

4 Konsistenzprüfungen und Editierungsmaßnahmen

4.1 Einleitung

Unter Dateneditierung versteht man sowohl die nachträgliche Veränderung von persönlich abgefragten und elektronisch erfassten Beobachtungen, die dazu dient, die bei der Erhebung entstandenen Fehler und logischen Inkonsistenzen zu berichtigen, wie auch die Aufbereitung (Aggregation) der Informationen, die meist aus Gründen der Verständlichkeit bzw. Benutzerfreundlichkeit des Fragebogens in sogenannten Hilfsvariablen erfasst wurden. Der Editierungsprozess trägt somit in elementarer Weise zur Verbesserung von Qualität und Konsistenz der Datensätze bei.¹

Bei Datenerhebungen spiegelt die Rohfassung eines Datensatzes nicht immer jene Informationen wider, die zu erfassen in den Fragen ursprünglich intendiert war. Da es bei den Befragten im HFCS fallweise zu Verständnisschwierigkeiten kam oder unzureichendes Wissen über den Erhebungsinhalt herrschte, konnten auch fehlerhafte Antworten resultieren. Ebenso konnte es durch eine nicht korrekte Dateneingabe der Interviewer (siehe auch Kapitel 3) und bei der Datenverarbeitung zu Fehlern kommen. Im HFCS wurde großes Augenmerk darauf gelegt, Fehler zu minimieren.

Das vorliegende Kapitel gibt Aufschluss über Konsistenzanalysen und Editierungsmaßnahmen im HFCS in Österreich. Zunächst wird die Anzahl der Editierungsmaßnahmen dargelegt (Abschnitt 4.2). Darauf folgen die Erläuterungen zu den Konsistenzprüfungen während bzw. nach dem Interview (Abschnitte 4.3 und 4.4). Im Abschnitt 4.5 werden die Flags beschrieben, in denen die Veränderungen, die an den erhobenen Beobachtungen vorgenommen wurden, transparent gemacht sind. Danach werden detailliert alle Ex-Post-Editierungsmaßnahmen (Abschnitt 4.6) dargestellt und das Kapitel mit Abschnitt 4.7 „Formattierung und Editierung nach den multiplen Imputationen“ und abschließenden Bemerkungen (Abschnitt 4.8) abgerundet.

4.2 Anzahl und Art der Editierungen

In der HFCS-Erhebung wurden von insgesamt rund 841.000 Beobachtungen rund 22.000 Editierungen durchgeführt, d. h., es wurden 2,6% aller Datenpunkte bearbeitet (siehe Tabelle 1, Spalte „Gesamt“).

Die Spalten I bis III geben die unterschiedlichen Arten von Editierungsmaßnahmen an. Lediglich rund 6.900 Beobachtungen wurden bei der Editierung im Wert verändert (siehe Spalte I). Dazu zählen vor allem inkonsistente Werte, die entweder über Nachrecherchen und/oder andere Informationen korrigiert werden konnten oder gelöscht und im Imputationsmodell wieder erstellt wurden. Zwei Drittel der Editierungen (siehe Spalte II), also rund 13.800 Beobachtungen, konnten aus Verbatim-Erfassungen und einer respondentenfreundlichen Abfrage (z. B. Informationen über Lebensversicherungen oder die Höhe des jährlichen Nettoeinkommens) abgeleitet werden. Veränderungen durch Editierungen dieser Art wurden an insgesamt etwa 1,6% der Beobachtungen durchgeführt. Dieser niedrige Wert kann als Indiz für den erfolgreich konzipierten Fragebogen und eine

¹ Siehe z. B. Kennickell (2011) und Bledsoe und Friess (2002) zu Editierungsmaßnahmen im Survey of Consumer Finance des Federal Reserve Systems der USA.

Tabelle 1

Anzahl und Art der Editierungsmaßnahmen

	Alle ¹	Editierungsart		
		Experten- basiert ²	aus Hilfs- variablen ³	Wert gelöscht ⁴
Beobachtungen insgesamt ⁵	840.714	840.714	840.714	840.714
Anzahl der Editierungen	21.837	6.867	13.767	1.203
Anteil der editierten Beobachtungen an der Gesamtanzahl in %	2,6	0,8	1,6	0,1

Quelle: HFCS Austria 2010, OeNB.

¹ Alle Editierungen.

² Anzahl / Anteil der Veränderungen auf Basis von Experteneinschätzungen.

³ Anzahl / Anteil der Veränderungen auf Basis von anderen erhobenen Informationen (z. B. Verbatim-Antworten).

⁴ Anzahl / Anteil der gelöschten Beobachtungen.

⁵ Enthält ausschließlich beobachtbare Werte. Filter-Missings sind nicht berücksichtigt.

intensive Schulung der Interviewer (siehe Kapitel 2 und 3) gewertet werden. In 1.203 Fällen (nur 0,1 % der Beobachtungen) wurden erhobene Werte gelöscht und – teilweise bedingt durch die Editierung einer Head-Variable – durch ein Filter-Missing („.“) ersetzt.

Ein Beispiel² in diesem Zusammenhang ist die Erfassung der Auskunft über das Pensionseinkommen sowohl in der Variable „Unselbstständiges Einkommen“ als auch bei „Pensionseinkommen“. Die Head-Variable des unselbstständigen Einkommens (PG0100) wurde auf „Nein“ editiert und der eingetragene Wert (PG0110) in dieser Einkommensvariable gelöscht, da dieses Einkommen in der Variable des Pensionseinkommens (PG0300 und PG0310) schon korrekt erfasst worden war.

4.3 Konsistenzprüfungen während der Befragung

Der HFCS basierte auf computergestützter persönlicher Befragung (Computer Assisted Personal Interviews – CAPI). Diese Erhebungsmethode bringt gegenüber der Verwendung von Papierfragebögen oder telefonischer Erhebung zahlreiche Vorteile. Der Interviewer verwendete bei der Befragung ein Notebook, auf dem die Erhebungs-Software lief, und durchlief den Fragebogen auf dem Bildschirm. Die erhobenen Informationen wurden bereits während des Interviews auf Zulässigkeit und Konsistenz geprüft. Verständnisfragen des Befragten konnten durch den Interviewer bzw. die bereitgestellten Unterlagen unmittelbar beantwortet und Fehler schon bei der Dateneingabe vermieden werden.

Konsistenzprüfungen im Interviewverlauf sind jedoch sowohl zahlenmäßig als auch inhaltlich beschränkt. Zu viele Konsistenzprüfungen würden das Interview übermäßig verlängern und die Befragten ermüden, wodurch wiederum die Qualität der erhobenen Daten sinken würde. Im Einzelfall müssten Interviews möglicherweise sogar abgebrochen werden.

Eine inhaltliche Einschränkung für mögliche Konsistenzprüfungen besteht darin, dass alle Informationen, die dafür herangezogen werden sollten, bereits vorhanden sein müssten. Für einfache Konsistenzprüfungen auf Basis von Informationen aus der jeweils abgefragten Variable gelten diese Einschränkungen

² Beispiele in diesem Kapitel werden in eingerückter Form dargestellt.

nicht. Über- oder unterschreitet ein vom Befragten angegebener Wert gewisse Grenzen, erscheinen unmittelbar auf dem Bildschirm Warnhinweise, die eine rasche Prüfung des eingegebenen Werts ermöglichen. Bei komplexeren Konsistenzprüfungen hingegen, die Informationen aus einer Vielzahl weiterer Fragen verwenden, liegen oft erst im fortgeschrittenen Interviewverlauf die notwendigen Antworten vor.

Im HFCS wurden in der digitalen Version des Fragebogens mehr als 150 Konsistenzprüfungen³ programmiert. Die überwiegende Mehrheit davon bildeten sogenannte Soft-Checks. Bei einer Verletzung der Prüfbedingung öffnete sich auf dem Bildschirm des Notebooks ein Fenster, in dem auf die unstimmgige Antwort hingewiesen wurde.

Gab etwa ein Haushalt mit einem verfügbaren monatlichen Haushalts-Nettoeinkommen von 1.000 EUR an, dass – neben Konsumausgaben in Höhe von insgesamt 900 EUR – die im vergangenen Jahr typischerweise pro Monat an Personen außerhalb des Haushalts geleisteten Zuwendungen 200 EUR betragen hatten, so erschien in der Fragebogenmaske folgender farblich abgehobener Hinweis:

„Die regelmäßigen Zuwendungen an Personen außerhalb des Haushalts zusammen mit den Konsumausgaben waren höher als das Haushalts-Nettoeinkommen. Stimmen die Angaben? Bitte bestätigen Sie bei Richtigkeit, andernfalls korrigieren Sie bitte die Angabe/n.“

Da sich die abgefragten Beträge auf unterschiedliche Zeitpunkte bezogen haben könnten, oder die Zuwendungen aus der Veräußerung von Vermögen finanziert worden sein könnten oder das Haushaltseinkommen etwa infolge des Jobverlusts eines oder mehrerer Mitglieder eines Haushalts gesunken sein könnte, konnte die Richtigkeit dieser Angaben nicht ausgeschlossen werden. Im konkreten Fall wurde der Befragte daher vom Interviewer aufgefordert, sowohl die angegebene Höhe des Haushaltseinkommens als auch jene der Zuwendungen und Konsumausgaben zu bestätigen oder zu korrigieren.

Im Gegensatz dazu wurden in der digitalen Version des HFCS-Fragebogens in Österreich auch Konsistenzprüfungen programmiert, bei denen ein Weiterkommen im Fragebogenverlauf nur dann möglich war, wenn eine als unmöglich oder inkonsistent erkannte Antwort korrigiert wurde. Diese sogenannten Hard-Checks wurden jedoch nur in jenen Fällen eingesetzt, in denen bestimmte Antworten definitiv ausgeschlossen werden konnten.

Gab etwa eine Person an, seit 40 Jahren in Österreich zu wohnen, aber erst 30 Jahre alt zu sein, so erschien auf dem Bildschirm die farblich abgehobene Fehlermeldung:

„Die befragte Person lebt länger in Österreich als sie alt ist. Das ist nicht möglich. Bitte korrigieren Sie die Angabe/n.“

Erst nach Abänderung des Alters auf zumindest 40 Jahre oder aber des Verweilzeitraums in Österreich auf maximal 30 Jahre (oder Änderungen in beiden Variablen) war im beschriebenen Fall ein Weiterkommen im CAPI-Fragebogen möglich.

³ Eine Liste aller Konsistenzprüfungen, die für die digitale Version des Fragebogens programmiert wurden, ist im Online-Anhang abzurufen.

4.4 Konsistenzprüfungen nach der Befragung

4.4.1 Expertenbasierte Analysen der Daten

Während der Feldphase des HFCS in Österreich wurden die Daten der vom Erhebungsinstitut als abgeschlossen angesehenen Haushalte in insgesamt sieben Tranchen an die OeNB übermittelt. Alle Daten wurden unmittelbar nach jeder Tranche expertenbasierten Analysen⁴ unterzogen. Diese Untersuchungen dienten zum einen der Erhöhung der Datenkonsistenz innerhalb eines Haushalts und zum anderen der Prüfung der Erhebungs-Software (insbesondere der Überwachung der Fragebogen-Programmierung) und des Aufbereitungsmechanismus des Erhebungsinstituts.

Die Datensätze erfolgreich interviewter Haushalte wurden wie auch jene von Haushalten, die eine Teilnahme verweigerten, einzeln betrachtet. Damit konnte auch das Engagement der Interviewer, die Haushalte von einer Teilnahme zu überzeugen hatten, kontrolliert und optimiert werden. Eine Selektion von einfachen bzw. leichter erreichbaren Haushalten durch die Interviewer und damit verbundene Verzerrungen (z. B. Hausfrauen- oder Pensionistenbias) konnten somit weitestgehend ausgeschlossen werden. Den Interviewern war bewusst, dass über die 4.436 gezogenen Haushalte der Bruttostichprobe hinaus keine weiteren Adressen zur Verfügung standen. So wurde sichergestellt, dass Interviewer nicht weniger schwierige Haushalte auswählten und dann auf zusätzliche Adressen zurückgreifen konnten. Der Anreiz der Interviewer, mit dem von vornherein strikt beschränkten Adressmaterial möglichst effizient umzugehen, wurde auch durch die leistungsorientierte Bezahlung und den relativ hohen Aufwand zur Teilnahme als Interviewer in der Erhebung unterstützt. Auch die Gebietsleiter wurden angewiesen zu vermeiden, dass Interviewer neue Haushalte zugeteilt bekamen, ohne vorher angemessenen Aufwand für die Abwicklung der bereits in Bearbeitung befindlichen Haushalte betrieben zu haben. Keine Nachziehungen (Ersatzhaushalte) zuzulassen, ist eines der wichtigsten Kriterien für eine erfolgreiche Erhebung. Dies ist auch zur Gewährleistung der Repräsentativität der Stichprobe unerlässlich (Vehovar, 1999).

Schon während der Feldphase wurden Informationen jedes einzelnen Haushalts zur geografischen Lage und Struktur, zu Finanz- und Sachvermögen, Verbindlichkeiten, Einkommen, Eigentumsübergang (Erbschaft bzw. Schenkung), allfällige Kommentare des Haushalts oder der Interviewer sowie Dauer, Datum und Zeitpunkt des Interviews analysiert. Diese Informationen ermöglichten eine rasche erste Einschätzung der Qualität des Interviews. Alle einen einzelnen Haushalt betreffenden Mikrodaten wurden auf inhaltliche Konsistenz geprüft und von zumindest zwei Ökonomen untersucht. Unklarheiten wurden im Team besprochen und weitere Vorgehensweisen festgelegt.

Darüber hinaus wurden in diesem Prozessschritt auch die Interviewer (siehe auch Kapitel 3) geprüft und auf Fehler oder Verständnisschwierigkeiten hingewiesen. Häufig waren dies kleinere Mängel, allerdings kam es auch zum Ausschluss von drei Interviewern, deren Interviewergebnisse die Qualitätskriterien (z. B. in Bezug auf Antwortverweigerung) nicht erfüllten.

⁴ Diese Evaluierungen erfolgten unter Zuhilfenahme externer Datenquellen wie etwa der OeNB-Immobilienvermögenserhebung 2008 oder EU-SILC (erhoben von Statistik Austria).

4.4.2 Nachträgliche Recherchen

War aus der Datenanalyse nicht eindeutig erkennbar, welcher Art ein Fehler war, wurden die Haushalte vom Erhebungsinstitut erneut kontaktiert, um die bestehenden Unklarheiten zu klären und eine korrekte Erfassung zu gewährleisten. Ein typischer Fall eines leicht zu erkennenden Problems, das keine Nachrecherche erforderte, war das Umschreiben eines negativen Girokontostands auf eine positive Verbindlichkeit (Kontoüberzug) bei gleichzeitiger Nullsetzung des Vermögens auf dem Girokonto (siehe auch Abschnitt 4.6). Dabei handelte es sich um eine einfache Konvention, an welcher Stelle derartige Verbindlichkeiten erfasst werden sollen. Grundlage der Entscheidung für die Nachrecherche war stets das Ziel, die Daten ex post möglichst wenig zu editieren, gleichzeitig aber auch den teilnehmenden Haushalten nicht unnötige Bürden aufzuerlegen. Insgesamt wurden Einzelinformationen von rund 400 Haushalten nachrecherchiert. Oft wurden auch ungewöhnliche Werte (etwa besonders hohe Vermögenswerte) durch Nachrecherchen bestätigt oder korrigiert.

4.4.3 Untersuchung von Extremwerten

Im Rahmen der individuellen Analysen wurde besonderes Augenmerk auf die Erkennung und Bearbeitung von sogenannten Outliern (überaus niedrige oder hohe Werte) gelegt. Insbesondere bei Vermögensvariablen, beim Einkommen oder etwa bei der Wohnungsgröße kamen Outlier vor. Diese Werte waren – soweit sie im Datensatz bestehen blieben – nicht auf Fehler im Zuge der Erhebung zurückzuführen, sondern wurden größtenteils in Nachrecherchen bestätigt. Es wird daher empfohlen, bei Untersuchungen auf Basis der HFCS-Daten sogenannte Outlier nicht generell von den Analysen auszuschließen, sondern diese unter Anwendung adäquater Methoden in die Berechnungen miteinzubeziehen.

4.4.4 Technische Prüfung der Filterführung bzw. Konsistenz

Neben den in der digitalen Version des Fragebogens programmierten Konsistenzprüfungen wurden die Daten aller Haushalte auch während der Feldphase zusätzlich zur Kontrolle durch Experten in einem automatisierten Prozess eingehend auf ihre Konsistenz geprüft.

So wurden durch den wiederholten Einsatz aller Hard-Checks die Beobachtungen daraufhin untersucht, ob Antworten ermöglicht wurden, die ein Weiterkommen im Fragebogenverlauf ausschließen und daher abgeändert hätten werden müssen. Kam es zu fehlerhafter Programmierung einzelner Hard-Checks, so wurde das Erhebungsinstitut darüber informiert und die Behebung der Programmierungsfehler veranlasst.

Zudem wurde die gesamte Filterstruktur des Fragebogens geprüft, um durch Programmierfehler bedingte umfangreiche und kostspielige Nacherhebungen zu vermeiden. Nach intensiven, vor Beginn der Feldphase durchgeführten Tests der Fragebogen-Programmierung wie auch einer Test-Feldphase mit 50 Haushalten konnten auf diese Art und Weise kleine Programmierfehler erkannt und behoben werden. So wurden z. B. in Haushalten, in denen zwei Personen die Angabe ihres Lebensalters verweigerten, für die zweite Person die meisten Personenfragen

Tabelle 2

Flags im HFCS in Österreich

Block I	0	Nicht zutreffend (kein Wert aufgrund eines Filters)
	1	Wert erhoben, vollständige Beobachtung
	2	Wert erhoben, aber in den Schleifen umgeordnet
Block II	1050	Nicht imputiert, erhoben: „Weiß nicht“
	1051	Nicht imputiert, erhoben: „Keine Angabe“
	1052	Nicht imputiert, nicht erhoben aufgrund von Missing in der Filtervariable
	1053	Nicht imputiert, erhoben als Intervall
	1054	Nicht imputiert, erhobener Wert gelöscht oder nicht erhoben aufgrund eines CAPI-Fehlers
	1055	Nicht imputiert, Wert Missing gesetzt aufgrund einer falschen Antwort in der Filtervariable
Block III	2050	Fehlender Wert, editiert aufgrund von Anonymisierung
	2051	Fehlender Wert, Daten wurden nicht erhoben
Block IV	3050	Editiert, erhobener Wert nicht korrekt
	3051	Editiert, Information aus anderen (national erhobenen Variablen) verwendet
	3052	Editiert, Information aus Verbatim-Erfassung verwendet
	3053	Editiert, auf Filter-Missing („.“) gesetzt
Block V	4050	Imputiert, erhoben: „Weiß nicht“
	4051	Imputiert, erhoben: „Keine Angabe“
	4052	Imputiert, nicht erhoben aufgrund von Missing in der Filtervariable
	4053	Imputiert aus erhobener Intervallangabe
	4054	Imputiert, erhobener Wert gelöscht, kein Wert aufgrund eines Fehlers erhoben
	4055	Imputiert, nicht erhoben aufgrund einer falschen Antwort in der Filtervariable

Quelle: HFCS Austria 2010, OeNB.

nicht erfasst (siehe Abschnitt 4.6.2.12).⁵ Mithilfe dieser Filterchecks wurde auch eine durchgehend konsistente Kodierung der Variablen sichergestellt.⁶

4.5 Flags

In sogenannten Flag-Variablen wurden alle Editierungen (und Imputationen, siehe Kapitel 5) dokumentiert. Diese Variablen enthalten sämtliche Informationen zur Entstehung jeder einzelnen Beobachtung im HFCS. Zur Klassifizierung wurden Flags verwendet, die in Tabelle 2 angeführt sind. Um den internationalen Vorgaben Rechnung zu tragen, wurden manche Flags für den international erhältlichen Datensatz aggregiert (siehe Abschnitt 4.7).

Block I

Die Flags in Block I wurden für Informationen verwendet, die erhoben wurden. So bekamen alle Werte, die während des Interviews in die Erhebungs-Software eingetragen wurden, das Flag 1 und alle Filter-Missings, also „.“-Beobachtungen das Flag 0. In Schleifen erfasste Informationen (siehe Abschnitt 4.6.2.4) wurden, wenn notwendig, in der Iteration der Schleife umgeordnet und bekamen das Flag 2. Dies bedeutet, dass Beobachtungen mit dem Flag 2 so im Datensatz enthalten sind, wie sie erhoben wurden; es wurde lediglich die Iterationsnummer der Schleife geändert.

⁵ Dieses Problem wurde bald nach der ersten Datenteillieferung eruiert und behoben. Daher beschränkt es sich auf wenige Haushalte.

⁶ Alle Variablen des HFCS enthalten Value-Labels, die die Kodierung erläutern. Auch der Fragebogen (im Online-Anhang abzurufen) enthält die Kodierung der einzelnen Variablen.

Block II

Für alle Beobachtungen, die Flags mit den Werten aus Block II haben, wurden unzureichende oder keine vollständigen Informationen erfasst. Darunter fallen Informationen, bei denen der Befragte die Antwort verweigerte („Weiß nicht“ oder „Keine Angabe“) oder keine genaue Betragsangabe machen konnte und stattdessen lediglich ein Intervall angab. Ebenfalls sind hier Beobachtungen enthalten, die aufgrund von Editierungsmaßnahmen an dieser oder an einer Head-Variable (Flags 1054 und 1055) nicht vorhanden waren. Beobachtungen mit Flags aus diesem Block wurden nicht imputiert (siehe Kapitel 5).

Block III

Beobachtungen bzw. Variablen, die mit den Flags aus Block III kodiert sind, wurden nicht erhoben bzw. aufgrund von Anonymisierungsvorschriften nach der Erhebung aus den Datensätzen entfernt.

Block IV

Die Flags in Block IV beschreiben eine Ex-Post-Editierungsmaßnahme des beobachteten Werts. Hier wird unterschieden zwischen logischen Inkonsistenzen (Flag 3050), Berechnungen aufgrund von anderen im Survey erhobenen Informationen (z. B. Lebensversicherungen, siehe Erläuterung im Abschnitt 4.6.2.9) mit dem Flag 3051, Ex-Post-Kodierungen von verbatim erfassten Informationen (siehe Abschnitt 4.6.2.3) mit dem Flag 3052 und Editierungsmaßnahmen, die einen Wert löschen und die Beobachtung auf Filter-Missing setzen (z. B. Doppelnennungen) mit dem Flag 3053. Alle Beobachtungen, die aufgrund von Nachrecherchen korrigiert werden konnten, wurden mit dem Flag 3050 versehen.

Block V

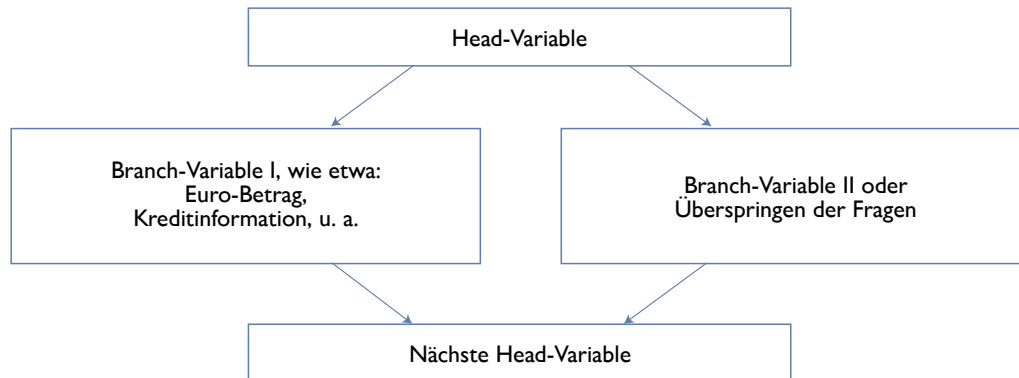
Die Flags in Block V stellen das Äquivalent zu den Flags von Block II dar. Konnten die fehlenden Werte imputiert werden, erhielten die Beobachtungen je nach Ursprung ein Flag, das mit der Ziffer 4 beginnt. Wurde z. B. von einem Befragten eine Betragsangabe im Intervall erfasst und dann imputiert, erhielt diese Beobachtung nach den multiplen Imputationen das Flag 4053. Damit sind alle Informationen auch nach den Imputationen noch nachvollziehbar.

In Grafik 3 lässt sich erkennen, wie im HFCS-Fragebogen die Struktur vieler Fragen aufgebaut ist. Anhand eines Beispiels (Einkommen aus unselbstständiger Beschäftigung) lässt sich sowohl der Aufbau einzelner Fragenblöcke⁷ als auch die Verwendung der Flags gut erkennen.

Die Head-Variable zur Erfassung des Einkommens aus unselbstständiger Beschäftigung eruiert, ob ein Haushalt Einkommen in dieser Form bezieht (Ja/Nein-Frage). Wurde diese Frage positiv beantwortet, wurde die Einkommenshöhe erfasst und gemäß Fragebogen mit der nächsten Head-Variablen, in diesem Fall der Frage nach dem Einkommen aus selbstständiger Beschäftigung, fortgesetzt. Bezog ein Haushalt kein Einkommen dieser Form oder verweigerte er die Aussage („Weiß nicht“ oder „Keine Angabe“), wurde die Befragung mit der Frage nach dem Selbstständigen-Einkommen (nächste Head-Variable) fortgeführt. Alle erhobenen Beob-

⁷ Für den Aufbau des gesamten Fragebogens siehe Kapitel 2.

Abfolge von Fragen



Quelle: HFCS Austria 2010, OeNB.

achtungen bekamen je nach Fragbogenverlauf zunächst entweder das Flag 1 oder 0. Wurde ex post (z. B. durch Informationen über die Beschäftigung) festgestellt, dass die Frage nach dem unselbstständigen Einkommen fälschlicherweise mit „Nein“ beantwortet worden war, wurde diese Antwort korrigiert, die Beobachtung mit dem Flag 3050 („Editiert, erhobener Wert nicht korrekt“) versehen, die folgende Angabe zur Höhe des Einkommens zur Imputation freigegeben und nachdem die Höhe imputiert worden war, mit Flag 4055 („Imputiert, nicht erhoben wegen falscher Antwort in Filtervariable“) versehen.

Wurde z. B. in der Frage nach dem höchsten erreichten Bildungsabschluss einer Person (Variable (A)PA0200) in einem Haushalt die Kategorie „Anderer Abschluss“ gewählt und konnte ex post einer der vorgegebenen Kategorien zugeordnet werden, trug diese Beobachtung in der Flag-Variable des Personen-Datensatzes die Nummer 3052 (also „Editiert aus den Informationen der Verbatim-Erfassung“).

Durch die Verwendung von Flags kann die Entstehung jeder einzelnen Beobachtung im HFCS nachvollzogen werden. Sowohl die Variablen zur Haushalts- und Personen-Identifikation als auch die Erfassung der Länderkennzeichnung und die Iterationsnummer der Imputation wurden aber nicht mit diesen Flags kodiert. Die beschriebenen Flags sind um einige Kategorien spezifischer als die Flags des von der EZB erhältlichen internationalen HFCS-Datensatzes. Die Aggregation der Flags aufgrund der internationalen Konsistenz wurde vor der Übermittlung an die EZB durchgeführt (siehe Abschnitt 4.7).

4.6 Ex-post-Editierung

4.6.1 Einzelfallbetrachtung

Da jeder Haushalt mit großer Sorgfalt einzeln untersucht wurde, konnten Inkonsistenzen gefunden und mithilfe von Nachrecherchen sowie ex post durch Editierungen bereinigt werden. Bei diesen individuellen Auswertungen spielten auch deskriptive Statistiken (wie etwa zur Höhe des durchschnittlichen Einkommens) der bis zu diesem Zeitpunkt im HFCS befragten Haushalte sowie externer Datenquellen eine Rolle in Bezug auf die Plausibilität der neu erhobenen Infor-

mationen im Vergleich zu bekannten Kennzahlen. Dabei war auch die Verwendung von Hilfsvariablen, in denen Werte in aggregierter Weise bzw. auf unterschiedliche Weise erhoben wurden, von wesentlichem Interesse.

Interviewer mit unregelmäßigen Interviews (siehe Kapitel 3) und Nachrecherchen des Erhebungsinstituts wurden genauer untersucht. Durch expertenbasierte Analysen wurden im Allgemeinen folgende Unklarheiten ex post editiert:

- Doppelnennungen: Wurde eine Erbschaft z. B. sowohl in „Hauptwohnsitz geerbt“ als auch im Kapitel „Erbschaft“ erfasst oder dasselbe Einkommen in zwei verschiedenen Kategorien von Einkommen angeführt, war dies zu korrigieren.
- Nullen-Problem: Interviewer gaben in wenigen Fällen bei Betragsangaben irrtümlich eine Null zu viel bzw. zu wenig ein.
- Eigentumsstruktur des Immobilienvermögens im Zusammenhang mit der Intra-Haushalts-Verteilung dieser Vermögensposition: Einige wenige Kompetenzträger bezogen die Frage nach den Eigentumsverhältnissen im Hauptwohnsitz nicht wie vorgesehen auf den Haushalt, sondern auf die Einzelperson des Kompetenzträgers.
- Unplausible Werte, die nicht erfolgreich nachrecherchiert werden konnten, wurden auf Missing editiert und danach imputiert.
- Fehlerhafte Eingaben von Interviewern wurden bereinigt: Zum Beispiel wurde ein Interview, das im Jahr 2011 stattfand, aber als Monat des Kontaktversuchs Oktober 2011 (also 10) auswies, auf Jänner 2011 (also 1) gesetzt, da die Feldphase im Juni beendet war und dieses Interview im Februar übermittelt wurde. Diese Editierungsmaßnahmen betrafen den gesamten Fragebogen und nicht nur einzelne Variablen. Die Eingriffe in die erhobenen Daten wurden minimiert und – sofern Nachrecherchen bzw. Hilfsvariablen (wie z. B. Verbatim-Erfassungen) keine andere Information erbrachten – eine inkonsistente Beobachtung auf Missing gesetzt und zur Imputation freigegeben. Inkonsistente oder unplausible Beobachtungen wurden mit großer Vorsicht bearbeitet und lediglich dann entfernt, wenn mit an Sicherheit grenzender Wahrscheinlichkeit von Inkonsistenz auszugehen war.

4.6.2 Strukturelle Editierung

4.6.2.1 Cleaning

Bei der Beantwortung von HFCS-Fragen kam es vor, dass Befragte sich irrten, dann im Verlauf der Fragebogenstruktur zurückgeführt wurden und Antworten entsprechend korrigierten. Aufgrund dieser Richtigstellungen musste auch der nachfolgende Verlauf des Fragebogens abgeändert werden, weil die neu gegebenen Antworten andere Filterführungen bedingten. Die ursprünglich falsch eingeschlagenen Wege durch den Fragebogen blieben allerdings zur Nachvollziehbarkeit bestehen und mussten nachträglich bereinigt werden.

4.6.2.2 Umrechnung von Währungsangaben

Alle Betragsangaben konnten von den Befragten in unterschiedlichen Währungen angegeben werden (siehe Kapitel 2). Im Folgenden beschriebene Editierungsmaßnahmen beziehen sich sowohl auf die Angabe von exakten Werten als auch von eigens gewählten Intervallen (fix vorgegebene Intervalle konnten nur in Euro angegeben werden).

Im Regelfall wurden alle Währungen entweder in Euro oder in Schilling angegeben. Insbesondere der Wert des Hauptwohnsitzes (sowohl der Kaufpreis als auch der aktuelle Wert) wurde häufig in Schilling angegeben. Für die nachfolgende Umrechnung wurde der unwiderrufliche Umrechnungskurs⁸ (1 EUR = 13,7603 ATS) verwendet. Darüber hinaus wurden einige Beträge in Deutscher Mark angegeben. Auch hier wurde der unwiderrufliche Umrechnungskurs der EZB (gleiche Referenz wie in Fußnote 8) verwendet (1 EUR = 1,95583 DEM).

Zu Fremdwährungskrediten wurden auch Angaben in japanischen Yen und Schweizer Franken gemacht. Der zum Zeitpunkt der Befragung angegebene Wert des ausstehenden Kreditvolumens wurde auf Basis des Mittelwerts des Jahres 2010 umgerechnet. Das Kreditvolumen zum Zeitpunkt der Kreditaufnahme wurde mit dem Mittelwert des Wechselkurses jenes Jahres, in dem der Kredit aufgenommen worden war, umgerechnet. Als Referenz dienen die auf der Website der OeNB⁹ publizierten Wechselkurse.

4.6.2.3 Verbatim-Erfassung

Bei vielen Fragen konnten die Befragten als Antwort die Kategorie „Sonstiges“ wählen und den jeweiligen Sachverhalt wörtlich in einer Verbatim-Variable erfassen lassen. Die Vorteile dieser Form der Erfassung lagen vor allem in der Benutzerfreundlichkeit des Fragebogens für den Befragten. War eine Zuordnung der Antwort des Befragten zu den vorgegebenen Kategorien (während des Interviews) nicht möglich, konnte eine wörtliche Beschreibung erfasst werden. Die wörtlichen Informationen wurden dazu genutzt, eine nachträgliche Zuordnung zu den einzelnen Antwortmöglichkeiten einer Frage vorzunehmen, was zumeist möglich war. Wenn nicht, wurde die Beobachtung in der Kategorie „Sonstiges“ belassen. Alle Beobachtungen, bei denen eine Verbatim-Information zur Ex-Post-Editierung genutzt wurde, wurden mit dem Flag 3052 (siehe Abschnitt 4.5 zu den Flags) versehen.

4.6.2.4 Schleifen-Bearbeitung

Wie in Abschnitt 2.6.1 ausführlich dargelegt, fand die Erfassung mancher Informationen in Form von Schleifen statt. Eine Schleife beschreibt die Abfrage eines identischen Sets von Fragen für jedes einzelne Item einer Gruppe von Items, die sich im Eigentum des Haushalts befinden. Folgende Items wurden in Form einer Schleife erhoben:

- mit dem Hauptwohnsitz hypothekarisch besicherte Kredite
- weitere Immobilien neben dem Hauptwohnsitz
- mit den weiteren Immobilien hypothekarisch besicherte Kredite
- unbesicherte Kredite
- Unternehmen im Eigentum des Haushalts
- Erbschaften und Schenkungen

Im Folgenden werden die Editierungsmaßnahmen, die aufgrund der Abfrage in Schleifenform notwendig wurden, beschrieben.

⁸ Siehe <http://www.oenb.at/isaweb/report.do?lang=DE&report=2.12> zu den unwiderruflichen Euro-Umrechnungskursen (abgerufen am 22. Jänner 2013).

⁹ Siehe http://www.oenb.at/de/stat_melders/datenangebot/zinssaetze/wechselkurse/wechselkurse.jsp#tcm:14-147886 (abgerufen am 22. Jänner 2013).

Reihenfolge

Die Reihenfolge bei der Erfassung der einzelnen Items in einer Schleife war vorgegeben. Zum Beispiel wurde mit dem wertmäßig höchsten noch ausstehenden und mit dem Hauptwohnsitz hypothekarisch besicherten Kredit begonnen, dann mit der Frageschleife für den nächsthöheren Kredit und schließlich für den dritthöchsten ausstehenden Kredit fortgesetzt. Diese Reihenfolge wurde von einigen Befragten in den in Schleifen abgefragten Vermögens- und Verbindlichkeitsbestandteilen nicht immer eingehalten. Im Zuge des Editierungsprozesses wurden solche Fälle entsprechend umkodiert. Einzig in der Erbschaftsschleife wurde auf eine derartige Umsortierung verzichtet. Denn hier waren die Befragten zu Beginn der ersten Schleife aufgefordert worden, die erhaltenen Erbschaften nach ihrer Bedeutung für die derzeitige Vermögenssituation zu ordnen. Die innerhalb der Erbschaftsschleifen abgefragten Werte sollten sich jedoch auf den Zeitpunkt des Erhalts der Erbschaft beziehen. Zwischen Erhalt der Erbschaft und Zeitpunkt der Erhebung konnten bestimmte Erbschaften mehr an Wert gewonnen (verloren) haben als andere oder es konnte z. B. eine geerbte Wohnimmobilie bereits an Kinder weitergegeben worden sein und damit für die momentane Vermögenssituation keine Bedeutung mehr haben.

In den Flags wurde jede Beobachtung einer Variable einer Schleife, die mit den Einträgen derselben Variablen aus einer anderen Schleife überschrieben wurde, mit dem Code 2 (siehe Abschnitt 4.5) versehen. Wurde ein Filter-Missing einer Variable innerhalb einer Schleife mit dem Filter-Missing derselben Variable aus einer anderen Schleife überschrieben, wurde der Code 0 „Nicht zutreffend – Übersprungen aufgrund eines Filters“ verwendet.

Überspringen von Fragen

Um den Abbruch eines Interviews innerhalb einer Schleife zu verhindern, wurde es den Befragten ermöglicht, Aspekte der Fragen in den Schleifen zu überspringen und direkt zur Zusammenfassungsfrage zu gelangen. Dort wurde entweder die Residualsumme der restlichen, noch nicht erfassten Kredite bzw. Unternehmen (mehr als drei Kredite bzw. Unternehmen) oder die Summe aller (maximal drei) Kredite bzw. Unternehmen erfasst. In der Schleife der Erbschaften und Schenkungen wurde in der Zusammenfassungsfrage nach dem Überspringen von Fragen in der Schleife immer nach der Summe aller Erbschaften gefragt. Da die Zusammenfassungsfrage in allen Abschnitten im an die EZB gelieferten Datensatz nur die über die ersten drei hinausgehenden Kredite, Immobilien, Erbschaften und Schenkungen enthalten sollte, mussten hier Editierungsmaßnahmen getroffen werden. Zur Erleichterung der Lesbarkeit werden diese beispielhaft anhand des Fragebogenabschnitts zu den unbesicherten Krediten beschrieben (siehe Abschnitt 2.5).

In den 21 Fällen, in denen ein Haushalt nur einen unbesicherten Kredit aufgenommen hatte und Fragen einer Schleife übersprang, hing die getroffene Editierungsmaßnahme davon ab, ob (1) nur in der Zusammenfassungsfrage ein Betrag zur Höhe des ausstehenden Kredits genannt wurde oder (2) sowohl in der ersten Kreditschleife als auch in der Zusammenfassungsfrage ausstehende Kreditbeträge angegeben wurden oder (3) weder innerhalb der Schleife zum ersten Kredit noch in der Zusammenfassungsfrage Beträge erfasst wurden. Wurde (1) kein Betrag bei der Frage

nach der Höhe des einzigen ausstehenden Kredits angegeben, dafür aber bei der Zusammenfassungsfrage, dann wurde der genannte Betrag als Antwort zur Frage nach der Höhe des einen ausstehenden Kredits (in der ersten Schleife) eingetragen. Die Angabe in der Zusammenfassungsfrage wurde dann auf Filter-Missing gesetzt. Sofern (2) der innerhalb der Kreditschleife genannte Betrag jenem der Zusammenfassungsfrage entsprach, wurde letzterer als Filter-Missing kodiert, da es sich um eine Doppelbenennung handelte.¹⁰ Wurde (3) weder innerhalb der Schleife noch an ihrem Ende ein Betrag genannt, wurde nur die Zusammenfassungsfrage als Filter-Missing editiert.

Gab ein Haushalt an, zwei unbesicherte Kredite aufgenommen zu haben und übersprang Fragen der Schleife,¹¹ so wurde die Editierungsmaßnahme in Abhängigkeit davon getroffen, ob (1) der Wert des höchsten ausstehenden Kredits und anschließend ein Betrag bei der Zusammenfassungsfrage angegeben wurde oder (2) sowohl in den zwei Kreditschleifen als auch in der Zusammenfassungsfrage Beträge genannt wurden oder (3) nur in der Zusammenfassungsfrage ein Wert erfasst wurde oder (4) weder innerhalb der Schleifen noch in der Zusammenfassungsfrage Angaben zur Höhe ausstehender Kredite gemacht wurden.

(1) wurde für die ausstehende Kredithöhe des zweithöchsten Kredits die Differenz zwischen dem in der Zusammenfassungsfrage und jenem in der ersten Kreditschleife genannten ausstehenden Kreditbetrag eingesetzt. Dies allerdings nur dann, wenn die angegebene Summe der zwei ausstehenden Kredite größer als jene des ersten Kredits war. War sie kleiner, so wurde angenommen, dass in der Zusammenfassungsfrage nicht die Summe der ausstehenden Kredite, sondern nur die Höhe des verbleibenden zweiten ausstehenden Kredits angegeben wurde. In beiden Fällen wurde die Zusammenfassungsfrage anschließend auf Filter-Missing gesetzt. (2) wurde der in der Zusammenfassungsfrage genannte Betrag als Filter-Missing kodiert. Wurde (3) nur die Summe der zwei ausstehenden Kredite genannt, dann bildete diese sowohl die Obergrenze des ersten als auch des zweiten noch ausstehenden Kredits für das Imputationsmodell. Wurde (4) weder die Höhe des ersten noch des zweiten ausstehenden Kredits noch deren Summe angegeben, wurde die Zusammenfassungsfrage als Filter-Missing editiert.

Bei drei Krediten und einem Überspringen der Fragen einer Schleife vor Nennung der einzelnen noch ausstehenden Kreditbeträge wurde bei den Editierungen ähnlich vorgegangen wie bei zwei Krediten, bei denen Fragen in den Schleifen übersprungen worden waren. Alle vorgenommenen Editierungen wurden wieder mit entsprechenden Flags versehen.

¹⁰ Bei abweichenden Angaben wurde dem innerhalb der ersten Kreditschleife erfassten Wert mehr Bedeutung zugewiesen als jenem in der Zusammenfassungsfrage. Dieses Vorgehen wird damit begründet, dass innerhalb der ersten Kreditschleife explizit nach dem ausstehenden Betrag des unbesicherten Kredits gefragt wurde und der Angabe daher mehr vertraut wird.

¹¹ Lediglich zwei Haushalte wählten diesen Weg.

Zusammenfassungsfragen

Am Ende jeder Schleife wurden Zusammenfassungsfragen gestellt (siehe Grafik 2 in Kapitel 2). Die Variablen zu diesen Fragen enthalten im Datensatz ausschließlich Informationen zu den über drei hinausgehenden Items eines Haushalts. Wie aus Grafik 2 ersichtlich ist, kamen im Zuge des Interviews auch all jene Befragten zu diesen Fragen, die die Antwort auf die Frage nach der Anzahl der Items verweigert hatten. Lag eine Verweigerung vor, wurde die Information bei der multiplen Imputation (Kapitel 5) verwendet und nachträglich aus dem Datensatz entfernt.

4.6.2.5 Personenvariablen auf Haushaltsebene

Verschiedene Variablen erfassen Informationen zu den Personen des Haushalts, sind aber im Haushalts-File gespeichert. Darunter fallen Informationen über die Verteilung von Immobilieneigentum innerhalb eines Haushalts, darüber, wer Kreditnehmer der einzelnen Kredite ist, oder über jenes Haushaltsmitglied, das im Unternehmen, das sich im Eigentum des Haushalts befindet, arbeitet.

Es wurden Variablen für bis zu 18 Personen erstellt, um auch einen ungewöhnlich großen Haushalt erfassen zu können. Die maximale Haushaltsgröße bei den erfolgreich durchgeführten Interviews in Österreich erreichte jedoch lediglich 9 Mitglieder, weshalb alle Variablen für weitere Personen aus dem Datensatz gelöscht wurden. Jede dieser Variablen wurde einzeln untersucht und gelöscht, wenn sie keine Information enthielt. War bei den Kreditnehmern etwa Person 6 der letzte Kreditnehmer, wurden die Variablen für Person 8 und alle weiteren gelöscht. Die Variable für Person 7 hatte in diesem Fall keinen echten Eintrag im dem Sinn, dass alle Haushalte auf „Keine weitere Person genannt“ kodiert sind, und verbleibt lediglich im Datensatz, um darzustellen, dass keine weitere Person genannt wurde.

4.6.2.6 Guthaben bzw. Überziehung auf den Girokonten

Es gab einige Haushalte, die den negativen Saldo auf den Girokonten des Haushalts fälschlicherweise als Guthaben auf den Girokonten (HD1110) auswiesen. Es kam in diesem Bereich mitunter sowohl zu Doppelnennungen als auch zu Erfassungen an der falschen Stelle, die in der Folge editiert werden mussten.

4.6.2.7 Variablen zu den Mieten

Im HFCS-Fragebogen wurde die Höhe der Miete sowohl exklusive als auch inklusive Betriebskosten abgefragt. Bei manchen Haushalten war die Miete ohne Einrechnung der Betriebskosten größer als oder gleich hoch wie jene mit Betriebskosten. Dies war logisch unmöglich, da Betriebskosten nicht kostenfrei zur Verfügung stehen. Ein Teil dieser Haushalte hatte in der Position „Miete inklusive Betriebskosten“ lediglich die Betriebskosten angegeben. Diese wurden zur Position „Miete exklusive Betriebskosten“ addiert und so als „Miete inklusive Betriebskosten“ editiert. Bei anderen Haushalten wurde die „Miete inklusive Betriebskosten“ auf Missing editiert und zur Imputation freigegeben, wobei die „Miete exklusive Betriebskosten“ als untere Schranke der zu imputierenden Werte der „Miete inklusive Betriebskosten“ diente.

Des Weiteren wurde die Position „Miete inklusive Betriebskosten“ als obere Schranke für die Variable „Miete exklusive Betriebskosten“ gesetzt und bei den Imputationen verwendet, wenn bei „Miete exklusive Betriebskosten“ ein fehlender Wert (also „Weiß nicht“ oder „Keine Angabe“) erfasst wurde.

4.6.2.8 Landwirtschaften

Gemäß der Definition des HFCS sind Landwirte Eigentümer eines Unternehmens in Form ihrer Landwirtschaft. Einige Landwirte sahen sich jedoch nicht als Unternehmer. Diese Fälle mussten deshalb gesondert untersucht und die Beobachtungen editiert werden. Die Basis der Editierung stellte die Definition der Klassifizierung eines Haushalts, der eine Landwirtschaft im Eigentum hielt, dar. Diese Klassifizierung erfolgte mithilfe der Beschäftigungsvariablen aller zum Haushalt gehörenden Personen. Gab zumindest eine Person im Haushalt an, (selbstständig) als Landwirt tätig zu sein, so wurde die Anzahl der Unternehmensbeteiligungen nur dann um eins erhöht, wenn noch keine Unternehmensbeteiligung für die Landwirtschaft erfasst worden war. Wurde die Unternehmensbeteiligung bereits zuvor angegeben, wurde keine Editierung vorgenommen. Der NACE-Code für dieses Unternehmen wurde auf Landwirtschaft gesetzt und zumindest jene Person, die angab, als Landwirt zu arbeiten, als in dieser Landwirtschaft beschäftigt betrachtet. Die Rechtsform wurde auf „Einzelunternehmen“ editiert. Sowohl der Eigentumsanteil des Haushalts als auch der Wert der Landwirtschaft wurden zur Imputation freigegeben.

In manchen Fällen wurde der Wert der Landwirtschaft fälschlicherweise beim Wert des Hauptwohnsitzes berücksichtigt. Diese Information wurde bei den Imputationen über die Grenzen bzw. durch die Verwendung einer Hilfsvariable, in der der Wert des Hauptwohnsitzes zusammen mit dem Wert der Unternehmensbeteiligung erfasst war, berücksichtigt (siehe Abschnitt 5.3.5).

Die Gruppe der Landwirte wurde gesondert expertenbasierten Einzelfallbetrachtungen unterzogen. Besonders komplexe Fälle wurden nachrecherchiert und gegebenenfalls logische Änderungen durchgeführt.

4.6.2.9 Lebensversicherungen

Zur Erfassung des Vermögens in Lebensversicherungen wurde eine Abfrage gewählt, die eine möglichst präzise und wenig fehleranfällige Antwort gewährleistete. So wurde nicht direkt nach der Höhe des Vermögens in Lebensversicherungen gefragt, sondern nach Einzahlungsbeginn, Laufzeit, Einzahlungsfrequenz (monatlich oder jährlich) und Höhe der laufenden Einzahlungen jeder einzelnen Lebensversicherung im Eigentum eines Haushalts.¹² Die Höhe des Vermögens, das in Lebensversicherungen gehalten wird, wurde als Summe der Einzahlungen errechnet. Wurde eine oder mehrere Informationen nicht angegeben, so wurden die übrigen Beobachtungen zur Eingrenzung des zu imputierenden Werts verwendet.

4.6.2.10 Einkommensvariablen

Das Einkommen wurde für jede Person ab 16 Jahre gesondert in folgenden Kategorien abgefragt:

- Einkommen aus unselbstständiger Arbeit (PG0110)
- Einkommen aus selbstständiger Arbeit (PG0210)
- Einkommen aus staatlicher Pension (PG0310)
- Einkommen aus privater (und betrieblicher) Pensionsvorsorge (PG0410)
- Einkommen aus Arbeitslosenunterstützung (PG0510)

¹² Die Möglichkeit einer Einmalzahlung zu Beginn der Lebensversicherung wurde über eine Verbatim-Erfassung ermöglicht.

Hinzu kamen folgende Einkommensarten, die auf Haushaltsebene abgefragt wurden:

- Einkommen aus öffentlichen Sozialtransfers (HG0110)
- Einkommen aus privaten Transfers (HG0210)
- Einkommen aus Immobilienbesitz (HG0310)
- Einkommen aus Finanzanlagen (HG0410)
- Einkommen aus Unternehmensbeteiligungen oder Personengesellschaften (HG0510)
- Einkommen aus sonstigen Quellen (HG0610)

Für die ersten vier personenbezogenen Einkommenskategorien konnten Befragte, wenn ihnen ihr jährliches Bruttoeinkommen nicht bekannt war, das Nettoeinkommen angeben (siehe Kapitel 2).

Wurde in einer Einkommenskategorie nur ein Nettobetrag genannt, wurde der Brutto-Netto-Rechner des Bundesministeriums für Finanzen in Österreich¹³ zur Umrechnung auf Bruttowerte verwendet. Dafür wurden Informationen zu Einkommensart, Haushaltsstruktur (hinsichtlich des Alleinerzieher- bzw. Alleinverdienerabsetzbetrags), Beschäftigungsstatus und Alter möglicher Kinder, Bundesland und Beschäftigungsstatus (Angestellte, Arbeiter, Pensionisten)¹⁴ verwendet.

Waren beide Eltern erwerbstätig, wurde der Alleinverdienerabsetzbetrag dem Hauptverdiener, das heißt jenem Elternteil mit dem höheren Einkommen, zugerechnet (sofern die gesetzlichen Vorgaben erfüllt waren und der Partner nicht mehr als 6.000 EUR jährlich verdiente).

Für die Umrechnung des Einkommens aus selbstständiger Beschäftigung wurde bei Selbstständigen aufgrund des erheblich größeren Spielraums bei der

Tabelle 3

Anzahl und Anteil der Editierungen des Bruttoeinkommens aus abhängiger Beschäftigung auf Basis der Flags

	Anzahl der Personen	Anteil in %
Anzahl der Personen, die Einkommen aus unselbstständiger Beschäftigung bezogen haben	2.166	100
Wert erhoben, vollständige Beobachtung (Flag 1)	1.263	58,3
Nicht imputiert, erhoben: „Weiß nicht“ (Flag 1050)	49	2,3
Nicht imputiert, erhoben: „Keine Angabe“ (Flag 1051)	76	3,5
Nicht imputiert, nicht erhoben aufgrund von Missing in der Filtervariable (Flag 1052)	14	0,7
Nicht imputiert, erhoben als Intervall (Flag 1053)	248	11,5
Nicht imputiert, erhobener Wert gelöscht oder nicht erhoben aufgrund eines CAPI-Fehlers (Flag 1054)	5	0,2
Nicht imputiert, Wert Missing gesetzt aufgrund einer falschen Antwort in der Filtervariable (Flag 1055)	207	9,6
Editiert, erhobener Wert nicht korrekt (Flag 3050)	28	1,3
Editiert, Information aus anderen (national erhobenen Variablen) verwendet (Flag 3051)	274	12,7
Editiert, Information aus Verbatim-Erfassung verwendet (Flag 3052)	2	0,1

Quelle: HFCS Austria 2010, OeNB.

¹³ Siehe http://www.bmf.gv.at/service/anwend/steuerberech/bruttonetto/_start.htm (abgerufen am 22. Jänner 2013).

¹⁴ Alle Lehrlinge wurden als Arbeiter eingestuft, Beamte hingegen aufgrund der günstigeren Stellung bei der Besteuerung als Angestellte bei der Berechnung gesehen.

steuerlichen Abzugsfähigkeit auf eine exakte Umrechnung verzichtet. Lediglich für Jahreseinkommen unter 11.000 EUR, die als steuerfrei eingestuft werden und somit der Bruttobetrag dem Nettobetrag entspricht, ist eine exakte Übertragung von brutto auf netto gegeben. Für alle anderen Werte (bei etwa 25 Personen) wurde die Umrechnung (auf Basis des Erwerbsstatus für Angestellte) durchgeführt und zu bzw. von diesem Wert 10.000 EUR hinzugerechnet bzw. abgezogen, um ein Intervall für die Imputation des exakten Werts zu erstellen. Damit wurde die Unsicherheit, die mit dieser Umrechnung einherging, abgebildet. Gleichzeitig ging die wertvolle Information über den tatsächlichen Bereich, in dem dieser Wert liegt, nicht verloren.

Wurde der Nettobetrag auch nur im Intervall angegeben, so wurden die Intervallgrenzen in Bruttobeträge umgerechnet und in der Folge bei den Imputationen verwendet. Alle umgerechneten Werte wurden mit dem Flag-Code 3051 versehen.

Private Pensionen (PG0410) wurden nicht umgerechnet, sondern auf Bruttowerte übertragen, da es sich um geringe Betragshöhen handelte.¹⁵

Tabelle 3 gibt beispielhaft auf Basis der Flags Aufschluss über die Anzahl der vorgenommenen Editierungsmaßnahmen beim unselbstständigen Einkommen. Auch lässt sich anhand dieser Tabelle eine Verwendung der Flag-Variablen (siehe auch Abschnitt 4.5) exemplarisch darstellen. Zur Frage nach der Höhe des Einkommens aus unselbstständiger Beschäftigung (Variable PG0110) gelangten insgesamt 2.166 Personen. 58,3 %, also 1.263 Personen, gaben das gewünschte jährliche Bruttoeinkommen an. Des Weiteren antworteten 49 (2,3 %) mit „Weiß nicht“ und 76 (3,5 %) mit „Keine Angabe“. Weitere 14 Personen beantworteten die Ja/Nein-Frage, ob sie ein Einkommen als Arbeitnehmer beziehen, mit „Weiß nicht“ oder „Keine Angabe“ und erreichten aufgrund der Filterführung deshalb nicht die Frage nach der Höhe des Einkommens (Flag-Code 1052). Rund 11,5 % (248 Personen) gaben das Einkommen in Form eines Intervalls an. Angaben zu insgesamt 212 Personen (9,8 %) wurden durch Editierungsmaßnahmen auf Missing und „Zu imputieren“ gesetzt, wobei der überwiegende Anteil (207 Personen) aufgrund einer falschen Head-Variable (Flag-Code 1055) editiert werden musste. Von 12,7 % der Personen (274) konnte das Einkommen lediglich netto angegeben und mit dem Brutto-Netto-Rechner des BMF umgerechnet werden. Die Angaben der verbleibenden 30 Personen (rund 1,4 %) wurden auf Basis der Experteneinschätzungen im Wert editiert. Dabei handelte es sich zumeist um Nachrecherchen bzw. das oben beschriebene Nullen-Problem.

4.6.2.11 ISCO-Klassifizierung

Gemäß den Vorgaben des Fragebogens der EZB wurde die hauptberufliche Tätigkeit der Befragten in der Berufssystematik der *International Standard Classification of Occupations* (ISCO88), einem Klassifikationsschema für Gruppen von Berufen, erhoben (erfasst in der Variable PE0300). Die Selbstzuordnung der einzelnen Haushaltsmitglieder hätte jedoch die Befragten, denen die ISCO-

¹⁵ Insgesamt sieben Personen gaben ihr Einkommen aus privater Pension als Nettobetrag an. Bei sechs Personen lag das Jahreseinkommen unter EUR 2.900, eine Person gab Nettoeinkommen aus privater Pension in Höhe von EUR 18.500 an. Alle diese Nettowerte wurden aufgrund der geringen Betragshöhe als Bruttowerte verwendet und somit nicht umgerechnet.

Klassifizierung nicht vorab bekannt war, vor große Schwierigkeiten gestellt und hätte so zu mangelhaften Beobachtungen geführt. Aus diesem Grund wurden die Berufsbezeichnung bzw. die für den Hauptberuf relevanten Tätigkeiten in Österreich in verbaler Form abgefragt und in einer Verbatim-Variable erfasst. Diesen Angaben wurden die entsprechenden ISCO-Codes aus dem „Alphabetikum zur OEISCO“ der Statistik Austria ex post zugeordnet. In Übereinstimmung mit den EZB-Vorgaben erfolgte die ISCO-Klassifizierung auf dem Aggregationsniveau der Berufsgruppen (Zweisteller). Dafür wurden, zusätzlich zur Verbatim-Abfrage der hauptberuflichen Tätigkeit, Personeninformationen, die für die ISCO-Klassifikation relevant sind, herangezogen (vor allem der Bildungsgrad des Befragten und die Hauptaktivität des Unternehmens, bei dem der Befragte jeweils beschäftigt war).

Um internationale Vergleiche aller Euroraumländer zu ermöglichen, wurden die von der EZB geforderten ISCO-Codes nach alter (ISCO88) Klassifikation erstellt. In den Flags wurde die an die EZB abzuliefernde Variable PE0300 vorerst mit dem Flag-Code 3051 versehen und in einem weiteren Schritt aggregiert (siehe Abschnitt 4.7). Dieser Flag-Code steht für „Editiert, Information aus anderen (national erhobenen) Variablen verwendet“.

4.6.2.12 Ausschluss erfolgreich durchgeführter Interviews

Aus unterschiedlichen Gründen wurden folgende drei Gruppen von insgesamt 28 Haushalten, die ein erfolgreiches Interview absolviert hatten, nicht in den endgültigen Daten verwendet:

- Haushalte außerhalb der Grundpopulation: Die Grundpopulation im HFCS in Österreich umfasste alle Haushalte, die nicht in Institutionen (wie etwa Kinder- oder Seniorenheimen, Gefängnissen) leben. Einige erfolgreich interviewte Haushalte wurden aufgrund der Tatsache, dass sie in einem Seniorenheim wohnten, von der Erhebung ausgeschlossen. Die vier zu dieser Gruppe zählenden Haushalte wurden als „nicht zur Bruttostichprobe gehörend“ editiert. Zwei Haushalte mit nicht erfolgreich absolvierten Interviews wurden ebenfalls mit dieser Begründung aus der Grundgesamtheit entfernt.
- Haushalte mit zu hohem Anteil an Item-Non-Response: In dieser Kategorie mussten 22 Haushalte aus dem Datensatz entfernt werden, da sie zu oft Antworten verweigerten.
- CAPI-Fehler: Bei zwei Haushalten kamen aufgrund eines Programmierfehlers im digitalen Fragebogen nicht alle Haushaltsmitglieder zu den Personenfragen. Die Haushalte der letzten beiden Gruppen wurden auf „Nichtteilnahme aus anderen Gründen“ editiert und bekamen ein Non-Response-Gewicht von null (siehe Abschnitt 7.2.3).

4.6.2.13 CAPI-Fehler: Probleme im Fragebogen

Kategorie „Sonstige“ bei der Frage nach dem Geburtsland (RA0400)

In dieser Kategorie wurde keine Verbatim-Erfassung vorgenommen. Das heißt, alle „Sonstigen Länder“ konnten ex post nicht zugeordnet werden und mussten als „Sonstige“ im Datensatz gespeichert bleiben. Insgesamt 11 Personen wählten diese Kategorie und blieben so im Datensatz enthalten.

Bildung der Eltern (APA02\$0¹⁶)

Diese Information sollte bei Kindern von Kompetenzträgern nicht erhoben werden, da die Information schon über die Bildung des Kompetenzträgers bekannt war. Dieser Filter funktionierte aber nicht zu 100 %: Vier Personen – Eltern eines Kompetenzträgers – kamen gar nicht zu dieser Variablen. Die Variable dieser Personen wurde auf Missing editiert und zur Imputation freigegeben. Weiteren 18 Personen, die Kinder eines Kompetenzträgers waren, wurde die Frage nach der Bildung der Eltern irrtümlicherweise gestellt und deshalb ex post aus den Daten gelöscht.

Nach oben offenes Intervall der Unternehmensbeteiligung (HD0801)

Am Anfang der Feldphase kam es in der Filterführung bei der Euro-Schleife für den Wert der ersten Unternehmensbeteiligung zu Mängeln. Bei der Angabe eines nach oben offenen Intervalls gelangte man nicht zur Währungsangabe und Bestätigungsfrage, sondern zu den fixen Intervallen. Im einzigen Fall, bei dem dies geschah, erfolgte die Angabe offensichtlich in Euro und die Angabe wurde auch so erfasst.

Beschäftigungsstatus bei der letzten Hauptbeschäftigung bzw. Dauer der Beschäftigung insgesamt (PE0900 bzw. PE1000)

Aufgrund falscher Filterführung im CAPI fehlten diese Variablen von manchen Personen, von anderen wurden sie teils unnötigerweise erfasst. So wurden im Editierungsprozess die Angaben in der Variable PE0900 von den irrtümlich zusätzlich erfassten Personen (das heißt von allen, die ein aufrechtes Arbeitsverhältnis vorwiesen) entfernt, um eine Übereinstimmung mit den internationalen Vorgaben zu erreichen. Diese Informationen können jedoch auf einfachem Weg über die Daten in der im Fragebogenverlauf folgenden Variable (PE1000) wiederhergestellt werden. Zusätzlich wurden die Beobachtungen von 14 Personen, die irrtümlicherweise nicht zur Frage nach der Dauer der Beschäftigung insgesamt gelangten, zur Imputation freigegeben.

Voraussichtliches Pensionseintrittsalter (PE1100)

Die Filterführung im CAPI des HFCS in Österreich war mit jener der Vorgabe der EZB nicht kongruent in der Hinsicht, dass sich der Filter der Frage nach dem Pensionseintrittsalter nicht nur (gemäß EZB-Vorgabe) auf den Hauptbeschäftigungsstatus (PE0100a), sondern auch (entsprechend der HFCS-Durchführung in Österreich) auf einen anderweitigen Beschäftigungsstatus bezog. Aus diesem Grund wurde diese Frage an neun Personen nicht gestellt. Die Beobachtungen zu den Personen in dieser Variable wurden auf Missing gesetzt und zur Imputation freigegeben, um eine Übereinstimmung mit der internationalen Vorgabe zu erreichen.

Regelmäßige Auszahlung der betrieblichen Altersvorsorge (PF0800)

Die Filterführung zur Frage nach den regelmäßigen Auszahlungen der betrieblichen Altersvorsorge wurde in Österreich nicht allen Personen, die einen Anspruch

¹⁶ Dies ist eine österreich-spezifische Non-Core Variable, die nicht im international erhältlichen HFCS-Datensatz vorhanden ist, sondern lediglich für Österreich vorliegt.

auf zukünftige betriebliche Pensionszahlungen haben, gestellt. Die Informationen zu den Personen, denen diese Frage fälschlicherweise nicht gestellt wurde, wurden imputiert.

4.7 Formatierung und Editierung nach den multiplen Imputationen

Alle in Österreich in feineren Kategorien abgefragten Informationen werden nach den Imputationen auf die international vorgegebene Aggregationsstufe aufbereitet. Wesentliche Aggregationen lassen sich wie folgt zusammenfassen:

- Familienstand: Die Kategorien „Verheiratet und mit Ehepartner zusammenlebend“ und „Verheiratet und getrennt lebend“ wurden zu „Married“ zusammengefasst
- Bildung: Die österreich-spezifischen Kategorien wurden den ISCED-Kategorien zugeordnet. Da die nationalen Bildungswege in manchen Fällen nicht leicht der ISCED-Kodierung zuzuordnen sind, wird die Verwendung der für Österreich erhobenen Kategorisierung empfohlen
- Beschäftigungsstatus/-verhältnis: Aggregation der feineren Kategorien
- Hauptwohnsitz – Eigentumsverhältnis: Aggregation der feineren Kategorien
- Kreditraten: Die Raten für die Rückzahlung von endfälligen Krediten (sowohl besicherter als auch unbesicherter) wurden auf „0“ gesetzt, da diese Kredite erst am Ende ihrer Laufzeit in einer Einmalzahlung getilgt werden. Die Höhe des Sparens für die Rückzahlung des gesamten Volumens kann in österreich-spezifischen Variablen analysiert werden.
- Anzahl der sonstigen Fahrzeuge: Fahrzeugtypen „Kleintransporter/Vans“ und „Wohnmobile/Wohnwägen“ wurden zu „Vans“ aggregiert
- Zweck eines Kredits: Die Kategorie „Zur Finanzierung des Genossenschaftsbeitrags“ wurde in die Erfassung der Kategorie „Other“ geschrieben
- Rechtsform des Unternehmens: Aggregation der feineren Kategorien
- Bausparguthaben und Lebensversicherung: Die Informationen zu diesen beiden Anlagemethoden wurden zum Sparguthaben (HD1200 und HD1210) aggregiert
- Anlageverhalten – Risikopräferenzen: Die Antwortmöglichkeit „Keine einheitliche Zuordnung möglich“ wurde mit „Weiß nicht“ kodiert
- Art der erhaltenen Vermögenswerte im Fragebogenabschnitt zu Erbschaften und Schenkungen: Auflösung der Reihung nach Vermögenswert
- Geber der Zuwendung im Fragebogenabschnitt zu Erbschaften und Schenkungen: Aggregation der feineren Kategorien
- Zweck des Sparens: Auflösung der Reihung nach Wichtigkeit der Sparzwecke
- Paradata: HR1100 und HR1200 wurden zu bzw. von einer Mehrfachantwort umkodiert; HR1400I wurde nicht erhoben und auf Filter-Missing gesetzt
- Flags: Die österreich-spezifischen feineren Flags werden zur internationalen Norm aggregiert; d. h.,
 - Flag 1055 wird zu 1054;
 - Flags 3051, 3052 und 2 werden zu 1;
 - Flag 3053 wird zu 0;
 - und Flag 4055 wird zu 4054.

Die zusätzlichen Informationen können voraussichtlich gesondert zu den HFCS-Datensätzen der EZB, die alle national erhobenen und von der EZB geforderten Variablen enthalten, von der OeNB ab Frühjahr 2013 bezogen werden. Darin sind insofern zusätzliche Informationen, als zusätzliche Variablen enthalten sind und

manche Variablen feiner kategorisiert sind. Ein Zusammenspielen der Datensätze auf Basis der Identifikationsnummern sowie der Imputationsnummer ist möglich.

4.8 Abschließende Bemerkungen und Online-Anhang

Grundsätzlich unterlagen alle Editierungsmaßnahmen der Maxime, sie lediglich auf jene Beobachtungen anzuwenden, die mit an Sicherheit grenzender Wahrscheinlichkeit nicht korrekt erfasst wurden. Bestand Unklarheit, so wurde zunächst die Möglichkeit einer telefonischen Nachrecherche in Erwägung gezogen. Durch diesen Schritt konnten viele Beobachtungen korrigiert oder als korrekt eingestuft werden.

Die Kenntnis der durchgeführten Maßnahmen zur Konsistenzprüfung der Daten ist sowohl für die Analyse der Daten als auch für das Verständnis der Entstehung von Beobachtungen unerlässlich. Des Weiteren wird dem Nutzer durch die Verwendung der Flags ermöglicht, ein individuelles Imputationsmodell zu erstellen, auf die Imputationen zu verzichten oder auf eine andere Art mit dem Problem des Antwortausfalls umzugehen.

Den Ausführungen zu den Editierungsmaßnahmen und Konsistenzprüfungen im HFCS in Österreich wird in Form eines Online-Anhangs eine Liste der in der digitalen Version des Fragebogens programmierten Konsistenzprüfungen beigelegt.¹⁷

¹⁷ Alle Dokumente des Online-Anhangs sind abrufbar unter www.hfcs.at.