Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

# Data Documentation

51

Joachim R. Frick • Markus M. Grabka •
Jan Marcus

## Editing und multiple Imputation der Vermögensinformation 2002 und 2007 im SOEP

Berlin, September 2010

IMPRESSUM



DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1861-1532

DIW BERLIN

**Data Documentation 51**

Joachim R. Frick*, Markus M. Grabka* und Jan Marcus*

# Editing und multiple Imputation der Vermögensinformation 2002 und 2007 im SOEP

Berlin, September 2010

* DIW Berlin, Das Sozio-oekonomische Panel (SOEP), jfrick@diw.de , mgrabka@diw.de, jmarcus@diw.de

# Inhaltsverzeichnis

**1 Einleitung ..... 1**

**2 Inzidenz und Selektivität von fehlenden Antwortangaben ..... 3**

2.1 Inzidenz ..... 3

2.2 Selektivität fehlender und inkonsistenter Angaben ..... 5

**3 Prinzipien von Editing und Imputation ..... 8**

**4 Editing und Imputation einzelner Vermögensarten ..... 12**

4.1 Vermögen aus Bausparverträgen oder privaten Versicherungen, Wertsachen, Betriebsvermögen und Schulden aus Konsumentenkrediten ..... 12

4.1.1 Querschnittsimputationen 2002 ..... 12

4.1.2 Längsschnittsimputationen 2002 und 2007 ..... 13

4.1.3 Imputation von Vermögen unterhalb der Geringfügigkeitsschwelle in 2002 ... 14

4.2 Selbstgenutztes Wohneigentum ..... 15

4.2.1 Konsistenzprüfungen ..... 15

4.2.2 Editing des Eigentumsanteils ..... 16

4.2.3 „Logische“ Imputationen ..... 16

4.2.4 Multiple Imputation der Markt- und Schuldenwerte ..... 18

4.3 Sonstige Immobilien ..... 22

4.3.1 Editing und „logische“ Imputationen ..... 22

4.3.2 Regressionsbasierte Imputationen ..... 23

4.4 Geldvermögen ..... 24

4.4.1 „Logische“ Imputationen ..... 24

4.4.2 Regressionsbasierte Imputationen ..... 24

**5 Einfluss der Imputation auf den Anteil der Vermögenden, die Vermögenshöhe und -verteilung ..... 26**

**6 Konventionen bei der Benennung der Variablen ..... 32**

6.1 Vermögensvariablen auf der individuellen Ebene im File PWEALTH ..... 33

6.2 Vermögensvariablen auf der Haushaltsebene im File HWEALTH ..... 35

**7 Das Arbeiten mit multiple imputierten Daten ..... 37**

**8 Literaturverzeichnis ..... 38**

# Tabellenverzeichnis

Abbildung 1: Die Messung von Vermögen im Personenfragebogen des SOEP am Beispiel des selbstgenutzten Wohneigentums ..... 8

Abbildung 2: Geschätzte Marktwerte selbstgenutzter Immobilien und die Verwendung zufallsbestimmter Residuen 2002 ..... 20

Abbildung 3: Vergleich mehrfach imputierter Fälle mit Item-Non-Response im Gegensatz zu den beobachteten Fällen 2002 ..... 21

Abbildung 4: Vergleich beobachteter und imputierter Werte für das Netto-Gesamtvermögen 2002 ..... 31

Abbildung 5: Auszug aus dem Personenfragebogen des SOEP des Erhebungsjahres 2002 ..... 39

Abbildung 6: Auszug aus dem Personenfragebogen des SOEP des Erhebungsjahres 2007 ..... 41

Tabelle 2-1: Item-Non-Response und Inkonsistenzen bei Vermögensangaben 2002 und 2007 ..... 4

Tabelle 2-2: Ergebnisse eines Probit-Modells zur Schätzung der Wahrscheinlichkeit fehlender oder inkonsistenter Antwortangaben 2002 und 2007 ..... 7

Tabelle 5.1: Der Einfluss von Editing und Imputation auf den Anteil der Population mit Vermögen und das Vermögensaggregat 2002 und 2007 ..... 27

Tabelle 5.2: Der Einfluss von Editing und Imputation auf mittlere individuelle Vermögen 2002 und 2007 ..... 28

Tabelle 5.3: Der Einfluss von Editing und Imputation auf die Ungleichheit von ausgewählten Vermögenskomponenten ..... 29

Tabelle 5.4: Der Einfluss von Editing und Imputation auf das Nettogesamtvermögen und auf relative Vermögensarmut 2002 und 2007 ..... 30

# 1 Einleitung

Dieser Artikel beschreibt die Aufbereitung der Vermögensinformationen, die im Rahmen des Sozio-oekonomischen Panels (SOEP) in den Jahren 2002 und 2007 erhoben wurden. Von besonderer Bedeutung sind dabei Umfang von und Umgang mit fehlenden sowie inkonsistenten Angaben. Daher wird im folgenden neben der Problematik der Selektivität fehlender Antwortangaben mit anschließender Imputation (Ersetzung) fehlender Informationen auch das Editing von inkonsistenten Angaben beschrieben.

Fehlende Antwortangaben sind ein weit verbreitetes Problem in Bevölkerungsbefragungen. Bestimmte Personengruppen sind weniger bereit oder in der Lage vollständige Antworten zu geben. Insbesondere bei sensitiven Informationen wie dem Einkommen oder dem Vermögen tritt verstärkt das Problem fehlender Antwortangaben auf. Dieses Antwortverhalten beruht häufig nicht auf einem Zufallsprinzip (missing at random – MAR, Rubin 1976), sondern korreliert für gewöhnlich mit bestimmten Charakteristika der Befragungspersonen als auch mit erhebungsbedingten Merkmalen. Ohne eine Korrektur des selektiven Antwortverhaltens kann es daher zu Verzerrungen bei Analysen der interessierenden Variablen (hier dem Vermögen) kommen – das Verfahren der multiplen Imputation (Rubin 1987) ist hierbei besonders geeignet, da es auch die Unsicherheit des Imputationprozesses angemessen berücksichtigt.

Im SOEP werden die folgenden Vermögenskomponenten erfragt (vgl. die entsprechenden Fragebogenmodule für 2002 und 2007 im Anhang): selbstgenutztes Wohneigentum, sonstiger Immobilienbesitz, Geldvermögen, Vermögen aus Bausparverträgen oder Lebensversicherungen[1], Betriebsvermögen und Wertsachen in Form von Gold, Schmuck, Münzen oder wertvollen Sammlungen. Als Verbindlichkeiten werden Schulden aus Hypotheken auf selbstgenutzte Immobilien sowie auf sonstige Immobilien als auch Angaben zur Höhe von Konsumentenkrediten erfasst. Nicht erhoben werden der Wert des Hausrats und von Kraftfahrzeugen sowie Anwartschaften an die diversen Alterssicherungssysteme (u. a. Gesetzliche Rentenversicherung, berufsständische Versorgungswerke, Betriebsrenten)[2].

[1] Bei der Befragung des Jahres 2002 wurde das Vermögen aus Bausparverträgen und Lebensversicherungen in einer Komponente gemeinsam erhoben; ab 2007 ist durch eine Revision des Frageschemas eine getrennte Analyse dieser Informationen möglich.

[2] Vgl. zur Berücksichtigung von Anwartschaften an Alterssicherungssysteme in der Vermögensverteilungsrechung Frick und Grabka 2010.

Im Gegensatz zu herkömmlichen Surveys erfasst das SOEP Vermögen auf der Individual-Ebene, d.h. alle erwachsenen Personen ab dem SOEP-Erstbefragungsalter von 17 Jahren werden nach ihrer persönlichen Vermögenslage befragt. Diese Befragungsmethode erlaubt die Analyse der haushaltsinternen als auch der geschlechtsspezifischen Vermögensverteilung[3]. Demgegenüber kann es zwar zu einer Untererfassung des Gesamtvermögens kommen, da das Vermögen von Minderjährigen nicht separat erfasst wird – es kann aber davon ausgegangen werden, dass dies nur zu einer marginalen Unterschätzung des Haushaltsaggregats führt.

Durch die individuelle Erfassung von Vermögensbeständen im Gegensatz zu einer haushaltsbasierten Befragung mit einer zentralen Referenzperson, besteht die Möglichkeit die Vermögenssituation eines Haushalts exakter zu erfassen, da eine einzelne Referenzperson nicht in jedem Falle über alle Anlagearten und die Höhe diese Vermögensbestände aller Haushaltsmitglieder gleichermaßen informiert sein dürfte. Zudem ist eine Antwortverweigerung durch die Referenzperson bei einer haushaltsbasierten Befragung gleichbedeutend mit fehlenden Angaben für *alle* Haushaltsmitglieder. Zwar birgt die individuelle Befragung das Problem selektiver Teilnahme einzelner Haushaltsmitglieder, diese sind unter Verwendung der Angaben Dritter gegebenenfalls aber besser zu imputieren als dies bei komplett fehlenden Werten der Fall wäre. Nachteilig erscheint ebenfalls die Tatsache der Möglichkeit inkonsistenter Angaben von Haushaltsmitgliedern zu gemeinsam gehaltenen Vermögensbeständen (z.B. bei Paaren mit jeweils 50% Anteil am gemeinsamen Wohneigentum). Faktisch sind solche Inkonsistenzen oder Messfehler aber nur aufgrund der individuellen Erhebung überhaupt feststellbar – und nach intensiver Datenprüfung und Editing (siehe Abschnitte 3 und 4) sind diese Angaben in unserer Interpretation deutlich verlässlicher als Informationen, die nur mittels einer Referenzperson für den Haushalt als Ganzes gesammelt wurden.

Im folgenden Abschnitt 2 wird zunächst die Selektivität fehlender Antwortangaben zu vermögensbezogenen Fragen beschrieben. Anschließend werden grundlegende Prinzipien der Aufbereitung der Vermögensdaten des SOEP erläutert (Abschnitt 3), bevor die Vorgehensweise für die einzelnen Vermögenskomponenten in Abschnitt 4 detailliert beschrieben wird. Abschnitt 5 behandelt den Einfluss von imputierten bzw. editierten Angaben auf Verteilungsergebnisse. Abschnitt 6 stellt die Bezeichnungen für die generierten Vermögensvariablen auf

3 Vgl. z.B. Sierminska, Frick & Grabka (2010).

Haushalts- und Personenebene vor. Abschließend enthält Abschnitt 7 Nutzerhinweise zum Umgang mit den multipel imputierten Daten des SOEP.

# 2 Inzidenz und Selektivität von fehlenden Antwortangaben

## 2.1 Inzidenz

Wie in allen Befragungen üblich, ist die Erhebung ökonomisch relevanter Größen wie Vermögen und Einkommen auch im SOEP mit Messfehlern und –problemen behaftet. Von besonderer Bedeutung in diesem Zusammenhang sind fehlende Werte aufgrund des Auslassens bestimmter Fragebogenteile („Item non-response“, INR) bzw. aufgrund von „Partial Unit non-response“ (PUNR). Das letztgenannte Phänomen beschreibt den Komplettausfall eines Individualinterviews in einem ansonsten befragungswilligen Haushalts mit mehreren Befragungspersonen. In diesen Fällen führt eine Aggregation der Vermögensangaben aller Haushaltsmitglieder definitionsgemäß zu einer Untererfassung des Gesamtvermögens eines Haushalts. Von diesem Phänomen sind in 2002 fast 1,200 Haushalte und in der Erhebung des Jahres 2007 rund 1,250 Fälle betroffen.[4]

Aufgrund der individuellen Erfassung der Vermögenssituation kann es zu inkonsistenten Angaben von Personen innerhalb eines Haushalts kommen, die eine bestimmte Komponente gemeinsam besitzen. Dieses Phänomen tritt vorrangig bei Ehepaaren mit gemeinsamem Immobilienbesitz auf. Mit dem SOEP-Fragebogen wird eine Selbsteinschätzung des Marktwertes einer Immobilie sowie des individuellen Besitzanteils erhoben. Idealerweise sollte der von allen Eigentümern jeweils genannte Marktwert übereinstimmen und die Summe der Besitzanteile sollte sich, unter der Annahme, dass es keine weiteren Miteigentümer außerhalb des Haushaltes gibt, entsprechend auf 100% addieren. Jede Abweichung von dieser Regel muss als Messfehler erkannt und entsprechend durch eine spezifische Form von „Editing“ korrigiert werden. Fehlende Werte aufgrund von INR (inkl. eventueller „Weiss nicht“-Angaben) werden hier hingegen mit Hilfe entsprechender Annahmen und Verfahren imputiert werden.

Der Anteil der Personen mit fehlenden bzw. inkonsistenten Vermögensinformationen schwankt je nach Komponente für die Gesamtpopulation zwischen 11,1% beim Betriebsver-

4 Für Fälle von Partial-uni non response, die nicht in Privathaushalten leben, wurde keine Imputation vorgenommen.

mögen und 19,7% beim Geldvermögen. Bei den Verbindlichkeiten reichen die entsprechenden Werte von 10,7% für Verbindlichkeiten bei sonstigen Immobilien bis zu 15,9% bei Hypotheken auf selbst genutztem Wohneigentum (Tabelle 2-1). Betrachtet man nur die Population derjenigen, die eine bestimmte Vermögenskomponente halten, so steigt definitionsgemäß der Anteil derer, die editiert oder imputiert werden müssen.[5]

Zwischen 2002 und 2007 sinkt der Anteil fehlender Antwortangaben bei allen Vermögenskomponenten. Dies kann als Beleg dafür gewertet werden, dass die Befragten aufgrund der Wiederholungsmessung mit dem inhaltlichen Fokus der Befragung vertrauter sind und sich ggf. entsprechend auf die Befragung vorbereiten (z.B. durch das Bereithalten oder vorherige Einsehen von Depotauszüge oder ähnlichem). Auch sind durch die verbesserte Vertrauenssituation die Befragten eher bereit einer dritten Person gegenüber entsprechende Auskünfte zu geben.

Tabelle 2-1: Item-Non-Response und Inkonsistenzen bei Vermögensangaben 2002 und 2007
(Anteil in %)

| | **2002** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Selbst genutzte Immobilie** | **Sonstige Immobilien** | **Geldvermögen** | **Private Versicherungen[2]** | **Betriebsvermögen** | **Wertsachen** | **Bruttogesamtvermögen** | **Hypotheken auf selbst genutzte Immobilien** | **Hypotheken auf sonstige Immobilien** | **Konsumentenkredite** | **Verbindlichkeiten insgesamt** | **Nettovermögen** |
| **Beobachtet** | **82,7** | **88,5** | **80,3** | **76,7** | **88,9** | **86,7** | **59,6** | **84,1** | **89,3** | **87,9** | **78,8** | **57,7** |
| **Inkonsistent** | **3,8** | **0,3** | **-** | **-** | **-** | **-** | **3,6** | **2,2** | **0,2** | **-** | **3,5** | **4,1** |
| **INR[1]** | **13,5** | **11,2** | **19,7** | **23,3** | **11,1** | **13,3** | **36,8** | **13,7** | **10,5** | **12,1** | **17,7** | **38,2** |
| **Insgesamt** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** |
| | **2007** | | | | | | | | | | | |
| **Beobachtet** | **82,3** | **90,0** | **83,2** | **77,3** | **90,8** | **91,1** | **60,1** | **85,3** | **90,9** | **91,1** | **81,2** | **58,2** |
| **Inkonsistent** | **5,1** | **0,5** | **-** | **-** | **-** | **-** | **4,7** | **2,4** | **0,1** | **-** | **4,0** | **5,1** |
| **INR[1]** | **12,6** | **9,5** | **16,8** | **22,7** | **9,2** | **8,9** | **35,2** | **12,3** | **9,0** | **8,9** | **14,8** | **36,7** |
| **Insgesamt** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** | **100** |

[1]: INR = Item-non-response, [2]: inklusive Bausparguthaben, in 2007 separat erhoben.
Basis: Gesamtpopulation (Personen im Alter von 17 und mehr Jahren): n = 25.091 (2002), n = 22.467 (2007)
Quelle: SOEP 2002 und 2007; Eigene Berechnungen.

5 Im Regelfalle wird bei der Imputation ein Wert größer Null generiert, ex-ante kann aber nicht ausgeschlossen werden, dass der Imputationsprozess, der ja ggf. auch die Imputation von Filterfragen mit einschließt, einen Wert von Null erzeugt.

## 2.2 Selektivität fehlender und inkonsistenter Angaben

Zur Identifikation eventueller Selektivität von Messfehlern (aufgrund von Inkonsistenz und Item-Non-Response) sowie der dadurch induzierten Notwendigkeit für Editing und Imputation werden auf Basis gepoolter Daten der beiden Erhebungsjahre 2002 und 2007 Schätzungen zur Wahrscheinlichkeit für derartige Messprobleme durchgeführt. Dabei muss die potentielle Selektivität in den Status „Eigentümer von Vermögenskomponente X“ zu sein, berücksichtigt werden. Beispielsweise ist die Wahrscheinlichkeit, Wohneigentum (und eventuell damit verbundene Schulden) zu halten, für gut ausgebildete Individuen höher. Diese Selektivität kann durch eine Korrektur der Selektionsverzerrung nach Heckman (1979) berücksichtigt werden.

Für jede zu erhebende Vermögenskomponente wurde ein Modell für die Wahrscheinlichkeit fehlender bzw. inkonsistenter Angaben geschätzt, wobei neben inhaltlich motivierten Kovariaten auch der Einfluss der Interviewsituation kontrolliert wird (siehe Tabelle 2-2).[6] Zentrale Ergebnisse des Selektionsmodells sind dabei:

- mit zunehmendem Alter steigt die Wahrscheinlichkeit, Vermögen (und Schulden) zu halten signifikant an (dieser Effekt flacht jedoch mit hohem Alter wieder ab). Ein derartig positiver Zusammenhang findet sich ebenso bei Verheirateten, Personen mit hohem Bildungsniveau und Beamten. Zudem nimmt mit steigendem Markteinkommen auch erwartungsgemäß die Wahrscheinlichkeit für das Vorhandensein sowohl von Aktiva als auch von Passiva zu. Dagegen ist die Wahrscheinlichkeit für das Halten von Vermögenskomponenten negativ korreliert mit den Merkmalen Migrationshintergrund, niedriges Bildungsniveau, Großstadt und insbesondere Arbeitslosigkeit.
- Bezüglich der Interviewmethode finden sich deutliche Effekte: Alle Personen, die nicht mit der Standardmethode PAPI („paper-and-pencil interview“) interviewt werden, haben eine höhere Wahrscheinlichkeit, Vermögen oder Schulden zu halten, was für eine entsprechende Selbstselektion in bestimmte Befragungsformen spricht. Die Zahl der in der bisherigen „SOEP-Karriere“ gegebenen Interviews weist dagegen keine klare Tendenz auf.

Bezüglich der Wahrscheinlichkeit für „Inkonsistenz bzw. Item-Non-Response“, gegeben die obige Selektionskontrolle, finden sich folgende zentralen Ergebnisse:

6 In diesen Schätzungen sind Fälle mit partiellem Unit-Non-response (PUNR) nicht enthalten, da deren Ausfallwahrscheinlichkeit anderen Mustern folgt als bei Item-Non-response oder Inkonsistenzen. Zudem ist die quantitative Bedeutung von PUNR im Vergleich zu INR und Inkonsistenzen relativ gering. Zur Selektivität von PUNR bei Einkommensmessungen bzw. -analysen vergleiche Frick, Grabka und Groh-Samberg (2009).

- Die Wahrscheinlichkeit für fehlende bzw. inkonsistente Werte ist für männliche Interviewpartner geringer.
- Höhere Bildungsabschlüsse reduzieren die Wahrscheinlichkeit derartiger Messprobleme, während geringe Bildung dieses Risiko erhöht.
- Für Beamte erhalten wir den erwarteten negativen Effekt, der auf eine erhöhte Wahrscheinlichkeit vollständiger und konsistenter Informationen in dieser Berufsgruppe hinweist.
- Für Selbständige scheint es besonders schwierig zu sein, vollständige Auskunft über den Wert ihres Vermögens zu geben, was oft auch der Schwierigkeit einer adäquaten Bewertung des Betriebsvermögens geschuldet ist.[7]
- Wiederum tragen Variablen bezüglich der Interview-Situation in signifikanter Art zur Varianzaufklärung bei: insbesondere Daten aus selbständig durchgeführten Interviews (ohne Präsenz eines Interviewers) müssen häufiger imputiert und editiert werden. Personen, die an der SOEP Befragung im Jahre 2007 teilnehmen, haben eine höhere Wahrscheinlichkeit, vollständige und konsistente Informationen über das Vermögen zu geben als dies in der ersten Befragung 2002 der Fall war. Dies ist ein Anzeichen für einen positiven Lerneffekt im Rahmen von Wiederholungsbefragungen und den Aufbau einer Vertrauensbeziehung zwischen Befragten und Interviewer (zur zunehmenden Datenqualität von Einkommensangaben im Rahmen wiederholter SOEP-Interviews vgl. Frick et al 2006). Dies steht auch im Einklang mit einer abnehmenden Wahrscheinlichkeit für die hier relevanten Messprobleme bei steigender Zahl von gegebenen Interviews. Letztlich zeigt sich bei CAPI-Interviews eine deutlich höheres Risiko für fehlende bzw. inkonsistente Werte. Dieser Befund ist aber das Ergebnis der spezifischen Interviewsituation mit Hilfe eines Laptops, welches explizit – falls bspw. ein Befragter keine metrische Angabe zum Marktwert einer Vermögenskomponente machen kann oder möchte – Nachfragen induziert („unfolding brackets“), mit der die Größenordnung des Vermögens in bestimmten Kategorien erfasst werden kann. In diesen Fällen liegt zwar keine direkte Messung eines metrischen Wertes vor, die entsprechenden Schwellenwerte erlauben aber eine qualitativ deutlich bessere Imputation der fehlenden Angabe.

[7] In den vorliegenden SOEP-Daten kann nicht zwischen fehlenden Werten aufgrund von fehlender Kenntnis („Weiss ich nicht“) bzw. mangelnder Auskunftsbereitschaft („Sage ich nicht“) unterschieden werden.

Tabelle 2-2: Ergebnisse eines Probit-Modells zur Schätzung der Wahrscheinlichkeit fehlender oder inkonsistenter Antwortangaben 2002 und 2007

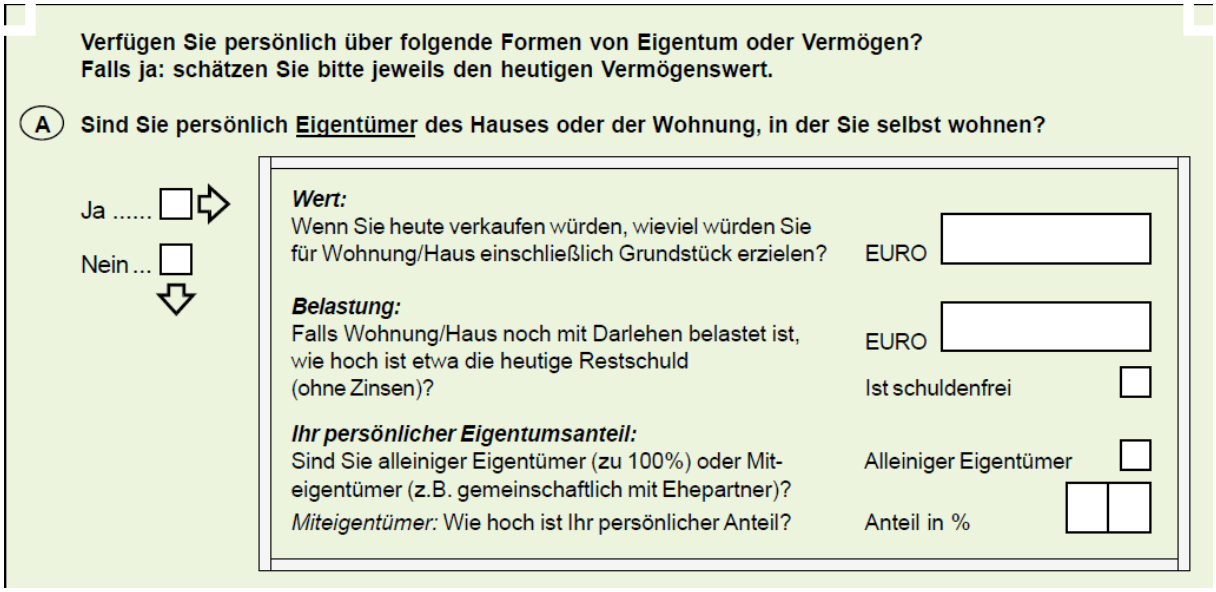
| | Selbst genutzte Immobilie | Hypotheken auf selbst genutzte Immobilien | Sonstige Immobilien | Hypotheken auf sonstige Immobilien | Geldvermögen | Private Versicherungen | Betriebsvermögen | Wertsachen | Konsumentenkredite |
|---|---|---|---|---|---|---|---|---|---|
| | **Wahrscheinlichkeit für Item-non-response oder Inkonsistenz** | | | | | | | | |
| **Alter** | -0,0314*** | -0,0253* | 0,0036 | 0,0167 | -0,0170*** | -0,0230*** | 0,0276 | -0,0184* | 0,0271 |
| **Alter²** | 0,0002*** | 0,0002* | 0,0000 | -0,0002 | 0,0002*** | 0,0002** | -0,0002 | 0,0002** | -0,0003 |
| **Mann** | -0,1068*** | -0,1118*** | -0,1457*** | -0,2034*** | -0,0184 | -0,1592*** | -0,0353 | -0,0848* | -0,0915 |
| **Migrant** | 0,1151*** | -0,1128** | 0,0216 | -0,0801 | -0,1187** | 0,1175*** | 0,0767 | -0,1321 | -0,0534 |
| **Ausbildungsniveau[1] niedrig** | 0,1472*** | 0,0171 | 0,1169 | -0,1736 | -0,0357 | 0,1004*** | -0,0991 | -0,0851 | -0,002 |
| **Ausbildungsniveau[1] hoch** | -0,1385*** | -0,1607*** | -0,0776 | 0,0008 | -0,0416 | -0,1601*** | -0,0993 | 0,0674 | -0,1088** |
| **HH mit Kindern <14Jahre** | -0,0486 | -0,0411 | -0,0829 | -0,1088 | -0,0875*** | -0,0082 | -0,0349 | -0,2218*** | -0,0439 |
| **1989 in DDR gelebt** | 0,2142*** | 0,0233 | 0,1713** | -0,1200 | -0,0815*** | -0,1522*** | -0,0872 | -0,1598* | -0,0461 |
| **Selbständig[2]** | -0,0274 | -0,0038 | 0,1688** | 0,2025** | 0,1030*** | -0,0034 | 1,0776*** | 0,0886 | 0,2502*** |
| **Beamte** | -0,0606 | -0,1377*** | 0,0535 | -0,1110 | -0,0783* | -0,1464*** | -0,3992 | 0,0718 | -0,0466 |
| **Zahl der Interviews** | -0,0027* | -0,0056*** | -0,0063** | -0,0088* | -0,0118*** | -0,0073*** | -0,0075* | -0,0033 | 0,0007 |
| **Selbstausfüller[3]** | 0,3166*** | 0,0947** | 0,2369*** | 0,0748 | -0,0428 | -0,0891*** | 0,1461* | 0,0046 | 0,1221* |
| **Capi Interview** | 0,6787*** | 0,5273*** | 0,5383*** | 0,2532*** | 0,6135*** | 0,6153*** | 0,4944*** | 0,7455*** | 0,5445*** |
| **Interview postalisch** | 0,5377*** | 0,4217*** | 0,4886*** | 0,4354*** | 0,1366*** | 0,1398*** | 0,2964*** | 0,3107*** | 0,2727*** |
| **Jahr 2007** | -0,0271 | -0,0949*** | -0,1467*** | -0,1963*** | -0,2013*** | -0,0605*** | -0,1269** | -0,6431*** | -0,6449*** |
| **Konstante** | 0,1853 | 0,2643 | -1,4020** | -1,6792** | -0,4866*** | 0,3989** | -2,5758*** | -0,9852** | -2,1651*** |
| | **Wahrscheinlichkeit für das Halten von Vermögenskomponente ...** | | | | | | | | |
| **Alter** | 0,1138*** | 0,1412*** | 0,0877*** | 0,1052*** | 0,0016 | 0,0616*** | 0,0538*** | 0,0256*** | 0,0867*** |
| **Alter²** | -0,0008*** | -0,0014*** | -0,0007*** | -0,0009*** | 0,0001*** | -0,0007*** | -0,0005*** | -0,0001*** | -0,0010*** |
| **Mann** | 0,0023 | 0,0103 | 0,0512** | 0,0348 | 0,0660*** | 0,1634*** | 0,4248*** | -0,0771*** | 0,1655*** |
| **Verheiratet** | 0,6634*** | 0,5908*** | 0,1799*** | 0,1084*** | 0,1736*** | 0,1863*** | 0,0068 | 0,1146*** | -0,0455** |
| **Migrant** | -0,5245*** | -0,1619*** | -0,2328*** | -0,2641*** | -0,5509*** | -0,4891*** | -0,3080*** | -0,2400*** | 0,0952*** |
| **Ausbildungsniveau[1] niedrig** | -0,2997*** | -0,2969*** | -0,1599*** | -0,1980*** | -0,2939*** | -0,3221*** | -0,1212** | -0,1929*** | -0,1732*** |
| **Ausbildungsniveau[1] hoch** | 0,1268*** | 0,0757*** | 0,2586*** | 0,2463*** | 0,4371*** | 0,0969*** | 0,0992** | 0,1979*** | -0,1229*** |
| **HH mit Kindern <14Jahre** | 0,3136*** | 0,4020*** | 0,0634** | 0,1173*** | -0,0821*** | 0,0034 | 0,2063*** | -0,0537** | 0,0865*** |
| **Dorf** | 0,4895*** | 0,2203*** | 0,0571** | -0,0985*** | 0,0271 | 0,0557*** | 0,1081*** | -0,0483* | -0,0575** |
| **Großstadt** | -0,4624*** | -0,2793*** | -0,0564* | -0,0795** | 0,0352 | -0,0911*** | -0,1371** | 0,1967*** | 0,0395 |
| **1989 in DDR gelebt** | -0,5428*** | -0,2398*** | -0,3295*** | -0,4388*** | -0,1183*** | 0,0517*** | -0,1382*** | -0,4168*** | 0,2237*** |
| **Arbeitslos** | -0,3783*** | -0,4322*** | -0,2170*** | -0,2410*** | -0,5128*** | -0,6460*** | -0,2472** | -0,1395*** | -0,1081*** |
| **Selbständig[2]** | 0,0978*** | 0,0415 | 0,3804*** | 0,3262*** | -0,0726** | -0,0900*** | 2,3027*** | 0,2138*** | 0,2121*** |
| **Rentner** | -0,1354*** | -0,1696*** | 0,0158 | -0,1144** | 0,1549*** | -0,4533*** | -0,2017*** | 0,0699* | -0,1678*** |
| **nicht erwerbstätig** | -0,1125*** | -0,2111*** | 0,0247 | -0,0747* | -0,0645*** | -0,5170*** | -0,0062 | 0,1917*** | -0,3458*** |
| **Beamte** | 0,2342*** | 0,2380*** | 0,0617 | -0,0134 | 0,1176*** | 0,1553*** | -0,3202*** | 0,1192*** | 0,0822** |
| **Zahl der Interviews** | 0,0001 | 0,0029** | -0,0052*** | -0,0072*** | 0,0007 | 0,0023** | -0,0065** | -0,0113*** | 0,0015 |
| **Selbstausfüller[3]** | 0,0823*** | 0,1310*** | 0,2139*** | 0,2612*** | 0,0863*** | 0,2162*** | 0,1288*** | 0,1613*** | 0,2933*** |
| **Capi Interview** | 0,0684*** | 0,0758*** | 0,1058*** | 0,0797** | 0,0821*** | 0,2314*** | 0,0472 | 0,1375*** | 0,1659*** |
| **Interview postalisch** | 0,0402 | 0,1052*** | 0,2100*** | 0,2539*** | 0,0699*** | 0,2453*** | 0,1926*** | 0,2356*** | 0,3732*** |
| **Markteinkommen[4]** | 0,0040*** | 0,0024*** | 0,0068*** | 0,0063*** | 0,0062*** | 0,0029*** | 0,0039*** | 0,0042*** | -0,0014*** |
| **Jahr 2007** | 0,0189* | 0,0052 | -0,0451*** | -0,0502*** | 0,0587*** | 0,0931*** | -0,0663** | -0,2917*** | 0,1711*** |
| **Konstante** | -4,0956*** | -4,5027*** | -4,1548*** | -4,6306*** | -0,6298*** | -1,1556*** | -3,9496*** | -2,2917*** | -2,9589*** |
| **athrho** | 0,2783*** | -0,1928*** | 0,2278* | 0,3433** | 0,2270** | -0,2652*** | 0,5571** | 0,6011*** | 0,5654* |
| **LR test on indep. equations** | 33,95 | 7,259 | 3,312 | 4,575 | 4,839 | 14,04 | 5,43 | 1 0,64 | 3,55 |
| **Number of obs.** | 45107 | 45107 | 45107 | 45107 | 45107 | 45107 | 45107 | 45107 | 45107 |
| **Censored obs.** | 26931 | 35277 | 39672 | 42460 | 22989 | 21333 | 42850 | 41305 | 38796 |
| **Uncensored obs.** | 18276 | 9830 | 5435 | 2647 | 22118 | 23774 | 2257 | 3802 | 6311 |
| **Wald chi2(15)** | 781,81 | 35 0,5 | 163,5 | 63,32 | 1287 | 1714 | 137,9 | 575,7 | 332,4 |
| **Log lik (full model)** | -33097 | -24564 | -16785 | -9527 | -38893 | -39874 | -5703 | -13752 | -18877 |

Quelle: SOEP gepoolte Informationen der Jahre 2002 und 2007; Probit-Modell inklusive Selektionskorrektur nach Heckman.
[1]: RF: Ausbildungsniveau mittel; [2]: RF: abhängig beschäftigt; [3]: RF: Paper and Pencil Interview; [4]: bedarfsgewichtet. *** p<0.01, ** p<0.05, * p<0.1.

# 3 Prinzipien von Editing und Imputation

Dieser Abschnitt erläutert die generelle Vorgehensweise bei der Aufbereitung der Vermögensdaten des SOEP der Erhebungsjahre 2002 und 2007. Diese Aufbereitung beinhaltet neben extensivem Editing zur Korrektur von Inkonsistenzen insbesondere die Behandlung fehlender Werte mittels multipler Imputation. Fehlende Antwortangaben können zum einen nur bei einem einzelnen Erhebungsmerkmal auftreten (*item non-response; INR*), zum anderen aber auch aus „Partial Unit non-repsonse“ (PUNR) resultieren, also durch die komplette Verweigerung einer Befragungsperson in einem ansonsten befragungswilligen Haushalt.

Abbildung 1: Die Messung von Vermögen im Personenfragebogen des SOEP am Beispiel des selbstgenutzten Wohneigentums




Die Messung von Vermögensbeständen im SOEP erfolgt in der Regel in einem mehrstufigen Prozess (Abbildung 1). Zu Beginn wird jede Befragungsperson gefragt, ob sie Eigentümer einer bestimmten Vermögensform ist. Wird diese Filterinformation mit „Ja“ beantwort, so erfolgt im Anschluss eine Frage nach der Höhe des aktuellen Marktwertes und der Höhe eventuell ausstehender Restschulden. Da die Erhebung der Vermögenskomponenten im SOEP auf individueller Ebene erfolgt, wird zudem abschließend der individuelle Vermögensanteil erfasst.

Editing wird bei inkonsistenten Antwortangaben angewendet. Inkonsistenzen liegen z.B. dann vor, wenn bei einer von einem Ehepaar gemeinsam gehaltenen Immobilie unterschiedlich

hohe Marktwerte oder eine unterschiedlich hohe Restschuld angegeben wird. Dabei kann im SOEP zu Zwecken der Plausibilitäts- und Konsistenzprüfung auf weitere Informationen aus dem Haushaltsfragebogen oder auf Angaben weiterer Haushaltsmitglieder zurückgegriffen werden.[8]

Die Imputation fehlender Angaben zum Vermögen wird im SOEP je nach Art der zu imputierenden Information mit verschiedenen Methoden betrieben. Hierzu zählen „logische" Imputationen, Imputationen auf Basis logistischer Regressionen und auf Basis von OLS-Regressionen mit Selektionskorrektur nach Heckman.

Das (deduktive) Ersetzen einer fehlenden Information durch das Ableiten von Informationen anderer Haushaltsmitglieder oder aus Informationen aus dem Haushaltsfragebogen wird hier als „logische" Imputation bezeichnet. Diese Art der Ersetzung findet beispielsweise dann Anwendung, wenn bei einer gemeinsam von einen Ehepaar gehaltenen Immobilie für einen der Partner der Marktwert der Immobilie fehlt, dieser aber von dem anderen Ehepartner direkt angegeben wurde. Hier wird unterstellt, dass die gegebene Information valide ist und dementsprechend dem Partner mit der fehlenden Antwortangabe direkt zugewiesen kann.

Die Imputation der Vermögensangaben beginnt grundsätzlich mit der Filterfrage zum Besitz einer bestimmte Vermögenskomponente. Liegt die Information zur Filterfrage nicht vor, so wird mit Hilfe einer logistischen Regression zunächst die Wahrscheinlichkeit bestimmt, eine entsprechende Vermögenskomponente zu halten.[9] Überschreitet dieser Schätzwert eine „Wahrscheinlichkeits"-Schwelle[10], so wird unterstellt, dass Vermögen in Form dieser Komponente gehalten wird – andernfalls liegt kein Vermögensbesitz vor.

Die Imputation fehlender metrischer Werte, d.h. der Höhe des Marktwertes einer Vermögenskomponente oder der ausstehenden Restschuld, wird mit Hilfe einer OLS-Regression mit Selektionskorrektur nach Heckman vorgenommen (vgl. Heckman 1979). Die Selektionskorrektur berücksichtigt, dass die Information über die Höhe des Markt-/Schuldenwerts einer

8 Editing wird eingesetzt für Vermögensinformationen zum selbstgenutzten und sonstigen Immobilienbesitz angewendet, da hierzu auch ausreichende Informationen aus dem Haushaltsfragebogen vorliegen.

9 Eine Ausnahme bildet das selbstgenutzte Wohneigentum, wo Informationen anderer Haushaltsmitglieder und Informationen aus dem Haushaltsfragebogen zur Ableitung der fehlenden Filterausprägung verwendet werden.

10 Um die Unsicherheit der Prädiktion zu berücksichtigen, wird keine feste Schranke verwendet, sondern zufällige Werte aus einer Normalverteilung mit dem Mittelwert 0,5 und einer Standardabweichung von 0,2. Diese Normalverteilung hat die Eigenschaft, dass ca. 99.98% der Werte im Intervall von 0 bis 1 liegen und dass Werte um den Mittelwert 0.5 entsprechend häufiger vorkommen.

bestimmten Vermögenskomponente nur für diejenigen Personen vorliegen kann, die auch tatsächlich Besitzer einer gegebenen Vermögenskomponente sind. Als Selektionsvariablen werden hier u. a. der Beamtenstatus sowie die Lebenszufriedenheit verwendet. Zudem erfolgt in allen Regressionsmodellen eine Berücksichtigung von regionalen Clustereffekten, die aufgrund des SOEP-Stichprobendesigns notwendig erscheint.[11] Als letzte Information wird der individuelle Vermögensanteil imputiert. Hierfür wird wiederum eine logistische Regression angewendet.[12]

Da die Höhe des Vermögens über die Zeit (d.h. in den Jahren 2002 und 2007) hinweg stark korreliert, werden bei der Imputation der Höhe des Marktwertes bzw. der Restschuld einer Vermögens-/Schuldenkomponente die Vermögensangaben wechselseitig in den Regressionsmodellen berücksichtigt, d.h. Vermögensinformationen aus 2002 werden genutzt, um fehlende Vermögensangaben aus dem Jahr 2007 zu imputieren und umgekehrt. Somit werden iterativ je drei Regressionsmodelle berechnet, in denen wechselseitig bereits imputierte Informationen des jeweils anderen Erhebungsjahres als zusätzliche Kovariate genutzt werden. Als Startmodell dient das rein auf Querschnittsdaten basierte Modell für 2002; diese Werte werden dann in der Schätzung für 2007 genutzt und letztlich sind die Ergebnisse des zweiten Modells wiederum Input für eine erneute Schätzung für 2002 – nun aber unter Berücksichtigung längsschnittlicher Informationen. Dies hat insbesondere Auswirkungen auf die mittels dieser Daten gemessene Vermögensmobilität, da bei einer reinen querschnittlichen Imputation nicht die Vermögensakkumulation in Abhängigkeit vom Vermögensbestand im Ausgangsjahr berücksichtigt werden kann.

Im Anschluss werden zufällig Residuen aus einer Prädiktion für die vollständig beobachteten Fälle gezogen und zu den mittels der Heckman-Korrekurmodellen vorhergesagten Werte für die Population mit fehlenden Werten addiert. Diese Varianz–erhaltende Maßnahme beugt regression-to-the-mean Effekten vor und berücksichtigt auch die Unsicherheit des Imputationsverfahrens. Das zufällige Zuweisen von Residuen wird insgesamt fünfmal durchgeführt,

11 Das SOEP ist eine mehrfach geschichtete Zufallsstichprobe, die regional geclustert ist und insofern ggf. zu einer Unterschätzung der tatsächlichen Variation einer interessierenden Information (z.B. Marktwert von selbstgenutztem Wohneigentum) führt. Dieses Clustern ergibt sich vorrangig durch die Optimierung des Interviewereinsatzes in der Feldphase.

12 Da nahezu ausschließlich entweder ein persönlicher Anteil von 50% oder 100% angegeben wird, schätzen wir dieses Modell als bivariates Probit, wobei die abhängige Variable angibt, ob eine Person einen hälftigen Eigentumsanteil hält. Geschätzte Eigentumsanteile von mehr als 50% werden auf 100% aufgerundet, ansonsten wird ein hälftiger Eigentumsanteil unterstellt.

und ergibt einen multipel imputierten Datensatz mit fünf verschiedenen Werte – so genannten Implicates.

Um die Belastung der Befragten gering zu halten wurde in der SOEP-Erhebung des Jahres 2002 eine untere Erfassungsschwelle für ausgewählte Vermögenskomponenten in Höhe von 2500 Euro verwendet. Dies betraf das Geldvermögen, Wertsachen und Schulden aus Konsumentenkrediten, d.h. erst oberhalb dieser Schwelle wurde der Marktwert bzw. die Höhe der Restschuld erfragt. Dies hatte sowohl eine Untererfassung der Zahl der Personen im Besitz dieser Komponenten als auch einen Bias im Hinblick auf das entsprechende Aggregat zur Folge. In einer Revision[13] des Fragebogens der Erhebung des Jahres 2007 wurde diese Erfassungsschwelle aufgehoben. Um Verzerrungen im intertemporalen Vergleich zu vermeiden, wird für die Befragten des Jahres 2002 eine rückwirkende Imputation der drei Vermögenskomponenten mit Hilfe eines zwei-stufigen Schätzmodells vorgenommen. Für jede der drei Komponenten wird auf Basis der 2007er Population für alle Personen, die maximal 2500 Euro dieser Komponente halten, ein logistisches Modell zur Wahrscheinlichkeit des Besitzes dieser Komponente geschätzt. Die Parameter dieser Schätzung werden unter Annahme der Konstanz der Zusammenhänge im Fünf-Jahres-Zeitraum auf die entsprechende Population des Befragungszeitpunkts 2002 übertragen, soweit diese die entsprechende Komponente nicht besaßen. Konditioniert auf das Ergebnis dieser out-of-sample Prädiktion wird allen 2002 Befragten ein zufällig gezogener („wahrer“) Wert aus der in 2007 erhobenen Verteilung zugewiesen..

Die genaue Vorgehensweise des Editing und der Imputation bei den einzelnen Vermögenskomponenten wird in den folgenden Abschnitten beschrieben. Zunächst wird das Verfahren bei den Komponenten vorgestellt, bei denen weder Editing noch „logische“ Imputationen vorgenommen werden. Im Anschluss wird der Aufbereitungsprozess beim selbstgenutzten Wohneigentum beschrieben und abschließend die Vorgehensweise bei den sonstigen Immobilien und dem Geldvermögen erläutert.

13 Die Revision des Fragebogens des Jahres 2007 umfasste auch eine nun getrennte Erfassung des Bausparguthaben und des Rückkaufswerts von privaten Versicherungen, die in 2002 in einer gemeinsamen Kategorie erfragt wurden.

# 4 Editing und Imputation einzelner Vermögensarten

## 4.1 Vermögen aus Bausparverträgen oder privaten Versicherungen, Wertsachen, Betriebsvermögen und Schulden aus Konsumentenkrediten

Das grundsätzliche Prinzip der multiplen Imputation der Vermögensinformationen im SOEP wird hier am Beispiel des Vermögens aus Lebensversicherungen beschrieben. Dieses Vorgehen findet auch Anwendung beim Betriebsvermögen, den Wertsachen, dem Bausparguthaben und den Konsumentenkrediten.

### 4.1.1 Querschnittsimputationen 2002

Für jede der Vermögenskomponenten Rückkaufswert privater Versicherungen, Wertsachen, Betriebsvermögen und Konsumentenkredite wird zunächst die Imputation der Filtervariablen für das Jahr 2002 durchgeführt. Die Imputation einer fehlenden Angabe zur Filterfrage basiert auf einer logistischen Regression: Anhand verschiedener Variablen auf der Personen- und Haushaltsebene wird die Wahrscheinlichkeit bestimmt, ob eine Person die jeweilige Komponente hält. Dabei werden unterschiedliche Regressionsmodelle für die Population mit INR bzw. mit PUNR berechnet, da für letztere nur ausgewählte Individualinformationen zur Verfügung stehen.

Die geschätzte Wahrscheinlichkeit wird wiederum mit einem Zufallswert aus einer Normalverteilung mit Mittelwert 0.5 und Standardabweichung 0.2. verglichen. Ist die Wahrscheinlichkeit größer als die Zufallszahl, wird angenommen, dass die Person die Vermögenskomponente besitzt. Ist die Wahrscheinlichkeit kleiner als der Referenzwert, so hat das Individuum diese Komponente nicht.

Für das Betriebsvermögen wird zusätzlich eine logistische Regression durchgeführt, um die Eigentümerstrukturen dieser Vermögensart zu berücksichtigen, da im SOEP danach gefragt wird, ob eine Person alleiniger Eigentümer ist oder nur beteiligt an z.B. einer GBR, GmbH oder KG. Diese Unterscheidung ist notwendig, da der Eigentumsanteil eine wichtige Kovariate für eine regressionsbasierte Imputation von fehlenden Angaben bezüglich des Marktwertes des Betriebsvermögens ist.

Anschließend wird ein auf einem Maximum Likelihood Prinzip basierendes OLS-Regressionsmodell mit Selektionskorrektur nach Heckman verwendet, um die metrischen Informationen zur Höhe des Markt-/Schuldenwertes der Vermögenskomponenten für 2002 zu imputieren.

### 4.1.2 Längsschnittsimputationen 2002 und 2007

Die Qualität von Imputationen ist zum einen abhängig von der Imputationsmethode (vgl. z.B. Starick & Watson 2007) zum anderen aber auch von der zur Verfügung stehenden Information. Spieß und Goebel (2003) weisen darauf hin, dass die Qualität einer Imputation durch die Verwendung von Längsschnittinformationen im Imputationsprozess signifikant verbessert wird. Daher wird für die Imputation von Vermögensangaben im SOEP wechselseitig Vermögensinformationen aus der anderen Erhebungswelle genutzt. So wird z.B. für die Imputation sowohl der Filtervariablen als auch des Markt-/Schuldenwertes des Jahres 2007 berücksichtigt, ob eine Person bereits im Jahre 2002 diese Vermögenskomponente besaß und wie hoch das Vermögen/Schulden war.[14]

Als Startwert für die längsschnittbasierte Imputation des Jahres 2007 werden die direkt beobachteten Vermögensinformationen des Erhebungsjahres 2002 verwendet.[15] In einem weiteren Schritt werden fehlende Angaben aus dem Jahre 2002 erneut imputiert, wobei nun Informationen sowohl für die beobachteten als auch die imputierten Informationen des Jahres 2007 als Kovariate genutzt werden. Dieser Vorgang wird weitere vier Mal für beide Erhebungsjahre wiederholt, um eine Konvergenz der Ergebnisse zu erzielen.[16] Zu den sich letztlich ergebenen Schätzwerten werden per Zufallsverfahren Residuen aus den Regressionsanalysen addiert, um die Unsicherheit des Imputationsverfahrens zu berücksichtigen, da die Varianz imputierter Werte auch trotz der Verwendung von Längsschnittinformationen für gewöhnlich

14 2007 wurden Bausparverträge und private Versicherungen separat erfragt, während sie 2002 als gemeinsame Vermögensposition geführt wurden. Bei der Imputation für 2007 wird für Bausparverträge und private Versicherungen jeweils der gesamte Wert beider Komponenten verwendet, da eine Disaggregation nicht möglich ist.

15 Für einige Regressionen in 2007 (Filter bei Geldvermögen, private Versicherungen, Betriebsvermögen und Schulden sowie beim metrischen Wert des Betriebsvermögens), wird zudem das durchschnittliche Haushaltseinkommen von 2002 bis 2007 berücksichtigt, da hierdurch der Anteil der erklärten Varianz zunimmt und Dynamiken über die Zeit abgebildet werden können.

16 Imputierte Vermögen/Schulden aus dem Jahr 2002 unterhalb der in 2002 geltenden Geringfügigkeitsschwelle in Höhe von 2.500 Euro, werden ab dem 2. Iterationsprozess zur Imputation des Markt-/Schuldenwertes ebenso herangezogen..

geringer ist als die der beobachteten Werte („regression to the mean"-Effekt).[17] Die Zuweisung von zufällig gezogenen Residuen wird insgesamt fünf Mal wiederholt, so dass ein multipel imputierter Datensatz zur Verfügung steht.

### 4.1.3 Imputation von Vermögen unterhalb der Geringfügigkeitsschwelle in 2002

In der Erhebung des Jahres 2002 wurde zur Reduktion der Belastung für die Befragten eine untere Erfassungsschwelle für das Geldvermögen, Wertsachen und für Konsumentenkredite in Höhe von 2.500 Euro verwendet. Diese Vorgehensweise führt zu einer systematischen Unterschätzung des Aggregats und der Zahl der Personen, die diese Vermögensart halten. Daher wurde im Rahmen einer Revision des Fragebogens für die Erhebung des Jahres 2007 die untere Erfassungsschwelle wieder verworfen. Es stehen damit Informationen im Jahre 2007 über die Charakteristika von Personen zur Verfügung, die ein Vermögen unterhalb von 2.500 Euro je Vermögenskomponente halten. Es wird dabei unterstellt, dass sich die Strukturen für den Besitz eher geringer Vermögen zwischen den beiden Erhebungsjahren nicht verändert haben.

Für alle Personen, für die in 2002 keine Angaben zum Geldvermögen, zu Wertsachen oder Konsumentenkrediten vorliegen, wird mittels einer logistischen Regression die Wahrscheinlichkeit bestimmt, ob diese die entsprechende Vermögenskomponente im Wert von weniger als 2.500 Euro besaßen. Diese Regression wird auf Basis der Population geschätzt, die in 2007 berichten, entweder kein oder weniger als 2.500 Euro je Vermögensart zu besitzen.[18]

Die vorhergesagte Wahrscheinlichkeit wird wiederum mit einem zufälligen Wert einer Normalverteilung (Mittelwert 0.5, Standardabweichung 0.2) verglichen. Ist die vorhergesagte Wahrscheinlichkeit kleiner als der zufällig gezogene Schwellenwert, so wird festgelegt, dass die Person keine dieser Vermögens-/Schuldenkomponente hält, andernfalls wird angenommen, dass ein geringfügiges Vermögen/Schulden im Jahre 2002 vorgelegen hat.

Die Höhe dieser Vermögens-/Schuldenkomponente wird bestimmt, indem Werte der Personen unterhalb der Schwelle von 2.500 Euro aus der Erhebung des Jahres 2007 gezogen und per Zufallsverfahren den entsprechenden Fällen in 2002 zugewiesen werden. Diese Vorge-

17 Um den Einfluss von extremen Ausreißern zu minimieren, wird die Verteilung der Residuen bei den Perzentilen 0.5 und 99.5 getrimmt.

18 Personen mit fehlenden Antwortangaben bezüglich der Filterinformation oder des Marktwertes des Vermögens/Schulden in 2007 blieben bei diesem Vorgehen unberücksichtigt.

hensweise wird fünfmal wiederholt um zum einen die Unsicherheit des angewendeten Imputationsverfahrens zu berücksichtigen und somit einen multipel imputierten Datensatz zu erhalten.

## 4.2 Selbstgenutztes Wohneigentum

### 4.2.1 Konsistenzprüfungen

Die quantitativ bedeutendste Vermögenskomponente der Privathaushalte in Deutschland ist das selbstgenutzte Wohneigentum. Für Plausibilitäts- und Konsistenzprüfungen stehen umfangreiche Angaben von anderen Haushaltsmitgliedern oder Informationen aus dem Haushaltsfragebogen zur Verfügung. So wird selbstgenutzter Immobilienbesitz für gewöhnlich z.B. von Ehepaaren gemeinschaftlich gehalten. Dadurch ist es möglich, Antwortangaben der Befragten auf inhaltliche Konsistenz zu prüfen. So sollten alle Eigentümer im selben Haushalt für das gemeinschaftlich selbstgenutzte Wohneigentum den gleichen Verkehrswert angeben; und mehrere Eigentümer in einem gegebenen Haushalt können zusammen nicht mehr als 100% ein und derselben Immobilie besitzen. Es liegen zudem Längsschnittinformationen vor, die zur Konsistenzprüfung der Angaben über die Zeit herangezogen werden können.

Bei geringfügigen Abweichungen zwischen zwei Personen in einem Haushalt hinsichtlich der Markt- und Schuldenwerte wird der Mittelwert aus beiden Angaben zugewiesen.[19] Sind die Unterschiede größer, wird im Einzelfall geprüft, ob bei einer Angabe möglicherweise Dezimalstellen zu viel oder zu wenig angegeben worden sind. Dazu werden u.a. Informationen zur Art, Größe, Renovierungsbedürftigkeit und Lage des Wohneigentums, ebenso wie Angaben zum Haushaltseinkommen, zum Einzugsjahr sowie zum Markt- und Schuldenwert aus anderen Erhebungswellen genutzt.[20] Gibt eine Person an „schuldenfrei“ zu sein und eine andere im selben Haushalt berichtet einen positiven Schuldenwert, wird nach den fallweisen Prüfungen i.d.R. der positive Schuldenwert beiden Personen zugewiesen, da anzunehmen ist, dass kleinere Schuldensummen fälschlicherweise leichter vergessen werden.

19 Als geringfügig wird eine Abweichung dann bezeichnet, wenn der kleinere Wert nicht weniger als 2/3 des größeren Wertes ausmacht.

20 Eine vollständige Auflistung aller Entscheidungsregeln kann von den Autoren für alle Vermögenskomponenten zur Verfügung gestellt werden.

Weiterhin werden alle selbstgenutzten Wohnungen und Häuser mit einem Marktwert von weniger als 10.000 Euro oder mit einem Schuldenwert von weniger als 2.000 Euro auf fehlende Dezimalstellen überprüft. Einzelprüfungen werden auch vorgenommen, falls eine Restschuld den Marktwert um das Doppelte übersteigt.

Sind die im Haushaltsfragebogen genannten monatlichen Zins- und Tilgungszahlen (SOEP-Variable SH32 in 2002 bzw. XH29 in 2007) ähnlich hoch wie der Schuldenwert des selbstgenutzten Wohneigentums (d.h. um nicht mehr als 20% abweichend), so wird der Schuldenwert auf „keine Angabe“ (also missing) gesetzt und anschließend imputiert, da hier davon auszugehen ist, dass die Frage zur Restschuld falsch interpretiert wurde.

### 4.2.2 Editing des Eigentumsanteils

Bei den Eigentumsanteilen ist zu berücksichtigen, dass mehrere Personen im Haushalt zusammen nicht mehr als 100% an ihrem selbstgenutzten Wohneigentum besitzen können. Daher werden Angaben von Personen, die angeben alleiniger Eigentümer zu sein, während eine andere Person im Haushalt angibt einen Anteil zu besitzen, korrigiert: Es wird hier der an 100% fehlende Teil des anderen Partners zugeordnet. Geben beide Partner an, das Wohneigentum zu 100% zu besitzen, wird davon ausgegangen, dass sie zusammen 100% besitzen, also jeder einen hälftigen Anteil besitzt. Wird das Wohneigentum zu mehr als 100% besessen (wobei keiner der Befragten angibt alleiniger Eigentümer zu sein), werden anhand von fallweisen Prüfungen Werte für beide Eigentümer ermittelt. Geben beispielsweise Eltern und ihre erwachsenen Kinder an Eigentümer zu sein und die Anteilssumme ist größer als 100, so wird der Besitz nach fallweiser Prüfung (insbesondere unter Berücksichtigung der Altersstruktur im Haushalt und den Besitzverhältnissen in den anderen Erhebungswellen) meist nur einer Generation zugesprochen. Dieses Vorgehen führt zum Editing weiterer Vermögensangaben, da der Verkehrswert als auch die Filterinformation entsprechend korrigiert werden. .

### 4.2.3 „Logische“ Imputationen

Bevor im nächsten Abschnitt auf die regressionsbasierte Imputation der Markt- und Schuldenwerte des selbstgenutzten Wohneigentums eingegangen wird, stellt dieser Abschnitt eine weitere Imputationstechnik vor. „Logische“ Imputationen basieren nicht auf Regressionen, sondern leiten sich direkt aus weiteren zur Verfügung stehenden Informationen aus dem Haushaltsfragebogen und/oder den Antwortangaben weiterer Haushaltsmitglieder ab. Es wird

angenommen, dass diese alternativen Informationen als valide gelten, und dementsprechend direkt zur Ersetzung eines fehlenden Wertes verwendet werden können.

Die Filterinformation für das selbstgenutzte Wohneigentum wird als einzige Vermögenskomponente nur auf Basis logischer Ersetzungen imputiert. Mit der Frage zum Eigentümerstatus im Haushaltsfragebogen (Variable SH22 bzw. OWNER02 in 2002 und XH20 bzw. OWNER07 in 2007) liegen Informationen vor, ob eine Immobilie gemietet, mietfrei bewohnt oder selbst genutzt ist.

Geben eine oder mehreren Personen im Haushalt an, zu 100% Eigentümer der selbstgenutzten Immobilien zu sein, so wird im Falle einer fehlenden Antwortangabe eines weiteren Haushaltsmitglieds bzgl. der Filterinformation angenommen, dass diese kein Miteigentümer ist. Geht aus dem Haushaltsfragebogen hervor, dass ein Haushalt zur Miete in einer Immobilie wohnt, so wird eine fehlende Filterinformation bzgl. des Besitzes einer selbstgenutzten Immobilie ebenso auf „Nein“ gesetzt.

Ist die Summe der Eigentumsanteile im Haushalt kleiner als 100% und sind eine oder mehrere Personen mit fehlender Filterinformation im Haushalt, so wird anhand des Alters der Befragten und der Stellung zum Haushaltsvorstand fallweise geprüft, ob eine Person Eigentümer ist. Es wird angenommen, dass sehr junge und sehr alte Personen keine Eigentümer selbstgenutzter Immobilien sind, und dass zudem keine Personen außerhalb des Haushaltes Anteile an dem Eigentum halten.

Im Falle fehlender Angaben zur Höhe des Markt- und Schuldenwerts einer selbstgenutzten Immobilie werden Informationen anderer Haushaltsmitglieder mit als valide anerkannten Antwortangaben zur Ersetzung der fehlenden Information herangezogen. Liegen keine Informationen anderer Haushaltsmitglieder vor, so wird davon ausgegangen, dass Haushalte schuldenfrei sind wenn sie einer der folgenden Bedingungen genügen: es werden keine monatlichen Zins- und Tilgungszahlungen (Variable SH32 in 2002 bzw. XH29 in 2007) geleistet; die Immobilie wurde geerbt; der Haushalt lebt länger als 25 Jahre in der Wohnung bzw. in dem Haus. Für Personen in Haushalten mit monatlichen Zins- und Tilgungszahlungen wird die Restschuld auf „fehlend“ gesetzt und anschließend ein positiver Wert regressionsbasiert imputiert.

Bei Haushalten, in denen Personen eine fehlende Information zum Eigentümeranteil aufweisen, wird angenommen, dass diese den restlichen Anteil aller Haushaltsmitglieder am selbst-

genutzten Wohneigentum besitzen. Es wird somit unterstellt, dass keine Personen außerhalb des Haushaltes Anteile an dem Wohneigentum halten. Für Einpersonenhaushalte mit fehlenden Angaben wird daher immer alleiniger Besitz zu 100% unterstellt. Hat ein Eigentümer einen Anteil genannt und ein anderer nicht, so bekommt derjenige ohne Angabe die an 100% fehlenden Eigentumsanteile zugewiesen. Geben zwei Miteigentümer keinen Anteil an, werden die Eigentumsansprüche hälftig aufgeteilt.

Nach der Durchführung des Editing und „logischer" Imputationen ist für alle Befragungspersonen geklärt, ob sie Eigentümer sind und wie hoch ihr eventueller Anteil am selbstgenutzten Wohneigentum ist. Außerdem liegt für alle Haushalte eine Information darüber vor, ob Schulden auf dem selbstgenutzten Wohneigentum bestehen. Sowohl die Höhe der Schulden wie auch der Marktwert werden für Befragungspersonen mit fehlenden Angaben im weiteren regressionsbasiert imputiert.

### 4.2.4 Multiple Imputation der Markt- und Schuldenwerte

Die Imputation der Markt- und Schuldenwerte selbstgenutzten Wohneigentums verläuft in ähnlicher Weise wie bei den o. g. Vermögenskomponenten: Zunächst wird als Startwert für eine iterative Imputation der fehlenden Angaben eine Imputation im Querschnitt für das Erhebungsjahr 2002 durchgeführt. Diese wird anschließend für eine längsschnittbasierte Imputation der Erhebung des Jahres 2007 genutzt und umgekehrt. Abschließend werden zu den vorhergesagten Werten zufällig gezogene Residuen addiert. Bei den Markt- und Schuldenwerten des selbstgenutzten Wohneigentums gibt es allerdings drei Veränderungen dieser Vorgehensweise: Erstens wird die Regression auf der Ebene der Haushalte durchgeführt, da die Angaben zum Markt- und Schuldenwert für alle Eigentümer im Haushalt identisch sind. Das bedeutet, dass pro Haushalt ein Repräsentant nach den folgenden Kriterien ausgewählt wird: Miteigentümer, Vorliegen von Informationen über die interessierenden personenbezogenen Kovariaten, höchster Eigentümeranteil, Stellung zum Haushaltsvorstand. Diese Auswahl führt dazu, dass mehrheitlich der Haushaltsvorstand als Haushaltsrepräsentant ausgewählt wird, welcher gemäß SOEP-Standards die Person sein sollte, die sich am besten mit den allgemeinen ökonomischen Bedingungen des Haushalts inkl. der Finanzen auskennt. .

Zweitens bedingen sich Markt- und Schuldenwert gegenseitig. Daher geht jede der beiden Komponenten wechselseitig in die Schätzung des jeweils anderen Wertes als Kovariate ein. In einem iterativen Regressionsprozess wird nur auf Basis der Angaben des Jahres 2002 zu-

nächst der Marktwert regressiert. Für Haushalte mit fehlender Angabe zur Schuldenhöhe wird als Startwert der Mittelwert der Schuldenwerte aller Haushalte verwendet. Anschließend wird die Schuldenhöhe regressiert. Hier wird der Marktwert (samt Editierungen und Imputationen) als Kovariate verwendet. Bei Haushalten mit fehlendem Marktwert wird der mittels der vorherigen Regression vorhergesagte Marktwert eingesetzt. Daraufhin wird wiederum der Marktwert 2002 regressiert. Diese wechselseitigen Regressionen werden wiederholt, bis Markt- und Schuldenwert 2002 jeweils fünf Mal regressiert worden sind.

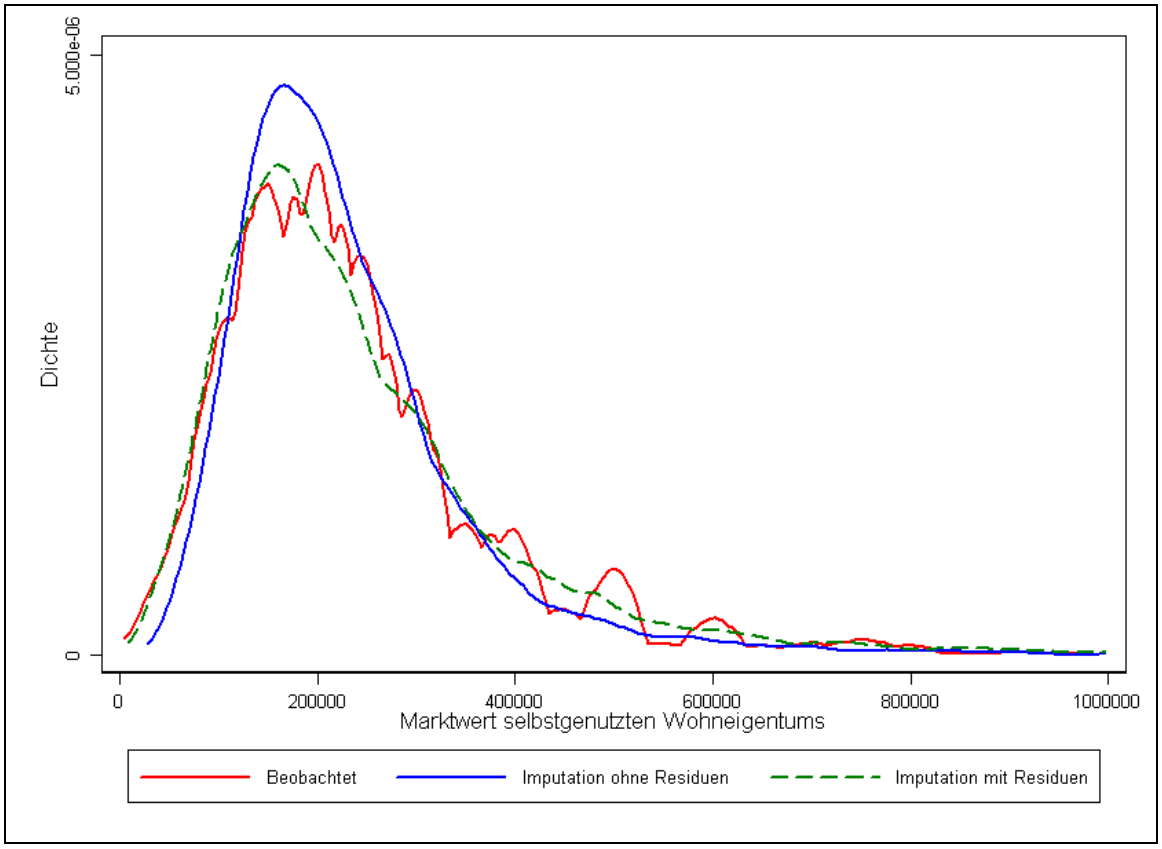
Die in der fünften Wiederholung vorhergesagten Werte für 2002 gehen in die Regressionen der Markt- und Schuldenwerte 2007 mit ein. Bei den Regressionen für 2007 wird wiederum wechselseitig der Markt- bzw. Schuldenwert als erklärende Variable verwendet, wobei nun die Längsschnittinformation zum Markt- bzw. Schuldenwert aus 2002 mit einfließt. Diese Werte werden jedoch nur für diejenigen Haushalte verwendet, die nach einer entsprechenden Kontrolle 2002 und 2007 im selben Wohneigentum wohnen. Wie bei den anderen Vermögenskomponenten werden die vorhergesagten Werte aus 2007 (hier: der fünften Wiederholung) wiederum für Regressionen der Werte in 2002 verwendet und umgekehrt; dieses Vorgehen wird für 2002 und 2007 jeweils fünf Mal im Längsschnitt wiederholt.

Der dritte Unterschied besteht darin, dass darauf geachtet werden muss, dass durch das Addieren von Residuen der Schuldenwert den Marktwert nicht zu stark übersteigt. Durch drei Maßnahmen wird versucht diese Bedingung zu erreichen: Als erste Maßnahme gehen bei der fünften Wiederholung der Schuldenwert-Regression in jedem Jahr die Marktwerte *inklusive* der hinzuaddierten Residuen als Kovariaten in die Schätzung mit ein. Das bedeutet, dass für jedes der (sieben) gezogenen Marktwert-Residuen eine separate Regression des Schuldenwertes berechnet wird. Zu den sieben vorhergesagten Schuldenwerten wird jeweils *ein* zufällig gezogenes Residuum aus der entsprechenden Regression addiert, sodass für Markt- und Schuldenwert jeweils sieben Versionen vorliegen. Die zweite Maßnahme ist, dass von diesen sieben Versionen jeweils die beiden Extremwerte, d.h. der größte und der kleinste imputierte Markt- bzw. Schuldenwert, verworfen werden, so dass insgesamt nur fünf Implicates vorliegen. Als dritte Maßnahme werden bis zu drei Mal neue Residuen für den Schuldenwert gezogen, wenn der Schuldenwert den Marktwert um das Eineinhalbfache übersteigt.

Um die Qualität des Imputationsverfahrens zu prüfen bzw. zu bestätigen wurden Personen mit validen Angaben fiktiv auf fehlend umgesetzt und anschließend imputiert. Abbildung 2 gibt die Kerndichte-Schätzer der ursprünglich beobachteten Information und jene der entsprechend

geschätzten Werte zum Marktwert der selbstgenutzten Immobilien für die identische Population im Erhebungsjahr 2002 wider.

Abbildung 2: Geschätzte Marktwerte selbstgenutzter Immobilien und die Verwendung zufallsbestimmter Residuen 2002



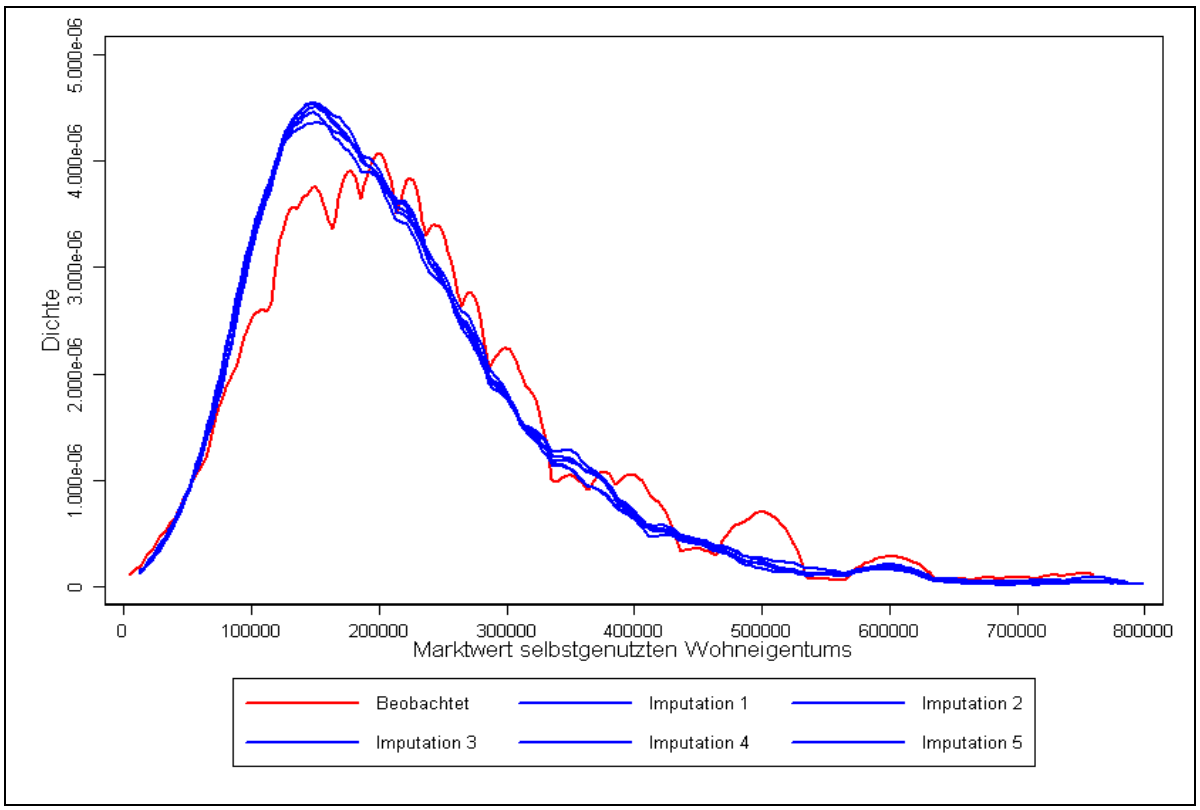

Anmerkung: Werte über einer Million Euro sind in dieser Abbildung gestutzt. Population: Haushalte mit einem valide beobachteten Marktwert (5,104 Haushalte).
Quelle: SOEP 2002.

Die gestrichelte Linie gibt die Dichte der (fiktiv) imputierten Werte inklusive zufallsbasierter Residuen an, die blaue Linie die Dichte der (fiktiv) imputierten Werte ohne zufallsbestimmte Residuen, die rote durchgezogene Linie gibt die tatsächlich beobachteten Marktwerte an. Wie deutlich wird, bewirkt eine Vernachlässigung der Residuen einen „*Regression-to-the-mean*“ Effekt mit entsprechend starker Unterschätzung der Varianz. Dagegen zeigen Vergleiche der

„wahren“ Verteilung mit derjenigen der Schätzung inklusive Störterme für die gleichen Beobachtungen einen hohen Grad an Übereinstimmung[21].

Im Gegensatz zur obigen fiktiven Imputation wird in Abbildung 3 wieder auf der Grundlage der Marktwerte selbstgenutzter Immobilien der Unterschied zwischen der Verteilung der beobachteten Werte und derjenigen, die aufgrund von Item-Non-Response tatsächlich imputiert werden mussten dargestellt.

Abbildung 3: Vergleich mehrfach imputierter Fälle mit Item-Non-Response im Gegensatz zu den beobachteten Fällen 2002




Anmerkung: Werte von über 1 Million Euro sind in dieser Abbildung nicht berücksichtigt. Die Imputationsversionen beinhalten zufällig aus der Verteilung der wahren Beobachtungen gezogene Residuen.
Quelle: SOEP 2002.

Zum einen variieren die fünf Schätzwerte der jeweiligen Verteilungen für die fünf (multiplen) Imputationen in Fällen von Item-Non-Response (blaue Linien). Zum anderen zeigt sich, dass sich die zu imputierenden Fälle deutlich in der Verteilung von den beobachteten Fällen unter-

[21] Die Verteilung der geschätzten Werte liegt in konsistenter Weise innerhalb des Zwei-Sigma-Konfidenzintervalls der wahren Verteilung (aus Lesbarkeitsgründen in der Abbildung nicht ausgewiesen).

scheiden: Die Verteilungen der fünf Imputationen sind kompakter und außerdem nach links verschoben. Dies deutet darauf hin, dass die Immobilien-Marktwerte im Falle von Item-Non-Response nicht zufällig fehlen (wie schon im vorherigen Abschnitt zur Selektivität fehlender Antwortangaben gezeigt wurde) und häufiger im unteren Teil der Verteilung auftreten. Tatsächlich finden sich diese niedrigeren (geschätzten) Marktwerte überproportional häufig für Immobilien von Haushalten in älteren Gebäuden, solchen mit geringerer Wohnfläche, in eher ländlichen Gegenden und sind häufiger von älteren Personen seit langem bewohnt.

## 4.3 Sonstige Immobilien

Die Imputation fehlender Angaben bei sonstigem Immobilienbesitz orientiert sich am Vorgehen zum selbstgenutzten Wohneigentum. Allerdings stehen hier deutlich weniger Informationen aus dem Haushaltsfragebogen als auch Angaben anderer Haushaltsmitglieder zur Verfügung, sodass das Editing und „logische“ Imputationen deutlich seltener eingesetzt werden können.

### 4.3.1 Editing und „logische“ Imputationen

Editing und „logische“ Imputationen werden in beiden Erhebungswellen nach demselben Muster und nur dann vorgenommen, wenn deutliche Hinweise darauf bestehen, dass mehrere Personen im Haushalt dieselbe/n Immobilie/n besitzen. Ein möglicher Hinweis auf diesen Sachverhalt besteht darin, dass die von den Befragungspersonen genannte Art und Anzahl der sonstigen Immobilien übereinstimmt und die Summe der Eigentumsanteile nicht mehr als 100% ausmacht. Editing und logische Imputationen des Markt- bzw. Schuldenwertes werden nur vorgenommen, wenn zudem die Werte um maximal ein Drittel von einander abweichen, da dann davon auszugehen ist, dass es sich vermutlich um die selbe/n Immobilie/n handelt. Den Befragungspersonen wird dann der Mittelwert der genannten Markt- bzw. Schuldenwerte der betroffenen Personen in dem gegebenen Haushalt zugeordnet. Weichen die Werte deutlicher voneinander ab, wird fallweise geprüft, ob die Unterschiede aus fehlenden Dezimalstellen resultieren.[22] Anders als beim selbstgenutzten Wohneigentum werden Filterinformationen weder editiert noch „logisch“ imputiert. Fehlt eine Angaben zur Höhe einer Restschuld, so

[22] Eine genauere Auflistung der einzelnen Editierungen und „logischen“ Imputationen findet sich für das Jahr 2002 in Frick et al. (2007) und kann von den Autoren auf Nachfrage auch für 2007 zur Verfügung gestellt werden.

wird geprüft, ob im Haushaltsfragebogen Informationen zu Zins- und Tilgungszahlungen für den sonstigen Immobilienbesitz vorliegen (Variable SH4202 in 2002 bzw. XH3902 in 2007). Werden Zins- und Tilgungszahlungen geleistet, wird eine Restschuld regressionsbasiert imputiert.[23]

### 4.3.2 Regressionsbasierte Imputationen

Die Imputation der Filterinformation zum Besitz sonstiger Immobilien erfolgt nach den gleichen Regeln wie in Abschnitt 4.1 für die anderen Vermögenskomponenten erläutert. Da die Art und Anzahl der sonstigen Immobilien wichtige Einflussgrößen für die Höhe des Markt- und Schuldenwertes sind, werden diese für Personen mit fehlenden Angaben ebenso imputiert. Dazu wird ein ordinales Logit-Modell für die Anzahl sonstiger Immobilien[24] und ein multinomiales Logit-Modell für die Art der Immobilie/n (Einfamilienhaus / Eigentumswohnung, Mehrfamilienhaus / Mietshaus, Ferienwohnung / Wochenendwohnung, unbebautes Grundstück, Sonstiges) geschätzt.[25]

Liegen nach dem Editing und eventueller logischer Imputationen noch fehlende Angaben über Restschulden auf sonstigen Immobilienbesitz vor, so wird mittels einer logistischen Regression die Wahrscheinlichkeit geschätzt, ob eine Restschuld besteht. Diese vorhergesagte Wahrscheinlichkeit wird mit einem zufälligen Wert verglichen. Der Zufallswert wird aus einer Normalverteilung mit Mittelwert 0.5 und Standardabweichung 0.2 gezogen. Ist die Wahrscheinlichkeit größer als die Zufallszahl, wird angenommen, dass die Person die Schuldenkomponente besitzt. Ist die Wahrscheinlichkeit kleiner, hat das Individuum diese Komponente nicht.

Fehlende Angaben zum Eigentumsanteil werden mit Hilfe einer OLS-Regression geschätzt. Dazu werden zusätzlich zu den Querschnitts-Kovariaten Anteilsinformationen aus den beiden Wellen in der oben beschriebenen Weise wechselseitig verwendet.

Die Imputation der Markt- und Schuldenwerte für sonstige Immobilien verläuft nach demselben iterativen Schema wie beim selbstgenutzten Wohneigentum. Mit zwei kleinen Unter-

23 Fehlt auch im Haushaltsfragebogen eine Information zur Zins- und Tilgungsleistung, so wird zuerst mittels logistischer Regression geschätzt, ob eine Restschuld vorliegt.

24 Elf und mehr Immobilien werden hierbei zu einer Kategorie zusammengefasst.

25 Vorher wurde zudem bei allen Personen, die zwei (oder mehr) verschiedene Arten von Immobilien besaßen, aber als Anzahl „eins“ angaben, die Anzahl „eins“ durch die Anzahl der verschiedenen Arten Immobilien ersetzt.

schieden: Zum einen wird kein Haushaltsrepräsentant ausgewählt, da Immobilienbesitzer im selben Haushalt unterschiedliche Immobilien halten können. Zum anderen werden nur dann neue Schuldenresiduen gezogen, wenn der Schuldenwert den Marktwert um das Doppelte übersteigt, um eine mögliche intensivere Kreditfinanzierung sonstiger Immobilien zu erlauben.

## 4.4 Geldvermögen

Der Imputationsprozess für das Geldvermögen folgt im Wesentlichen dem im Abschnitt 4.1 vorgestellten Schema. Es können jedoch zusätzliche “logische“ Imputationen zur Höhe des Marktwertes sowie des Eigentumsanteils vorgenommen werden.

### 4.4.1 „Logische“ Imputationen

„Logische“ Imputationen werden nur dann vorgenommen, wenn ein deutlicher Hinweis besteht, dass zwei Personen ihr Geldvermögen teilen. Geben beide Personen die gleiche Höhe des Geldvermögens an, aber nur eine Person einen Anteil von 50% und die andere Person keine Information zum Anteil, so wird auch der Anteil der Person ohne Angabe auf 50% gesetzt. Nennen zwei Person einen Anteilswert von 50% und eine gibt einen Wert zum Geldvermögen an, die andere Person jedoch keinen Wert, so wird dieser Person der gleiche metrische Wert zugeordnet.

### 4.4.2 Regressionsbasierte Imputationen

Die Imputation des Filters basiert wie für die anderen Vermögenskomponenten für jede Welle zunächst getrennt auf einer logistischen Regression für Personen mit INR und für PUNR. Die vorhergesagten Werte werden wieder mit einem zufälligen Schwellenwert verglichen. Auch für die Filtervariable des Geldvermögens werden Informationen aus den beiden Erhebungswellen wechselseitig genutzt und dieser Prozess fünfmal wiederholt.

Nahezu alle Befragungspersonen geben in beiden Erhebungsjahren entweder einen persönlichen Anteilswert von 50% oder 100% an. Daher wird bei einer fehlenden Angabe zum Anteilswert eine logistische Regression durchgeführt, um zu schätzen, ob eine Person alleiniger oder hälftiger Anteilsnehmer ist. Die Vorgehensweise ist analog zu den anderen logistischen Regressionen und verwendet zum einen Längsschnittinformationen und nutzt zudem das bereits beschriebene iterative Vorgehen der Imputation.

Die Imputation der Höhe des Geldvermögens verläuft ebenfalls nach dem eingangs beschriebenen Schema: Zunächst wird der Wert für 2002 im Querschnitt mittels einer Regression bestimmt. Diese Vorhersage wird dann gemeinsam mit denselben Kovariaten wie im 2002er-Modell für den Wert in 2007 genutzt. Anschließend wird auch der Wert für 2002 längsschnittsimputiert und der gesamte Prozess wird wiederholt, bis die Werte für beide Wellen fünfmal im Längsschnitt regressiert sind. Als varianz-erhaltende Maßnahme erfolgt abschließend wiederum fünfmal eine Addition zufällig gezogener Residuen (Ergebnisse einer in-sample Prädiktion).

# 5 Einfluss der Imputation auf den Anteil der Vermögenden, die Vermögenshöhe und -verteilung

Die folgenden Analysen zum Einfluss der oben beschriebenen Editing- und Imputationsroutinen vergleichen jeweils Ergebnisse auf Basis des multipel imputierten Datensatzes mit jenen, die sich aus der (selektiven) Population der Befragten mit vollständig beobachteten Daten ergeben (complete case analysis). Die Auswirkung von Imputationen[26] auf den Anteil der Population, der den jeweiligen Vermögensbestandteil hält, ist in Tabelle 5.1 dargestellt. Der Anteil der Vermögensbesitzer steigt aufgrund der Imputation um bis zu 50% im Jahre 2002. Er erreicht gut 17% beim Geldvermögen, rund 22% beim selbstgenutzten Immobilienbesitz und knapp 54% beim Betriebsvermögen. Für 2007 fallen die entsprechenden Anteile nur geringfügig niedriger aus.

Während der Anteil imputierter oder editierter Vermögensdaten in der Population je nach Vermögensbestandteil variiert, wird im Durchschnitt aller Vermögensbestandteile etwa 25-30% des Vermögens imputiert (Tabelle 5.1). Die Ausnahme sind Betriebsvermögen, bei denen mehr als 40% des Vermögens in 2002 entweder imputiert oder editiert wurde. Dabei ist die niedrige Antwortquote wohl ein Produkt sowohl „klassischer Antwortverweigerung“ und – wohl noch bedeutsamer – unzureichender Kenntnis des wahren Vermögenswertes bzw. der Unfähigkeit diesen Wert zu schätzen. Bei den Schulden umfassen die imputierten oder editierten Werte gut 30% des Aggregats der Verbindlichkeiten (im Jahr 2002); darunter 32% im Falle von Schulden im Zusammenhang mit selbst genutztem Wohneigentum, 33% bei Schulden für sonstige Immobilien und 19% bei Konsumentenkrediten. Gegenüber 2002 ergeben sich in 2007 vor allem bei den Konsumentenkrediten deutliche Veränderungen. Der Anteil der faktisch erhobenen Beobachtungen steigt deutlich gegenüber der Vorperiode an, was entsprechend zur Folge hat, dass der Anteil zusätzlich als verschuldet imputierter Fälle und auch die Veränderung des Aggregats (mit einem Zuwachs von knapp 9%) deutlich geringer ausfällt als in 2002 (Zuwachs von knapp 19%).

[26] Im Folgenden bezieht sich “Imputation” sowohl auf das Editing als auch auf Imputation.

Tabelle 5.1: Der Einfluss von Editing und Imputation auf den Anteil der Population mit Vermögen und das Vermögensaggregat 2002 und 2007

| | 2002 | | | | 2007 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Beobachtet[1]** | **Total[2]** | **Veränderung[3] In %** | **Veränderung im aggregierten Vermögen in %** | **Beobachtet[1]** | **Total[2]** | **Veränderung[3] in %** | **Veränderung im aggregierten Vermögen in %** |
| **Selbst genutzte Immobilie** | 29,6 | 36,0 | 21,6 | 29,5 | 29,2 | 35,9 | 22,9 | 32,5 |
| **Sonstige Immobilien** | 8,0 | 10,0 | 25,0 | 33,3 | 8,2 | 10,3 | 25,6 | 22,3 |
| **Geldvermögen** | 38,9 | 45,3 | 16,5 | 25,9 | 44,1 | 48,9 | 10,9 | 24,9 |
| **Private Versicherungen** | 38,3 | 47,1 | 23,0 | 31,1 | 44,6 | 53,0 | 18,8 | 39,5 |
| **Betriebsvermögen** | 2,8 | 4,3 | 53,6 | 42,4 | 3,1 | 4,5 | 45,2 | 36,0 |
| **Wertsachen** | 6,2 | 9,5 | 53,2 | 35,9 | 5,2 | 6,1 | 17,3 | 11,2 |
| **Brutto-Gesamtvermögen** | 63,4 | 74,9 | 18,1 | 31,2 | 66,0 | 77,3 | 17,1 | 31,1 |
| **Hypotheken auf selbst genutzte Immobilien** | 13,9 | 18,1 | 30,2 | 31,7 | 14,0 | 18,0 | 28,6 | 32,5 |
| **Hypotheken auf sonstige Immobilien** | 3,8 | 4,6 | 21,1 | 33,4 | 4,0 | 4,8 | 20,0 | 28,8 |
| **Konsumentenkredite** | 9,7 | 11,9 | 22,7 | 18,9 | 16,3 | 17,0 | 4,3 | 8,9 |
| **Verbindlichkeiten insgesamt** | 23,4 | 29,4 | 25,6 | 30,1 | 28,6 | 32,9 | 15,0 | 27,9 |
| **Netto-Vermögen** | -- | -- | -- | 31,5 | -- | -- | -- | 31,7 |

Population: Erwachsene Population (17 Jahre oder älter) mit Interview. Gewichtet.
[1] Nur Personen mit beobachteten Werten.
[2] Nach Editing und multipler Imputation.
[3] Berechnet als (Total-Beobachtet)/Total).
Quelle: SOEP 2002 und 2007.

Den Einfluss von Editing und Imputation verdeutlicht letztlich die Veränderung des Aggregats des Nettogesamtvermögens: gegenüber den beobachteten Fällen nimmt dieses in beiden Erhebungsjahren um knapp 32% zu. Aufgrund dieses vergleichsweise hohen Anteils von imputierter Vermögensmasse ist es um so wichtiger die Imputation mehrfach durchzuführen, um die Unsicherheit der Imputation im Sinne der Variation des unbeobachteten wahren Wertes auszudrücken.

In Tabelle 5.2 werden die Auswirkungen der Imputation auf die Höhe ausgewählter Vermögenskomponenten aufgeführt. Da keine der hier untersuchten Komponenten von mehr als der Hälfte der Beobachtungen in der SOEP-Stichprobe gehalten wird, liegt der Median der einzelnen Vermögens- und Schuldenkomponenten für die gesamte Bevölkerung bei Null und

wird daher hier nicht ausgewiesen. Die Imputation hat hingegen signifikant positive Auswirkungen auf die Durchschnittswerte der Vermögensbestandteile der gesamten Population. Für Betriebsvermögen nimmt der Wert um mehr als die Hälfte zu und für die anderen Bestandteile immerhin um Werte zwischen 5% und 29%. Signifikante Veränderungen liegen aber nur beim selbstgenutzten Immobilienbesitz und entsprechenden Hypotheken vor.

Unter den jeweiligen Besitzern der Vermögensbestandteile ist die durch Imputation und Editing bewirkte Änderung des Mittelwerts seltener signifikant und wesentlich geringer, bzw. im Falle von Geldvermögen und selbstgenutztem Immobilienbesitz fallen die Veränderungen zum Teil sogar negativ aus. Für die gesamte Population hat die Imputation einen signifikanten Effekt bei der Reduzierung von Ungleichheit gemessen am Gini-Koeffizienten und HSCV (half-squared coefficient of variation) (vgl. Tabelle 5.3). Der HSCV reagiert als top-sensitives Maß definitionsgemäß stets wesentlich stärker als der Gini. Dies kann durch den Imputationsprozess erklärt werden, bei dem Werte im oberen Bereich der Verteilung hinzugefügt werden, wodurch die Zahl der Beobachtungen steigt und die Ungleichheit am oberen Ende der Verteilung sinkt.

Tabelle 5.2: Der Einfluss von Editing und Imputation auf mittlere individuelle Vermögen 2002 und 2007

| | **2002** | | | | | | **2007** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Insgesamt** | | | **Nur Eigentümer** | | | **Insgesamt** | | | **Nur Eigentümer** | | |
| | **Mittelwert** | | | **Mittelwert** | | | **Mittelwert** | | | **Mittelwert** | | |
| | **Beo.[1]** | **Total[2]** | **Ver.[3] in %** | **Beo.[1]** | **Total[2]** | **Ver.[3] in %** | **Beo.[1]** | **Total[2]** | **Ver.[3] in %** | **Beo.[1]** | **Total[2]** | **Ver.[3] in %** |
| **Immobilien [4]** | 43.616 | 50.229 | +15,2* | 147.479 | 139.656 | -5,3* | 43.895 | 51.802 | +18,0* | 150.241 | 144.229 | -4,0 |
| Se | (818) | (809) | | (2,081) | (1,450) | | (1,143) | (938) | | (2,333) | (2,072) | |
| **Geld-Vermögen** | 9.472 | 9.982 | 5,4 | 24.371 | 22.015 | -9,7 | 12.328 | 13.100 | 6,3 | 27.984 | 26.801 | -4,2 |
| Se | (353) | (293) | | (828) | (660) | | (722) | (635) | | (1,601) | (1,160) | |
| **Betriebs-Vermögen** | 5.468 | 8.812 | 61,2 | 192.824 | 204.611 | 6,1 | 6.537 | 9.773 | 49,5 | 209.814 | 216.391 | 3,1 |
| Se | (1,081) | (1,351) | | (38,073) | (32,124) | | (957) | (2,157) | | (28,482) | (47,880) | |
| **Hypotheken [5]** | 12.094 | 15.164 | +25,4* | 86.902 | 83.883 | -3,5 | 12.909 | 16.659 | +29,0* | 92.256 | 92.676 | 0,5 |
| **Se** | (375) | (350) | | (1,632) | (1,270) | | (427) | (488) | | (1,990) | (1,862) | |

Standardfehler in Klammern. . (*) bezeichnet signifikante Abweichungen (95%-Niveau).
[1] Nur Personen mit beobachtetem individuellem Anteil und Marktwert
[2] Nach Editieren und multipler Imputation
[3] Veränderung von (Total-Beobachtet)/Total
[4] Selbst genutzte Immobilien
[5] Hypotheken auf selbst genutzte Immobilien
Quelle: SOEP 2002 und 2007.

Tabelle 5.3: Der Einfluss von Editing und Imputation auf die Ungleichheit von ausgewählten Vermögenskomponenten

| | 2002 | | | | | | 2007 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Insgesamt | | | Nur Eigentümer | | | Insgesamt | | | Nur Eigentümer | | |
| | Beo.[1] | Total[2] | Ver.[3] in % | Beo.[1] | Total[2] | Ver.[3] in % | Beo.[1] | Total[2] | Ver.[3] in % | Beo.[1] | Total[2] | Ver.[3] in % |
| **Immobilien[4]** | | | | | | | | | | | | |
| - Gini | 0,806 | 0,763 | -5,4* | 0,345 | 0,341 | -1,4 | 0,815 | 0,771 | -5,4* | 0,367 | 0,362 | -1,4 |
| - HSCV | 2,198 | 1,710 | -22,2* | 0,298 | 0,295 | -1,1 | 2,568 | 1,915 | -25,4* | 0,396 | 0,368 | -7,3 |
| **Geld-vermögen** | | | | | | | | | | | | |
| - Gini | 0,858 | 0,832 | -3,1* | 0,636 | 0,630 | -0,9 | 0,873 | 0,855 | -2,1 | 0,712 | 0,703 | -1,2 |
| - HSCV | 11,023 | 8,966 | -18,7 | 3,979 | 3,792 | -4,7 | 23,145 | 19,552 | -15,5 | 9,917 | 9,301 | -6,2 |
| **Betriebs-vermögen** | | | | | | | | | | | | |
| - Gini | 0,994 | 0,992 | -0,3 | 0,793 | 0,806 | +1,6 | 0,995 | 0,992 | -0,3 | 0,824 | 0,818 | -0,7 |
| - HSCV | 946,67 | 783,74 | -17,2* | 26,360 | 33,274 | +26,2 | 369,11 | 274,24 | -25,7 | 11,015 | 11,909 | +8,1 |
| **Hypotheken[5]** | | | | | | | | | | | | |
| - Gini | 0,920 | 0,897 | -2,4 | 0,422 | 0,432 | +2,3 | 0,916 | 0,892 | -2,6 | 0,402 | 0,401 | -0,1 |
| - HSCV | 6,112 | 4,958 | -18,9 | 0,420 | 0,487 | +15,8 | 5,342 | 4,117 | -22,9* | 0,317 | 0,330 | +4,0 |

* bezeichnet signifikante Abweichungen (95%-Niveau).
[1] Nur Personen mit beobachtetem individuellem Anteil und Marktwert
[2] Nach Editieren und multipler Imputation
[3] Veränderung von (Total-Beobachtet)/Total
[4] Selbst genutzte Immobilien
[5] Hypotheken auf selbst genutzte Immobilien
Quelle: SOEP 2002 und 2007.

Für die Gruppe der jeweiligen Besitzer einer bestimmte Vermögenskomponente zeigt sich ein uneinheitliches Bild. Bei selbstgenutzten Immobilien und Geldvermögen bewirkt die Imputation eine leichte Abnahme der Ungleichheit. Ungleichheitsfördernd wirkt die Imputation bei Betriebsvermögen sowie bei Hypotheken auf selbstgenutzten Immobilien. Die Auswirkungen sind aber in allen Fällen nicht statistisch signifikant.

Bisher wurde die Auswirkung von Editing und Imputation auf die einzelnen Bestandteile des Nettovermögens untersucht. Im Folgenden wird das für wohlfahrtsökonomische Analysen relevantere Nettogesamtvermögen analysiert, das sich als Summe aller Vermögenswerte abzüglich der Verbindlichkeiten ergibt. Tabelle 5.4 gibt einen vergleichenden Überblick über zentrale Verteilungsinformationen basierend auf den faktisch beobachteten und den nach Korrektur um Messprobleme (Editing und Imputation) verfügbaren Nettovermögensmessungen im SOEP der Jahre 2002 und 2007.

Tabelle 5.4: Der Einfluss von Editing und Imputation auf das Nettogesamtvermögen und auf relative Vermögensarmut 2002 und 2007

| | 2002 | | | 2007 | | |
|---|---|---|---|---|---|---|
| | **Beobachtet[1]** | **Total[2]** | **Veränderung in %[3]** | **Beobachtet[1]** | **Total[2]** | **Veränderung in %[3]** |
| **Mittelwert** | 61.372 | 81.123 | 32,2* | 64.376 | 88.799 | 37,9* |
| **Standardfehler** | (1,687) | (2,171) | | (3,263) | (3,362) | |
| **Mittelwert falls >0** | 105.492 | 114.564 | 8,6 | 110.490 | 123.585 | 11,9 |
| **Standardfehler** | (2,966) | (3,277) | | (5,829) | (4,348) | |
| **1. Perzentil** | -20.000 | -19.913 | -0,4 | -33.000 | -30.000 | -9,1 |
| **5. Perzentil** | -3.500 | -1.829 | -47,7 | -7.000 | -4.069 | -41,9 |
| **10. Perzentil** | 0 | 0 | | -70 | 0 | |
| **25. Perzentil** | 0 | 0 | | 0 | 0 | |
| **50. Perzentil (Median)** | 5.000 | 14.786 | 195,7* | 4.000 | 15.000 | 275,0* |
| **75. Perzentil** | 60.500 | 95.150 | 57,3* | 50.500 | 96.307 | 90,7* |
| **90. Perzentil** | 175.000 | 206.628 | 18,1* | 165.000 | 220.306 | 33,5* |
| **95. Perzentil** | 279.000 | 315.346 | 13,0* | 273.460 | 335.227 | 22,6* |
| **99. Perzentil** | 600.000 | 746.579 | 24,4* | 702.500 | 816.804 | 16,3 |
| **Gini** | 0,835 | 0,787 | -5,7* | 0,883 | 0,804 | -8,9* |
| **HSCV** | 6,737 | 16,218 | 140,7 | 14,270 | 8,727 | -38,8 |
| **P90/P50-Verhältnis** | 35,0 | 14,0 | -60,1* | 41,3 | 14,7 | -64,4* |
| **Zahl der Beobachtungen** | 14.038 | 23.135 | 64,8 | 12.901 | 20.965 | 62,5 |

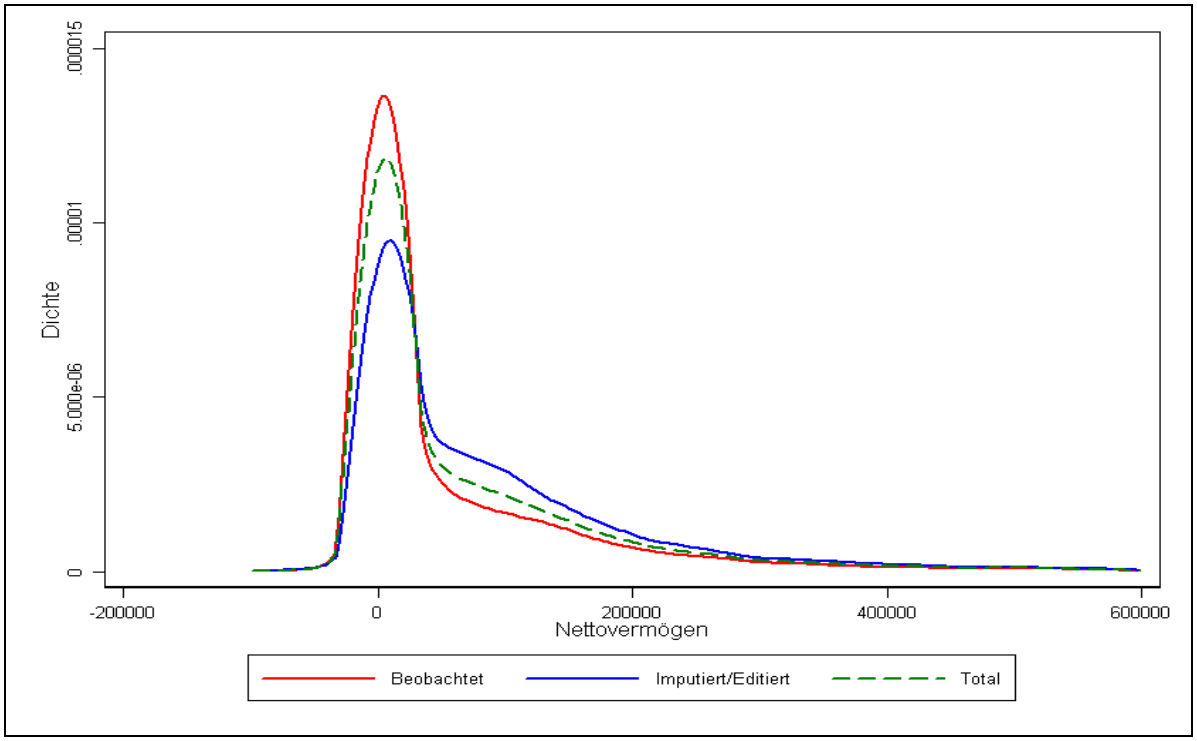
Standardfehler in Klammern. . (*) bezeichnet signifikante Abweichungen (95%-Niveau).
[1] Nur Personen mit beobachtetem individuellem Anteil und Marktwert.
[2] Nach Editieren und multipler Imputation
[3] (Total-Beobachtet)/Total
Quelle: SOEP 2002 und 2007 individuelle Vermögensinformationen

Wie es schon bei den Vermögensbestandteilen der Fall war, erhöht sich auch bei dieser aggregierten Betrachtung das durchschnittliche Nettovermögen aufgrund des Imputationsprozesses signifikant um rund ein Drittel für die Gesamtbevölkerung und um rund 10% für diejenigen mit positivem Nettovermögen. Bei Betrachtung der Verteilung auf Basis der Perzentilsgrenzen sind die Beobachtungen nahe des Median am stärksten betroffen. Die Auswirkung auf die insgesamt gemessene Ungleichheit ist uneindeutig. Während der Gini-Koeffizient durch die Berücksichtigung der imputierten Fälle reduziert wird, steigt der top-sensitive HSCV im Jahre 2002 um mehr als das Doppelte; hier sind offensichtlich auch Ausreißer-Effekte zu berücksichtigen (siehe dazu Frick, Grabka, Sierminska 2007).

Alternativ wird das Nettogesamtvermögen nach Imputationsstatus in Abbildung 4 mit Hilfe von Kerndichteschätzern dargestellt. Die Verteilung der editierten bzw. imputierten Werte ist in blau und die der beobachteten Werte in rot ausgewiesen. Das daraus resultierende gesamte Nettovermögen wird in seiner Verteilung für alle Beobachtungen in blau wieder gegeben. Es

finden sich deutliche Unterschiede in den Kerndichteschätzern rund um den Nullpunkt sowie im Bereich von 50,000 bis 150,000 Euro, daher ist die Verteilung der imputierten Fälle und damit auch die „gesamte“ Verteilung nach rechts verschoben[27].

Abbildung 4: Vergleich beobachteter und imputierter Werte für das Netto-Gesamtvermögen 2002




Anmerkung: Werte von weniger als -200.000 € und mehr als +600.000€ sind in dieser Abbildung getrimmt.
Quelle: SOEP 2002.

Zusammenfassend lässt sich somit festhalten: die Erfassung von Vermögen(skomponenten) in Bevölkerungssurveys ist quasi zwangsläufig mit Messfehlern, Item-Non-Response und Inkonsistenzen behaftet. Diese sind zudem zumindest in Teilen selektiv, d.h., ein „missing completely at random (MCAR)“ Mechanismus im Sinne Rubins (1987) kann weitestgehend ausgeschlossen werden. Multiple Imputation und Editing ermöglichen eine Korrektur dieser Selektivität und erhalten die komplette Analysepopulation. Dabei ergibt sich konsequenterweise ein insgesamt signifikanter Einfluss dieser Korrektur auf zentrale Vermögensverteilungsergebnisse. Dies gilt sowohl für einzelne Vermögenskomponenten als auch für das Nettogesamtvermögen.

[27] Die Verteilungen der fünf Imputationsversionen sowie der sich ergebenden zusammengesetzten Gesamtvermögensvariablen sind mehr oder weniger identisch und aus Gründen der besseren Lesbarkeit hier nicht komplett dargestellt.

# 6 Konventionen bei der Benennung der Variablen

Die verschiedenen Imputationsversionen einer Vermögenskomponente werden für jede Person und für beide Erhebungswellen jeweils der Größe nach sortiert. Anschließend werden die fünf verschiedenen Versionen per Zufallsverfahren angeordnet, sodass jede der Versionen gleich häufig jeden Rangplatz innehält. Dabei wird aber beachtet, dass die Version mit dem größten Wert in 2002 für jede Person auch die Version mit dem größten Wert in 2007 entspricht; die zweitgrößte Version in 2002 entspricht der zweitgrößten Version in 2007 usw. Diese Anordnung ist jedoch für eine Vermögenskomponente nur innerhalb einer Person gleich angeordnet, nicht aber über alle Personen oder Vermögenskomponenten hinweg, sodass sich für die fünf verschiedenen Implicates im Aggregat keine Rangordnung ergibt.

Um eine Identifikation von Fällen zu erleichtern, die mittels Editing bzw. Imputation ersetzt wurden, wird allen generierten Vermögensvariablen eine „Flag“-Variable zugeordnet. Diese Flaginformation nimmt den Wert „0“ an, wenn die dazugehörige Information weder imputiert noch editiert wurde, den Wert „1“, wenn Editing stattgefunden hat, und den Wert „2“ falls eine Imputation bei einem fehlenden Wert vorgenommen wurde.

Trotz der hier beschriebenen umfangreichen Arbeiten zur Qualitätssicherung der SOEP-Vermögensdaten wird Nutzern dringend empfohlen, im Rahmen eigener (Regressions-)Analysen die SOEP-Beobachtungen mit imputierten Daten mit Hilfe der genannten Flag-Variablen als zusätzlicher Kontrollvariable zu berücksichtigen.

Die generierten Vermögensinformationen des SOEP sind auf der Personenebene in dem File PWEALTH und in aggregierter Weise auf der Haushaltsebene im File HWEALTH abgelegt. Diese beiden Datensätze liegen im so genannten Longformat vor, d.h. für jede Beobachtungseinheit findet sich je nach Erhebungsjahr genau eine Zeile mit Vermögensinformationen. Um die Vermögensinformationen über die Zeit hinweg voneinander zu separieren, wird eine zusätzliche Identifikatorvariable zum Erhebungsjahr zur Verfügung gestellt (SVYYEAR).

## 6.1 Vermögensvariablen auf der individuellen Ebene im File PWEALTH

**Identifiers**

| | |
|---|---|
| PERSNR | Individual identifier |
| HHNRAKT | Wave specific household identifier |
| SVYYEAR | Survey year |

**Owner-occupied property**

| | |
|---|---|
| p10000 | Filter information |
| p20000 | Imputation flag for filter information |
| p0100x | Market value (x = implicate a,b,…,e) |
| p02000 | Imputation flag for market value |
| p0010x | Debts (x = implicate a,b,…,e) |
| p00200 | Imputation flag for debts |
| p00010 | Individual share |
| p00020 | Imputation flag for individual share |
| p0110x | Net market value (p0100x - p0010x; x = implicate a,b,…,e) |
| p02200 | Imputation flag for net market value |
| p0101x | Individual share of market value (p0100x * p00010/100; x = implicate a,b,…,e) |
| p02020 | Imputation flag for individual share of market value |
| p0011x | Individual share of debts (p0010x * p00010/100; x = implicate a,b,…,e) |
| p00220 | Imputation flag for individual share of debts |
| p0111x | Individual share of net market value (p0100x-p0010x)*p00010/100; x = implicate a,b,…,e) |
| p02220 | Imputation flag for individual share of net market value |

**Other property**

| | |
|---|---|
| e10000 | Filter information |
| e20000 | Imputation flag for filter information |
| e0100x | Market value (x = implicate a,b,…,e) |
| e02000 | Imputation flag for market value |
| e00010 | Individual share |
| e00020 | Imputation flag for individual share |
| e0010x | Debts (x = implicate a,b,…,e) |
| e00200 | Imputation flag for debts |
| e0110x | Net market value (e0100x – e0010x; x = implicate a,b,…,e) |
| e02200 | Imputation flag for net market value |
| e0101x | Individual share of market value (e0100x*e00010/100; x = implicate a,b,…,e) |
| e02020 | Imputation flag for share of market value |
| e0011x | Individual share of debts (e0010x*e00010/100; x = implicate a,b,…,e) |
| e00220 | Imputation flag for individual share |
| e0111x | Individual share of net market value (e0100x-e0010x)*e00010/100; x = implicate a,b,…,e) |
| e02220 | Imputation flag for individual share of net market value |
| e00001 | Type: single-family house |
| e00002 | Type: apartment building |
| e00003 | Type: holiday home |
| e00004 | Type: undeveloped real estate |
| e00005 | Type: other property |
| e00007 | Number of properties |
| e00026 | Imputation flag for the type of property |
| e00027 | Imputation flag for the Number of properties |

**Financial Assets**

| | |
|---|---|
| f10000 | Filter information |
| f20000 | Imputation flag for filter information |
| f0100x | Market value (x = implicate a,b,…,e) |
| f02000 | Imputation flag for market value |
| f00010 | Individual share |
| f00020 | Imputation flag for individual share |
| f0101x | Individual share of market value (f0100x*f00010/100; x = implicate a,b,…,e) |
| f02020 | Imputation flag for individual share of market value |

**Building Loan Contract (available since 2007)**

| | |
|---|---|
| l10000 | Filter information |
| l20000 | Imputation flag for filter information |
| l0100x | Market value (x = implicate a,b,…,e) |
| l02000 | Imputation flag for market value |

**Private Insurances (available since 2007)**

| | |
|---|---|
| h10000 | Filter information |
| h20000 | Imputation flag for filter information |
| h0100x | Market value (x = implicate a,b,…,e) |
| h02000 | Imputation flag for market value |

**Private Insurances & Building Loan Contracts**

| | |
|---|---|
| i10000 | Filter information |
| i20000 | Imputation flag for filter information |
| i0100x | Market value (x = implicate a,b,…,e) |
| i02000 | Imputation flag for market value |

**Business Assets**

| | |
|---|---|
| b10000 | Filter information |
| b20000 | Imputation flag for filter information |
| b0100x | Market value (x = implicate a,b,…,e) |
| b02000 | Imputation flag for market value |
| b00001 | Ownership status |
| b00002 | Imputation flag for ownership status |

**Tangible Assets**

| | |
|---|---|
| t10000 | Filter information |
| t20000 | Imputation flag for filter information |
| t0100x | Market value (x = implicate a,b,…,e) |
| t02000 | Imputation flag for market value |

**Consumer Debts**

| | |
|---|---|
| c10000 | Filter information |
| c20000 | Imputation flag for filter information |
| c0100x | Market value (x = implicate a,b,…,e) |
| c02000 | Imputation flag for market value |

**Overall wealth**

| | |
|---|---|
| w0101x | Gross overall wealth (p0101x + e0101x + f0101x + i0100x + b0100x + t0100x02; x = implicate a,b,…,e) |
| w02020 | Imputation flag for gross overall wealth |
| w0011x | Overall debts (p0011x + e0011x + c0100x; x = implicate a,b,…,e) |
| w00220 | Imputation flag for overall debts |
| w0111x | Net overall wealth (w0101x - w0011x; x = implicate a,b,…,e) |
| w02220 | Imputation flag for net overall wealth |

## 6.2 Vermögensvariablen auf der Haushaltsebene im File HWEALTH

**Identifiers**

| | |
|---|---|
| HHNRAKT | Wave-specific household identifier |
| SVYYEAR | Survey year |

**Property, primary residence**

| | |
|---|---|
| p100h0 | HH filter information (max of p10000 over all HH-members) |
| p200h0 | HH imputation flag for filter information |
| p010hx | HH market value (sum of p0101x over all HH-members; x = implicate a,b,…,e) |
| p020h0 | HH imputation flag for market value |
| p001hx | HH debts (sum of p0011x over all HH-members; x = implicate a,b,…,e) |
| p002h0 | HH imputation flag for debts |
| p011hx | HH net value (p010Hx-p001Hx; x = implicate a,b,…,e) |
| p022h0 | HH imputation flag for net value |

**Other property**

| | |
|---|---|
| e100h0 | HH filter information (max of e10000 over all HH-members) |
| e200h0 | HH imputation flag for filter information |
| e010hx | HH market value (sum of e0101x over all HH-members; x = implicate a,b,…,e) |
| e020h0 | HH imputation flag for market value |
| e001hx | HH debts (sum of e0011x over all HH-members; x = implicate a,b,…,e) |
| e002h0 | HH imputation flag for debts |
| e011hx | HH net value (e010Hx-e001Hx; x = implicate a,b,…,e) |
| e022h0 | HH imputation flag for net value |

**Financial assets**

| | |
|---|---|
| f100h0 | HH filter information (max of f10000 over all HH-members) |
| f200h0 | HH imputation flag for filter information |
| f010hx | HH market value (sum of f0101x over all HH-members; x = implicate a,b,…,e) |
| f020h0 | HH imputation flag for market value |

**Building Loan Contracts (available since 2007)**

| | |
|---|---|
| l100h0 | HH filter information (max of i10000 over all HH-members) |
| l200h0 | HH imputation flag for filter information |
| l010hx | HH market value (sum of i0100x over all HH-members; x = implicate a,b,…,e) |
| l020h0 | HH imputation flag for market value |

**Private Insurances (available since 2007)**

| | |
|---|---|
| h100h0 | HH filter information (max of i10000 over all HH-members) |
| h200h0 | HH imputation flag for filter information |
| h010hx | HH market value (sum of i0100x over all HH-members; x = implicate a,b,…,e) |
| h020h0 | HH imputation flag for market value |

**Private insurances and Building Loan Contracts**

| | |
|---|---|
| i100h0 | HH filter information (max of i10000 over all HH-members) |
| i200h0 | HH imputation flag for filter information |
| i010hx | HH market value (sum of i0100x over all HH-members; x = implicate a,b,…,e) |
| i020h0 | HH imputation flag for market value |

**Business assets**

| | |
|---|---|
| b100h0 | HH filter information (max of b10000 over all HH-members) |
| b200h0 | HH imputation flag for filter information |
| b010hx | HH market value (sum of b0100x over all HH-members; x = implicate a,b,…,e) |
| b020h0 | HH imputation flag for market value |

**Tangible assets**

| | |
|---|---|
| t100h0 | HH filter information (max of t10000 over all HH-members) |
| t200h0 | HH imputation flag for filter information |
| t010hx | HH market value (sum of t0100x over all HH-members; x = implicate a,b,…,e) |
| t020h0 | HH imputation flag for market value |

**Consumer Debts**

| | |
|---|---|
| c100h0 | HH filter information (max of c10000 over all HH-members) |
| c200h0 | HH imputation flag for filter information |
| c010hx | HH market value (sum of c0100x over all HH-members, x = implicate a,b,…,e) |
| c020h0 | HH imputation flag for market value |

**Overall wealth**

| | |
|---|---|
| w010hx | HH gross overall wealth (w010hx = p010Hx + e010Hx + f010Hx + i010Hx + b010Hx + t010Hx; x = implicate a,b,…,e) |
| w020h0 | HH imputation flag for gross overall wealth |
| w001hx | HH overall debts: (w001Hx = p001Hx + e001Hx + c010Hx; x = implicate a,b,…,e) |
| w002h0 | HH imputation flag for overall debts |
| w011hx | HH net overall wealth (w011Hx = w010Hx - w001Hx; x = implicate a,b,…,e) |
| w022h0 | HH imputation flag for net overall wealth |

# 7 Das Arbeiten mit multipel imputierten Daten

Die aufbereiteten Vermögensinformationen des SOEP bestehen aus multipel imputierten Daten, d.h. bei metrisch erhobenen Informationen liegen jeweils fünf so genannte Implicates vor. Während für die Fälle mit direkt beobachteten Informationen in allen Implicates identische Werte abgelegt sind, unterscheiden sich die Implicates bei denjenigen mit fehlenden Antwortangaben. Dies soll die Unsicherheit des Imputationsprozesses widerspiegeln.

Für die Bestimmung von einfachen Kennziffern wie dem Mittelwert oder den Median müssen entsprechende Analysen fünf Mal für jedes Implicate getrennt berechnet werden. Die daraus resultierenden Ergebnisse werden aufsummiert und durch fünf geteilt. Das Ergebnis ist der entsprechende Schätzer auf Basis der multipel imputierten Daten. Anders verhält es sich bei der Berechnung der Varianz. Da die fünf Implicates untereinander korrelieren muss diese Abhängigkeit berücksichtigt und korrigiert werden. Hierbei ist zwischen einer „within" und „between" Komponente zu unterscheiden. Die „Within-Komponente" der Varianz *W* eines Koeffizienten *ß* ist das arithmetische Mittel der auf Basis der *x* = 1 bis 5 Implicates geschätzten Varianzen: $\overline{W} = \frac{1}{5}\sum_{X=1}^{5} \text{var}(\beta')$

Die „Between-Komponente" der Varianz *B* ist die Varianz der *x* = 1 bis 5 geschätzten Koeffizienten *ß*: $B = \frac{1}{5-1}\sum_{X=1}^{5}\left(\hat{\beta}' - \overline{\beta}\right)^2$

Die gesamte Varianz V für den Koeffizienten ß ergibt sich damit aus der Between- und Within-Komponete (bei 5 implicates) wie folgt (siehe auch Leopold und Schneider 2010: 268): $\tilde{V}_{\beta} = \overline{W} + \left(1 + \frac{1}{5}\right) * B$

Ein Beispielprogramm von Arthur B. Kennickell zur Berücksichtigung von multipel imputierten Daten des US Survey of Consumer Finances (SCF) – insbesondere bei Regressionsanalysen – findet sich unter folgender Webadresse (siehe auch Kennickell 1998): <http://www.federalreserve.gov/pubs/oss/oss2/2004/codebk2004.txt>

# 8 Literaturverzeichnis

Frick, J.R., Goebel, J.; Schechtman, E.; Wagner, G.G. and Yitzhaki, S. (2006): Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: The German Socio-Economic Panel Study (SOEP) Experience. Sociological Methods & Research, Vol. 34 (4), S. 427-468.

Frick, J.R., Grabka, M.M. and Sierminska, E.M. (2007): Representative Wealth Data for Germany from the German SOEP: The Impact of Methodological Decisions around Imputation and the Choice of the Aggregation Unit. DIW discussion paper no. 562, Berlin, March.

Frick, J.R. und Grabka, M.M. (2010): Alterssicherungsvermögen dämpft Ungleichheit – aber große Vermögenskonzentration bleibt bestehen. Wochenbericht des DIW, Nr. 3/2010, S. 2-12.

Frick, J.R., Grabka, M.M. und Groh-Samberg, O. (2009): Dealing with incomplete household panel data in microsimulation models. Paper prepared for the 2nd General Conference of the International Microsimulation Association (IMA) Ottawa, June 8-10, 2009. Berlin, http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.334116.de/soep_punr_2009.pdf.

Heckman, James J. (1979): Sample Selection Bias as a Specification Error. Econometrica, Vol. 47(1), S. 153-161.

Kennickell, A. (1998): Multiple Imputation in the Survey of Consumer Finances. Working paper, http://www.federalreserve.gov/pubs/OSS/oss2/papers/impute98.pdf

Leopold, Th. und Schneider, Th. (2010): Schenkungen und Erbschaften im Lebenslauf. Vergleichende Längsschnittanalysen zu intergenerationalen Transfers. Zeitschrift für Soziologie, Vol. 39(4), S. 258-280.

Rubin, D.B. (1976): Inference and missing data. Biometrika, Vol. 63, p. 581-592.

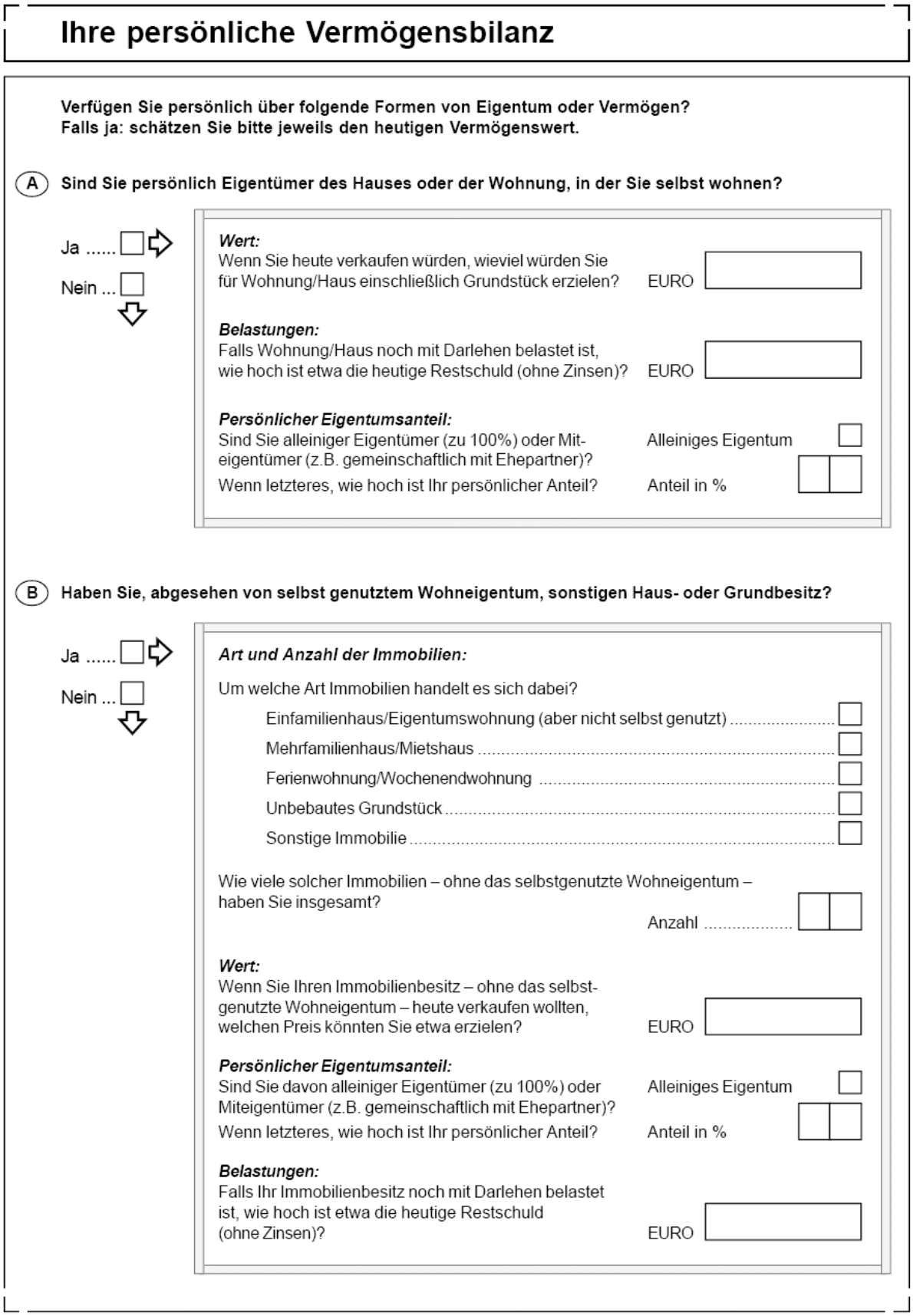
Rubin, D.B. (1987): Multiple Imputation for Nonresponse in Surveys. John Wiley and Sons, New York.

Joachim R. Frick, Markus M. Grabka, Eva M. Sierminska (2010): Examining the gender wealth gap. In: Oxford economic review. Im Erscheinen.

Spiess, M., Goebel, J. (2003): Evaluation of the ECHP imputation rules. CHINTEX Working Paper No. 17.

Starick, R. & N. Watson (2007): Evaluation of Alternative Income Imputation Methods for the HILDA Survey, HILDA Project Technical Paper Series No. 1/07, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Abbildung 5: Auszug aus dem Personenfragebogen des SOEP des Erhebungsjahres 2002



## Ihre persönliche Vermögensbilanz

**Verfügen Sie persönlich über folgende Formen von Eigentum oder Vermögen?**
**Falls ja: schätzen Sie bitte jeweils den heutigen Vermögenswert.**

(A) **Sind Sie persönlich Eigentümer des Hauses oder der Wohnung, in der Sie selbst wohnen?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Wenn Sie heute verkaufen würden, wieviel würden Sie für Wohnung/Haus einschließlich Grundstück erzielen? EURO ☐☐☐☐☐☐

***Belastungen:***
Falls Wohnung/Haus noch mit Darlehen belastet ist, wie hoch ist etwa die heutige Restschuld (ohne Zinsen)? EURO ☐☐☐☐☐☐

***Persönlicher Eigentumsanteil:***
Sind Sie alleiniger Eigentümer (zu 100%) oder Mit-eigentümer (z.B. gemeinschaftlich mit Ehepartner)? Alleiniges Eigentum ☐
Wenn letzteres, wie hoch ist Ihr persönlicher Anteil? Anteil in % ☐☐

(B) **Haben Sie, abgesehen von selbst genutztem Wohneigentum, sonstigen Haus- oder Grundbesitz?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Art und Anzahl der Immobilien:***

Um welche Art Immobilien handelt es sich dabei?

- Einfamilienhaus/Eigentumswohnung (aber nicht selbst genutzt) ..... ☐
- Mehrfamilienhaus/Mietshaus ..... ☐
- Ferienwohnung/Wochenendwohnung ..... ☐
- Unbebautes Grundstück ..... ☐
- Sonstige Immobilie ..... ☐

Wie viele solcher Immobilien – ohne das selbstgenutzte Wohneigentum – haben Sie insgesamt?
Anzahl ..... ☐☐

***Wert:***
Wenn Sie Ihren Immobilienbesitz – ohne das selbst-genutzte Wohneigentum – heute verkaufen wollten, welchen Preis könnten Sie etwa erzielen? EURO ☐☐☐☐☐☐

***Persönlicher Eigentumsanteil:***
Sind Sie davon alleiniger Eigentümer (zu 100%) oder Miteigentümer (z.B. gemeinschaftlich mit Ehepartner)? Alleiniges Eigentum ☐
Wenn letzteres, wie hoch ist Ihr persönlicher Anteil? Anteil in % ☐☐

***Belastungen:***
Falls Ihr Immobilienbesitz noch mit Darlehen belastet ist, wie hoch ist etwa die heutige Restschuld (ohne Zinsen)? EURO ☐☐☐☐☐☐

**C** **Verfügen Sie über Geldanlagen von mehr als 2.500 EURO, etwa in Form von Sparguthaben, Spar- oder Pfandbriefen, Aktien oder Investmentanteilen?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Wie hoch schätzen Sie den Wert Ihrer Geldanlagen insgesamt? EURO ☐

***Persönlicher Eigentumsanteil:***
Sind diese Geldanlagen alle auf Ihren Namen angelegt oder laufen sie teilweise auch auf Gemeinschaftskonten mehrerer Personen, etwa bei Ehepaaren? Alleiniges Eigentum ☐
Wenn letzteres, wie hoch ist Ihr persönlicher Anteil? Anteil in % ☐☐

**D** **Besitzen Sie gegenwärtig Lebensversicherungen oder private Rentenversicherungen oder Bausparverträge?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Wie hoch schätzen Sie den derzeitigen Rückkaufwert dieser Versicherungsverträge bzw. Geldanlagen? EURO ☐

**E** **Sind Sie Eigentümer eines gewerblichen Betriebes, d.h. einer Firma, eines Geschäfts, einer Kanzlei, einer Praxis oder eines landwirtschaftlichen Betriebes, oder an einem solchen Betrieb beteiligt?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Persönlicher Eigentumsanteil:***
Sind Sie in diesem Betrieb alleiniger Unternehmer, oder beteiligter Unternehmer, z.B. nach GBR, GmbH oder KG? Alleiniger Unternehmer ☐ Beteiligter Unternehmer ☐

***Wert:***
Wie hoch schätzen Sie den heutigen Vermögenswert Ihres Betriebes bzw. Ihrer Beteiligung? Das ist der Preis vor Steuern, den Sie bei einem Verkauf des Betriebes bzw. Ihrer Beteiligung erzielen könnten, unter Berücksichtigung eventueller bestehender Kreditbelastungen. EURO ☐

**F** **Verfügen Sie über Sachvermögen von mehr als 2.500 EURO (ohne Kraftfahrzeuge) in Form von Gold, Schmuck, Münzen oder wertvollen Sammlungen?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Angenommen, Sie könnten diese Sachvermögen veräußern:
Wie hoch schätzen Sie den Gesamtwert ein? EURO ☐

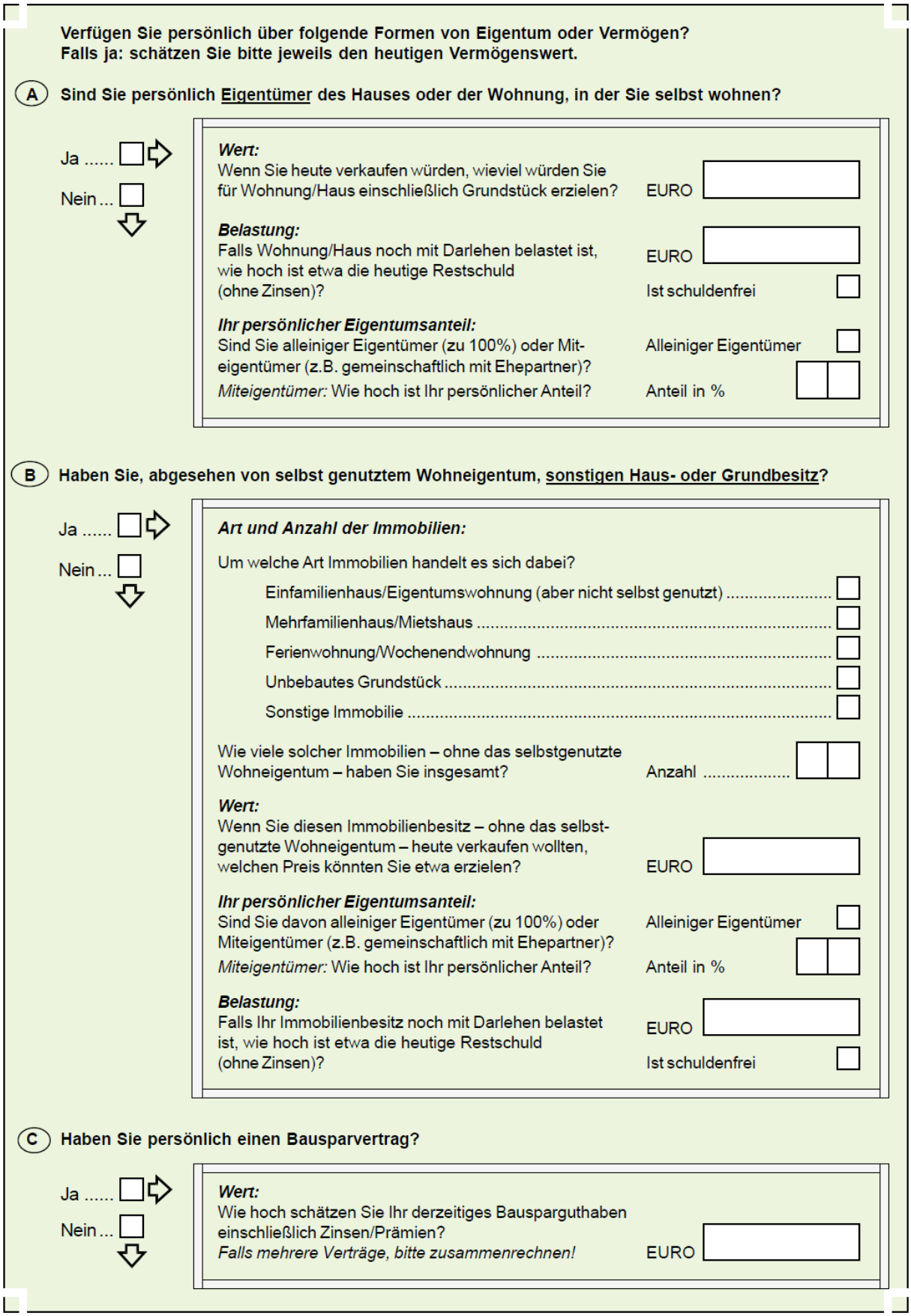
**G** Einmal abgesehen von Hypotheken für Haus- und Grundbesitz oder Baudarlehen: **Haben Sie zur Zeit noch Schulden aus Krediten, die Sie als Privatperson bei einer Bank, einer sonstigen Einrichtung oder bei einer Privatperson aufgenommen haben, und für die Sie privat haften?**
*Gemeint sind nur größere Schulden von 2.500 EURO oder mehr. Ohne Hypotheken und Baudarlehen!*

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Belastung:***
Derzeitige Restschulden (ohne Zinsen): EURO ☐

*Frage **86***
***nächste Seite!***

Abbildung 6: Auszug aus dem Personenfragebogen des SOEP des Erhebungsjahres 2007



**Verfügen Sie persönlich über folgende Formen von Eigentum oder Vermögen?**
**Falls ja: schätzen Sie bitte jeweils den heutigen Vermögenswert.**

**A** **Sind Sie persönlich Eigentümer des Hauses oder der Wohnung, in der Sie selbst wohnen?**

Ja ..... ☐
Nein ... ☐

***Wert:***
Wenn Sie heute verkaufen würden, wieviel würden Sie für Wohnung/Haus einschließlich Grundstück erzielen? EURO

***Belastung:***
Falls Wohnung/Haus noch mit Darlehen belastet ist, wie hoch ist etwa die heutige Restschuld (ohne Zinsen)? EURO
Ist schuldenfrei ☐

***Ihr persönlicher Eigentumsanteil:***
Sind Sie alleiniger Eigentümer (zu 100%) oder Mit-eigentümer (z.B. gemeinschaftlich mit Ehepartner)? Alleiniger Eigentümer ☐
*Miteigentümer:* Wie hoch ist Ihr persönlicher Anteil? Anteil in % ☐☐

**B** **Haben Sie, abgesehen von selbst genutztem Wohneigentum, sonstigen Haus- oder Grundbesitz?**

Ja ..... ☐
Nein ... ☐

***Art und Anzahl der Immobilien:***

Um welche Art Immobilien handelt es sich dabei?

Einfamilienhaus/Eigentumswohnung (aber nicht selbst genutzt) ..... ☐
Mehrfamilienhaus/Mietshaus ..... ☐
Ferienwohnung/Wochenendwohnung ..... ☐
Unbebautes Grundstück ..... ☐
Sonstige Immobilie ..... ☐

Wie viele solcher Immobilien – ohne das selbstgenutzte Wohneigentum – haben Sie insgesamt? Anzahl ..... ☐☐

***Wert:***
Wenn Sie diesen Immobilienbesitz – ohne das selbst-genutzte Wohneigentum – heute verkaufen wollten, welchen Preis könnten Sie etwa erzielen? EURO

***Ihr persönlicher Eigentumsanteil:***
Sind Sie davon alleiniger Eigentümer (zu 100%) oder Miteigentümer (z.B. gemeinschaftlich mit Ehepartner)? Alleiniger Eigentümer ☐
*Miteigentümer:* Wie hoch ist Ihr persönlicher Anteil? Anteil in % ☐☐

***Belastung:***
Falls Ihr Immobilienbesitz noch mit Darlehen belastet ist, wie hoch ist etwa die heutige Restschuld (ohne Zinsen)? EURO
Ist schuldenfrei ☐

**C** **Haben Sie persönlich einen Bausparvertrag?**

Ja ..... ☐
Nein ... ☐

***Wert:***
Wie hoch schätzen Sie Ihr derzeitiges Bausparguthaben einschließlich Zinsen/Prämien?
*Falls mehrere Verträge, bitte zusammenrechnen!* EURO

**D** **Verfügen Sie über Geldanlagen, etwa in Form von Sparguthaben, Spar- oder Pfandbriefen, Aktien oder Investmentanteilen?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Wie hoch schätzen Sie den Wert Ihrer Geldanlagen insgesamt? EURO ☐

***Ihr persönlicher Eigentumsanteil:***
Sind diese Geldanlagen alle auf Ihren Namen angelegt oder laufen sie teilweise auch auf Gemeinschaftskonten mehrerer Personen, etwa bei Ehepaaren? Alleiniger Eigentümer ☐
Wenn letzteres, wie hoch ist Ihr persönlicher Anteil? Anteil in % ☐☐

**E** **Haben Sie eine Lebensversicherung oder eine private Rentenversicherung, die Sie abgeschlossen haben oder Ihr Arbeitgeber für Sie abgeschlossen hat?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Wie hoch schätzen Sie den derzeitigen Rückkaufwert dieser Versicherungsverträge? EURO ☐

**F** **Sind Sie Eigentümer eines gewerblichen Betriebes, d.h. einer Firma, eines Geschäfts, einer Kanzlei, einer Praxis oder eines landwirtschaftlichen Betriebes, oder an einem solchen Betrieb beteiligt?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Ihr persönlicher Eigentumsanteil:***
Sind Sie in diesem Betrieb alleiniger Unternehmer, oder beteiligter Unternehmer, z.B. nach GBR, GmbH oder KG?
Alleiniger Unternehmer ☐
Beteiligter Unternehmer ☐

***Wert:***
Wie hoch schätzen Sie den heutigen Vermögenswert Ihres Betriebes bzw. Ihrer Beteiligung? Das ist der Preis vor Steuern, den Sie bei einem Verkauf des Betriebes bzw. Ihrer Beteiligung erzielen könnten, unter Berücksichtigung eventueller bestehender Kreditbelastungen. EURO ☐

**G** **Verfügen Sie persönlich über nennenswertes Sachvermögen in Form von Gold, Schmuck, Münzen oder wertvollen Sammlungen?**

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Wert:***
Angenommen, Sie könnten diese Sachvermögen veräußern:
Wie hoch schätzen Sie den Gesamtwert ein? EURO ☐

**H** Einmal abgesehen von Hypotheken für Haus- und Grundbesitz oder Baudarlehen: **Haben Sie zur Zeit noch Schulden aus Krediten, die Sie persönlich bei einer Bank, einer sonstigen Einrichtung oder einer Privatperson aufgenommen haben und für die Sie privat haften?**
*Ohne Hypotheken und Baudarlehen!*

Ja ..... ☐ ⇨
Nein ... ☐
⇩

***Belastung:***
Wie hoch sind die derzeitigen Restschulden? EURO ☐

*Frage **127***
***nächste Seite!***