)	(b)	(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*

Wealth Inequality in individual and HH perspective: Lorenz curves for net worth



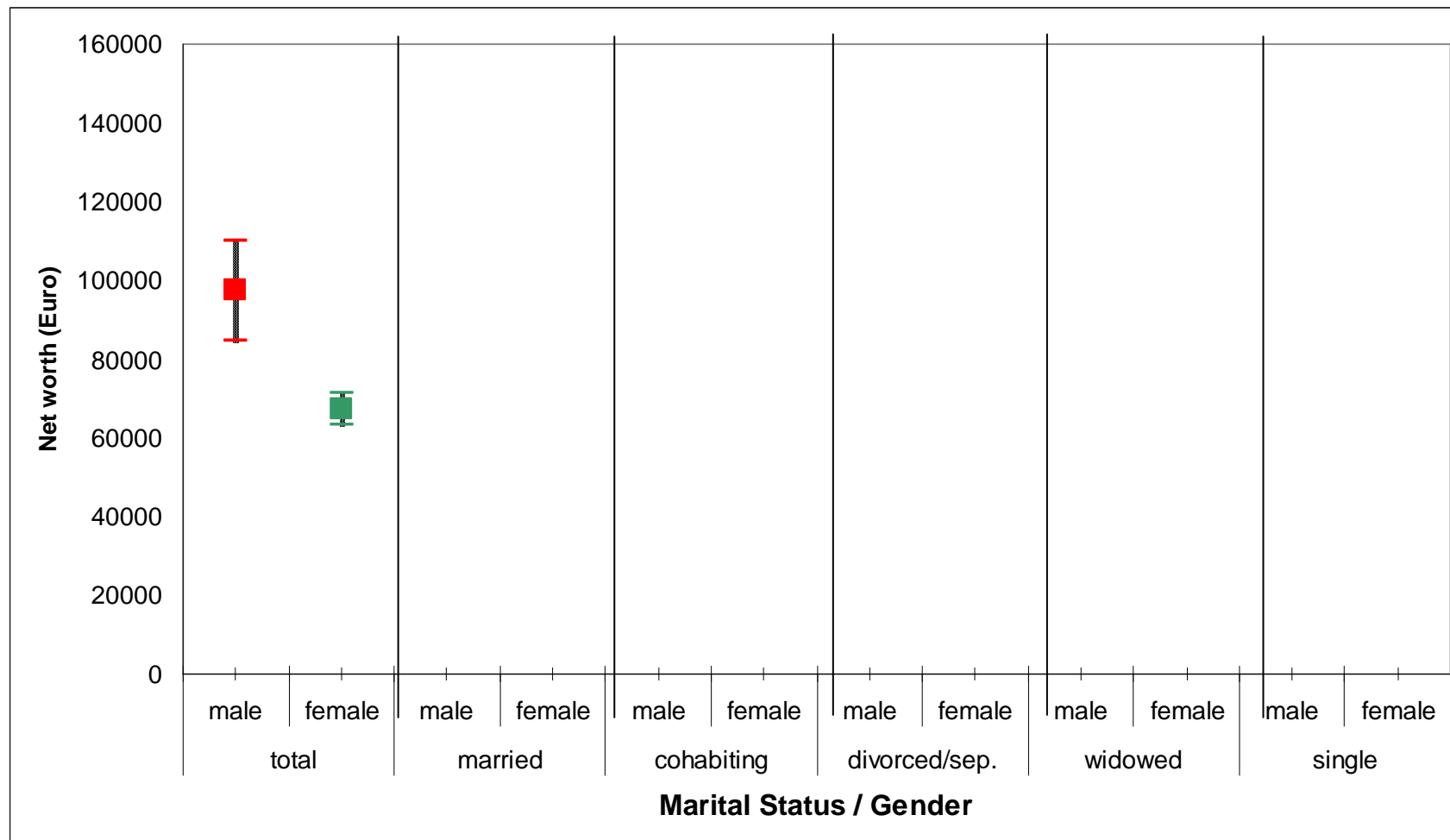
Aggregation unit and subgroup inequality (net worth)

Gini	Unit of Analysis: Individual ¹		
	Indiv.	HH per capita	% change
<i>Gini (total)</i>	0,756	0,697	-7,8*
<i>by AGE</i>			
<=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
<i>by SEX</i>			
Female	0,766	0,704	-8,1
Male	0,743	0,689	-7,3

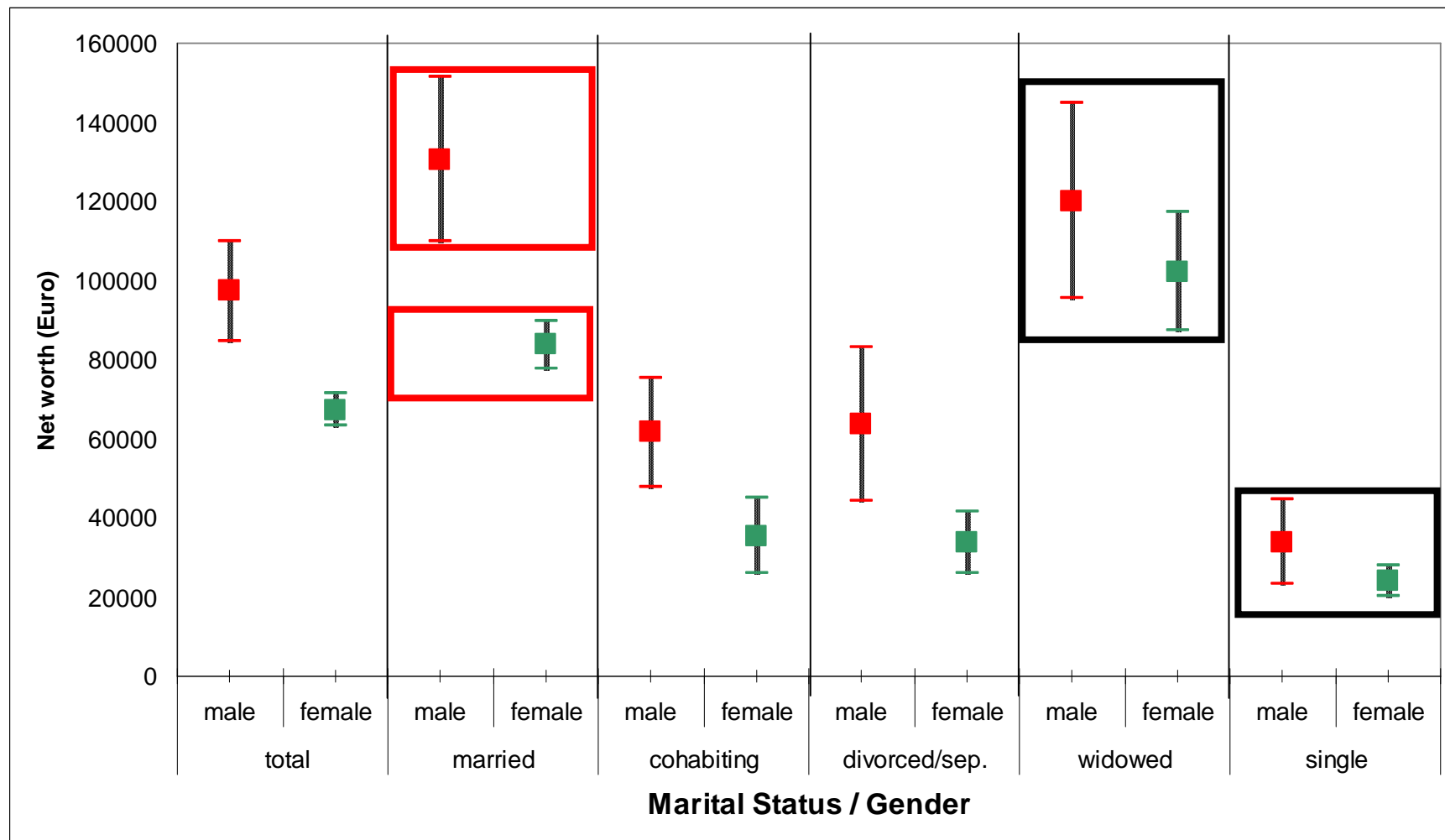
Aggregation unit and subgroup inequality (net worth)

Gini	Unit of Analysis: Individual ¹			Half SCV	Unit of Analysis: Individual ¹		
	Indiv.	HH per capita	% change		Indiv.	HH per capita	% change
Gini (total)	0,756	0,697	-7,8*	HSCV (total)	1,598	1,195	-25,2*
by AGE				by AGE			
<=24	0,968	0,746	-23,0	<=24	17,560	1,653	-90,6
25-34	0,946	0,849	-10,3	25-34	3,876	2,353	-39,3
35-44	0,754	0,708	-6,0	35-44	1,790	1,269	-29,1
45-54	0,684	0,654	-4,5	45-54	1,113	1,017	-8,6
55-64	0,645	0,617	-4,4	55-64	0,925	0,813	-12,1
65-74	0,656	0,628	-4,2	65-74	0,964	0,836	-13,3
75+	0,700	0,666	-4,9	75+	1,172	0,967	-17,4
by SEX				by SEX			
Female	0,766	0,704	-8,1	Female	1,727	1,235	-28,5
Male	0,743	0,689	-7,3	Male	1,456	1,151	-20,9

The „Gender Wealth Gap“: Net Worth by marital status and gender¹



The „Gender Wealth Gap“: Net Worth by marital status and gender¹



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 (see LWS initiative)

- complement current SOEP wealth measure by simulating pension entitlements (ongoing project)
- imputation strategy matters !
- sacrifice the “superior” information at individual level for the sake of comparability ?

!!! comments welcome !!!

Joachim R. Frick

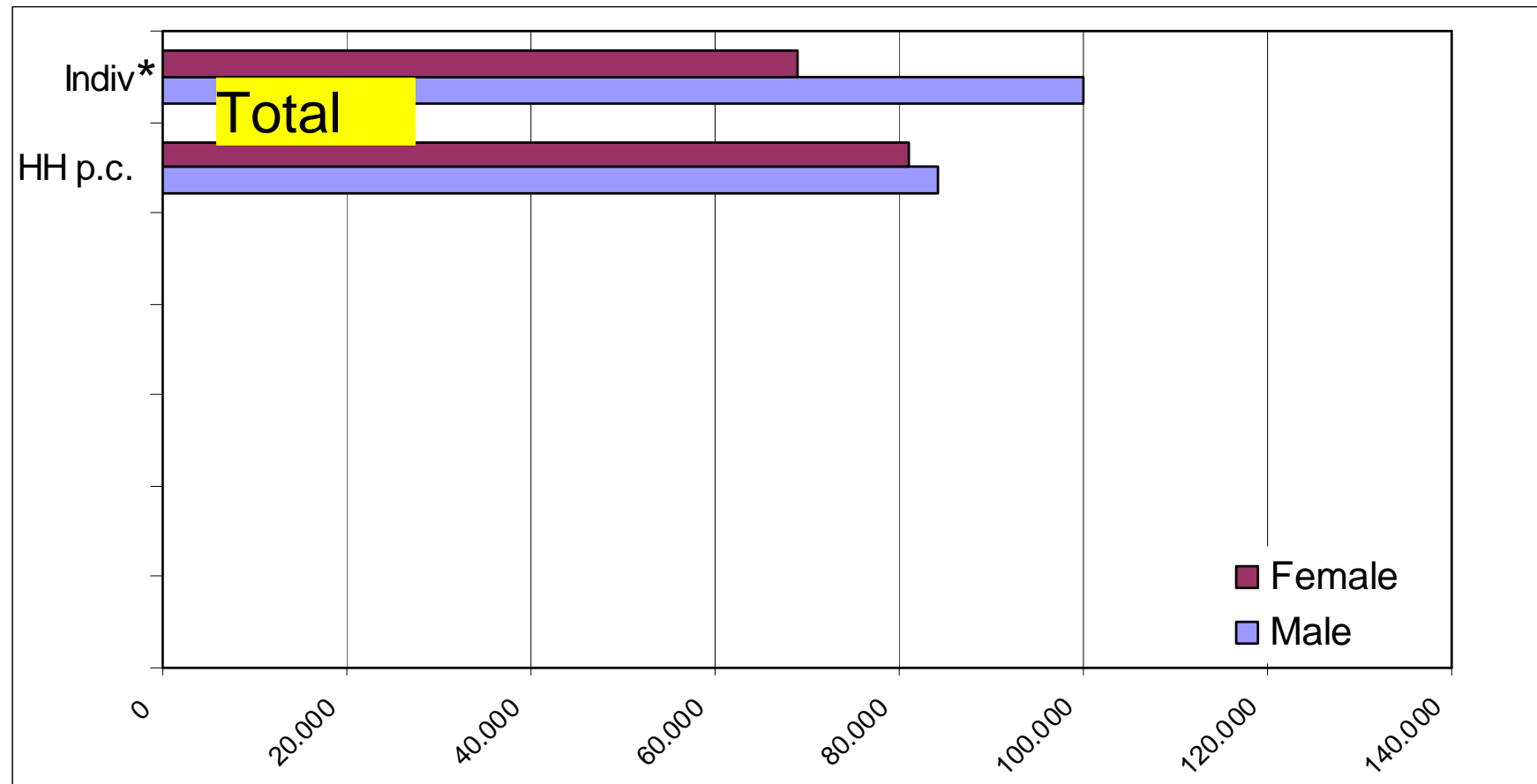
jfrick@diw.de

Appendix

Comparison of total wealth of private households with national balance sheet 2002

	National balance sheet (1)	SOEP ¹ (2)	(2) / (1) in %
Gross wealth (excluding durables)	9.025	6.493	71,9
Property	4.640	4.526	97,5
Financial assets I	3.730	1.284	34,4
Financial assets II ²	(2.630)	(1.284)	(48,8)
Net business assets ³	655	683	104,3
Liabilities ⁴	1.206	1.119	92,8
Mortgages ⁴	1.002	939	93,7
Other debts ^{4 5}	204	180	88,2
Net Wealth (excluding durables)	7.819	5.374	68,7
Net Wealth (excl.durables, based on financial assets II)	6.719	5.374	80,0
Durables ⁶	968	95	9,8

Individual vs. HH-perspective



Individual vs. HH-perspective

