

Paradaten im HFCS Austria 2010 – Teil 2: Evaluierung von Messfehlern

In dieser Studie wird das umfangreiche Paradaten-Set des HFCS Austria 2010 zur Evaluierung von Messfehlern herangezogen. Die Komplexität einer Erhebung zu Haushaltsfinanzen verlangt, dass auch Hilfsvariablen erhoben werden. Aus den im Zuge der ersten Welle des HFCS Austria 2010 gewonnenen Erfahrungen resultiert die Notwendigkeit einer Verwendung von disaggregierten Zeitstempeln und diversifizierteren Interviewerkommentaren. Für eine exakte Beantwortung von Betragsfragen durch die Respondenten ist zudem die Verwendung von Unterlagen (Kreditverträgen, Sparbüchern usw.) von großer Bedeutung.

Nicolás Albacete,
Martin Schürz

Paradaten sind Hilfsvariablen bei Haushaltserhebungen „*at all stages of the survey process and with very different granularities*“ (Kreuter, 2013). Sie dienen dem Monitoring während der Datenerhebung und der nachfolgenden Evaluierung. Mithilfe von Paradaten – etwa die Interviewdauer, die Anzahl der Kontaktversuche und die Zeitpunkte der Kontaktierung – können Verzerrungen in der Stichprobe ausgeglichen und nachfolgende Erhebungen verbessert werden.

Das Paradaten-Set der OeNB zum HFCS Austria 2010 dient der internen statistischen Analyse und ist der Forschungsgemeinschaft nicht zugänglich. Auch die anderen am HFCS teilnehmenden Zentralbanken publizieren ihre Paradaten zum HFCS nicht.

Im ersten Teil der Studie zu den Paradaten des HFCS Austria 2010 (Albacete und Schürz, 2014) wurde deren Bedeutung für die Evaluierung von Non-Response-Fehlern evaluiert. Die Informationen zu Haushalten, welche die Teilnahme verweigern, erlauben eine Korrektur des Non-Response-Bias. Eine wichtige Schlussfolgerung von Teil 1 war, dass dem Umstand der Verweigerung mehr Aufmerksamkeit in der Datenanalyse geschenkt werden muss. So wird etwa das Oversampling von Wiener Haushalten in der zweiten Welle des HFCS erhöht werden (siehe Albacete und Schürz, 2014). Eine Befragung der Interviewer erlaubte zudem eine extensive Analyse möglicher Kom-

munikationsverzerrungen in der Befragung (siehe Albacete und Schürz, 2013).

Im zweiten Teil der Studie zu den Paradaten des HFCS Austria 2010 steht die Evaluierung von Messfehlern im Zentrum der Betrachtung. Ein Messfehler besteht, wenn sich die Antwort eines Respondenten von einer wahren Angabe unterscheidet. Abweichungen zwischen Antworten und Wahrheit ergeben sich etwa bei kognitiven Fehlleistungen. Der kognitive Prozess beim Antworten verläuft mehrstufig: Am Anfang steht die Wahrnehmung der Frage, danach folgt das Verständnis der Worte und des Sinngehalts. Danach setzt die Erinnerung ein und das relevante Material wird abgerufen. Die Erinnerungen werden dann beurteilt und in einen Antwortrahmen gebracht. Diese Vielzahl von Bausteinen indiziert mögliche Fehlerquellen bei der Beantwortung. Treten diese Fehler systematisch in derselben Richtung über alle Respondenten auf, resultiert daraus ein Messfehler-Bias. Entstehen diese Fehler in variierendem Ausmaß und in unterschiedlichen Richtungen über die Respondenten, ergibt sich eine Messfehler-varianz.

Wie kann man Messfehler bei Haushaltserhebungen messen?

Bound et al. (2001) geben einen detaillierten Überblick zu Messfehlern in Erhebungen, einschließlich jener bei Angaben zum Vermögen. Ein Messfehler

besteht, wenn sich der wahre Wert von dem gemessenen unterscheidet. Das exakte Ausmaß des Messfehlers einer Variable kann nur berechnet werden, wenn der wahre Wert bekannt ist. Dies ist aber nur selten der Fall. Oft werden daher als Annäherung an den wahren Wert Administrativdaten herangezogen. Diese gibt es in Österreich z. B. auch für den Mikrozensus, der seit 2009 Einkommensdaten aus Verwaltungsdaten hinzugefügt hat.¹ Allerdings sind den Autoren keine Studien bekannt, die diese Informationen im österreichischen Mikrozensus verwendet haben, um bei dieser Variable Messfehler zu schätzen.

In den USA vergleichen Brickler und Engelhardt (2008) Einkommensdaten von Männern und Frauen aus dem Health and Retirement Study (HRS) mit Einkommensdaten derselben Männer und Frauen aus Sozialversicherungsdaten. Sie schätzen einen Messfehler im HRS von 6% beim Einkommen der Männer und von ca. 7% beim Einkommen der Frauen.

Es können aber auch Administrativdaten selbst Messfehler enthalten, wie Kapteyn und Ypma (2007) untersuchen. Bei einem Vergleich des Swedish Longitudinal Individual Data Base (LINDA) mit dem Survey of Health, Ageing, and Retirement in Europe (SHARE) finden die Autoren nicht nur Messfehler bei den SHARE-Erhebungsdaten, sondern auch bei den LINDA-Administrativdaten. Anhand verschiedener Beispiele zeigen sie, dass auch die Administrativdaten verzerrt sind und also nicht angenommen werden darf, dass solche Daten die Wahrheit vollkommen erfassen.

Für den Household Finance and Consumption Survey (HFCS) ist aufgrund fehlender wahrer Referenzwerte

keine exakte Berechnung von Messfehlern zu den Haushaltsangaben möglich. Dies ist weder bei den Vermögenskomponenten noch bei den meisten sonstigen abgefragten Variablen machbar. Es gibt jedoch Substitute, die auf potenzielle Messfehler verweisen, weil sie mit diesen üblicherweise korrelieren. Dies sind z. B. Informationen über Probleme beim Beantworten von Fragen, über die Geschwindigkeit beim Antworten, über Änderungen von Antworten seitens des Respondenten, über das Interesse des Respondenten an der Teilnahme, über die Verwendung von Unterlagen für das Interview, über das Misstrauen des Respondenten gegenüber dem Interviewer oder der HFCS-Befragung, über die Interviewerkommentare oder über die Interaktion Interviewer–Respondent. Letzteres wurde bereits in einem früheren Artikel analysiert (siehe Albacete und Schürz, 2013 und 2014).

Wie können Messfehler im HFCS Austria anhand von Paradaten evaluiert werden?²

Wenigstens zweierlei kann zu Meßfehlern führen: zum einen die Komplexität der Erhebung (Fragen, Fragedauer) und zum anderen die emotionale Verfasstheit und die kognitiven Fähigkeiten der Befragten. So kann sich das Alter, die Bildung, aber auch die Stimmung der Respondenten in den Antwortzeiten niederschlagen. Möglicherweise wird schnell, aber unaufmerksam geantwortet oder langsam und überfordert. Im Folgenden werden Schwierigkeiten beim Antworten auf die HFCS-Fragen, unterschiedliche Antwortgeschwindigkeiten und divergierende Motivationslagen der Respondenten beschrieben.

¹ www.statistik.at/web_de/services/mikrodaten_fuer_forschung_und_lehre/zusaetzliche_merkmale_mz/index.html

² Für eine ausführliche Literaturübersicht und Beschreibung der Verwendung von Paradaten im Hinblick auf die Evaluierung von Messfehlern siehe Olson und Parkhurst (2013) und Yan und Olson (2013).

Zudem werden auch Wahrnehmungen der Interviewer referiert.

Item-Non-Response

Ausdrücke des Zweifels oder der Unsicherheit des Respondenten (z. B. „weiß nicht“, „ich glaube“, „vielleicht“) können auf Ungenauigkeiten in seiner Antwort hindeuten. Obwohl solche Ausdrücke per se keine Messfehler oder Ursache von Ungenauigkeiten sind, werden sie in der Literatur oft als Indikatoren für Probleme bei der Beantwortung von Fragen und für potenzielle Messfehler verwendet. Im HFCS Austria 2010 kann die Nicht-Beantwortung einzelner Fragen mittels „weiß nicht“ oder „keine Angabe“ (Item-Non-Response) erfolgen. Non-Response-Fehler durch die Nicht-Teilnahme am Interview (Unit-Non-Response) wurden bereits im ersten Teil der Studie zu den Paradaten des HFCS-Austria 2010 evaluiert.

Die Item-Non-Response hängt vom Inhalt der Frage ab (siehe Tabelle 1). Respondenten beantworten problemlos Fragen nach ihrem Beschäftigungsstatus, ihren Ausgaben für Lebensmittel zu Hause oder nach ihrer monatlichen Miete. Werden sie jedoch nach der Höhe ihres Einkommens aus Finanzanlagen gefragt, so kann nur etwa ein Drittel der Befragten eine Betragsangabe machen. Sehr hilfreich für die Respondenten ist, wenn es alternativ möglich ist, ein Intervall anzugeben. Die Intervallsangaben werden auch für Imputationen genutzt. Für die Befragungen im Rahmen der zweiten Welle des HFCS Austria wurden daher die Möglichkeiten für Intervallsangaben weiter optimiert, sodass der Interviewer eine bestimmte Intervallsliste vorlegen kann, die für die jeweilige Frage adaptiert wurde.

Tabelle 1

Item-Non-Response bei ausgewählten Variablen (ungewichtet)

	Haushalt verfügt über das Item		Angaben jener Haushalte, die über das Item verfügen			
	Ja	Unbekannt	Betrag	Intervall	„Weiß nicht“/ „Keine Angabe“	Sonstiges Missing ¹
	(1)	(2)	(3)	(4)	(5)	(6)
	in %					
Wert des Hauptwohnsitzes ²	49,6	0,0	75,5	15,3	6,4	2,7
Durch Hauptwohnsitz besicherte Hypothek 1: ausstehender Kapitalbetrag	15,1	1,4	63,5	21,2	12,3	3,1
Monatliche Miete	44,1	0,0	97,1	2,3	0,5	0,1
Sonstiges Immobilieneigentum 1: Marktwert	12,6	0,2	74,1	15,3	9,6	1,0
Durch sonstige Immobilien besicherte Hypothek 1: ausstehender Kapitalbetrag	1,7	0,4	70,7	7,3	14,6	7,3
Guthaben auf Girokonten	98,9	0,0	72,0	13,3	14,4	0,3
Guthaben auf Sparkonten	86,0	1,6	64,6	18,6	16,0	0,8
Wert börsennotierter Aktien	5,4	0,4	71,1	12,5	16,4	0,0
Geldschulden gegenüber dem Haushalt	9,3	0,5	90,5	5,0	4,5	0,0
Beschäftigungsstatus (Hauptbeschäftigung) (Person 1)	100,0	0,0	99,9	0,0	0,1	0,0
Bruttoeinkommen aus abhängiger Beschäftigung (Person 1)	48,8	0,1	76,7	9,9	3,4	9,9
Bruttoeinkommen aus der Arbeitslosenunterstützung (Person 1)	6,1	0,1	83,3	9,7	6,3	0,7
Bruttoeinkommen aus Finanzanlagen	70,9	6,6	34,3	40,7	24,0	0,9
Schenkung/Erbschaft 1: Wert	21,4	1,3	71,1	16,3	10,0	2,6
Ausgaben für Lebensmittel zu Hause	100,0	0,0	96,3	3,4	0,3	0,0

Quelle: Albacete et al. (2012).

¹ Missing Values aufgrund von Editierungsmaßnahmen und dem Ausstieg aus Schleifen.

² Hierfür wurde die Variable HB0900 verwendet.

Langsame Antworten

Zeitstempel geben indirekt Aufschluss darüber, ob manche Fragenabschnitte oder einzelne Fragen schwieriger zu beantworten waren als andere. Diese Information hilft bei der Identifikation von Problemgruppen unter den Respondenten oder bei der Formulierung von Fragen in der Erhebung. Manchmal werden von Experten formulierte Fragen als zu kompliziert empfunden und treffen bei bestimmten Respondenten auf Unverständnis. Zeitverzögerungen bei der Beantwortung weisen möglicherweise auf die Notwendigkeit hin, Fragen umzuformulieren.

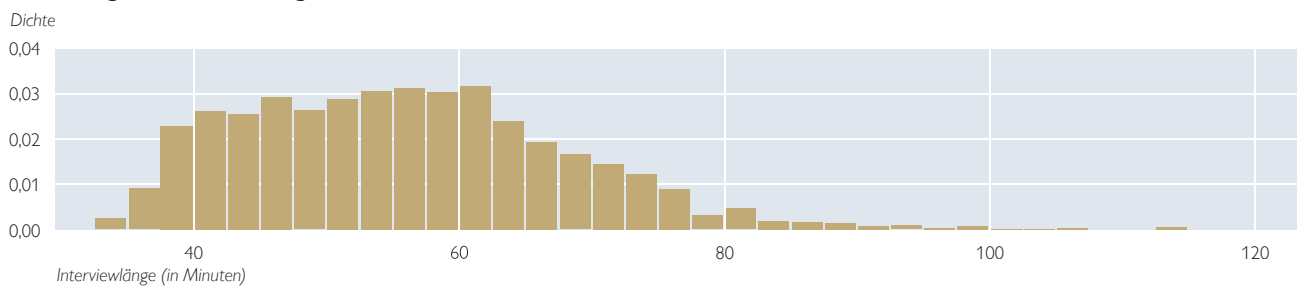
Wurden in der ersten Welle des HFCS Austria 2010 vom durchführenden Meinungsforschungsinstitut keine Zeitstempel zur Verfügung gestellt, so sind solche für die nächste Welle des HFCS Austria 2014 jedenfalls vorgesehen. Informationen zur gesamten Interviewlänge finden sich aber bereits in den Paradaten der ersten Welle.

Grafik 1 (oberes Panel) zeigt die Verteilung der Interviewlänge über alle Haushalte, die zwischen 33 und 114 Minuten variiert und im Mittelwert 56 Minuten beträgt. In Verbindung mit der Information über die Anzahl der gestellten Fragen pro Respondenten

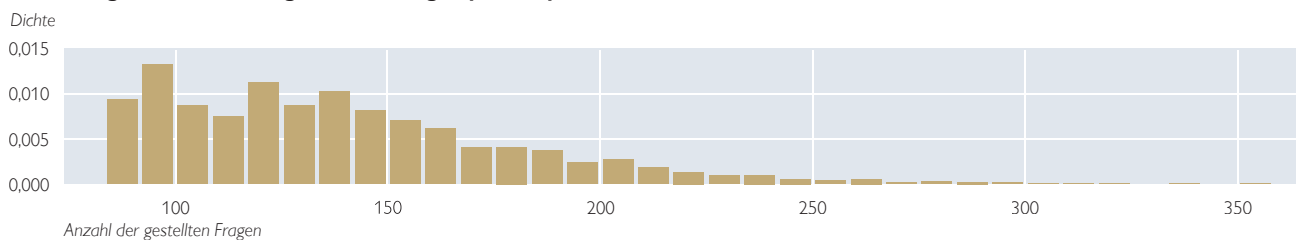
Grafik 1

Interviewlänge und Antwortdauer pro Frage im HFCS Austria 2010

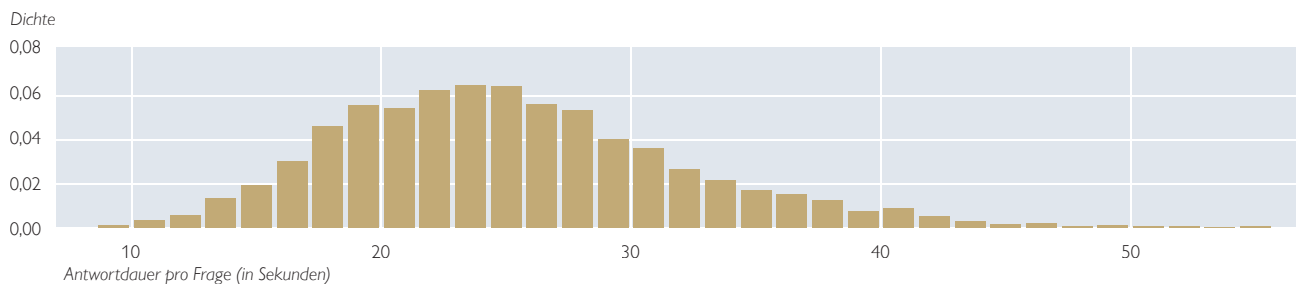
Verteilung der Interviewlänge



Verteilung der Anzahl der gestellten Fragen pro Respondenten



Verteilung der Antwortdauer pro Frage



Quelle: HFCS Austria 2010.

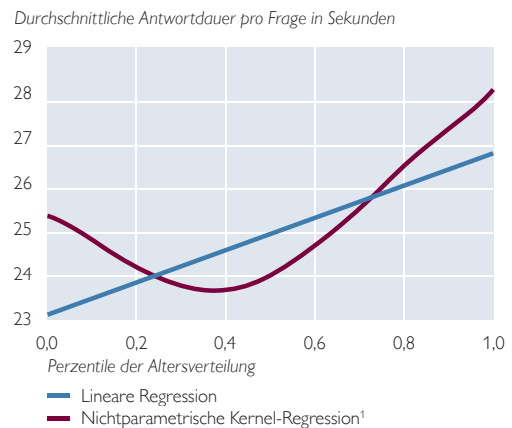
(siehe Grafik 1 Mitte) kann daher die durchschnittliche Antwortdauer pro Frage berechnet werden (siehe Grafik 1 unteres Panel). Auswertungen zeigen, dass die Interviewdauer pro Frage mit dem Anteil an Betragsfragen im Interview steigt. Fragen nach Beträgen zu beantworten ist zeitintensiver und schwieriger als die Beantwortung anderer Fragen, etwa jener nach Einstellungen.

Neben der Komplexität von Fragen können langsame Antworten auch auf die kognitive Kapazität der Respondenten zurückgeführt werden. Grafik 2 zeigt einen U-förmigen Verlauf beim Alter. Je älter der Respondent ist, desto höher ist die durchschnittliche Antwortdauer pro Frage. Doch auch bei jüngeren Respondenten ergeben sich teils zeitintensive Interviews, da jene aufgrund ihres Alters mitunter noch nicht über die Informationen verfügen, die für die Beantwortung bzw. für eine schnelle Beantwortung der Fragen notwendig wären.

Auch geringe Bildung kann die Ursache für langsame Antworten sein (siehe Tabelle 2). Würden die Probleme beim Antworten nach den Bildungsabschlüssen stark streuen, wäre dies ein Hinweis auf die Notwendigkeit, Fragen einfacher zu formulieren. Das uneinheitliche Muster und die nur geringen Unterschiede bei der Interviewlänge pro Frage zeigen aber, dass die Fragen des

Grafik 2

Korrelation zwischen Alter der Referenzperson und Beantwortungszeit



Quelle: HFCS Austria 2010, OeNB.

¹ Es wird für jeden Datenpunkt eine separate gewichtete Regression geschätzt, wo der jeweilige Datenpunkt das höchste Gewicht bekommt und die weiter entfernten Datenpunkte niedrigere Gewichte.

HFCS Austria 2010 vermutlich zu keiner von der Bildung des Respondenten abhängigen Verzerrung geführt haben.

Auch unter den Berufsgruppen gibt es Unterschiede in der Beantwortungszeit (Tabelle 3). So haben z. B. selbstständig beschäftigte Referenzpersonen der Haushalte im Durchschnitt um 5 Sekunden kürzere Antwortdauern pro Frage als Pensionisten, was je nach Anzahl der gestellten Fragen eine Verminderung der gesamten Interviewlänge um mindestens 7 bis maximal 30 Minuten bedeutet.

Tabelle 2

Beantwortungszeit und Bildung der Referenzperson

Höchster Bildungsabschluss	Durchschnittliche Antwortdauer pro Frage in Sekunden
Pflichtschule	26
Berufsschule	25
Berufsbildende mittlere Schule	25
Matura	25
Universität	24

Quelle: HFCS Austria 2010, OeNB.

Tabelle 3

Beantwortungszeit und Beruf der Referenzperson

Beschäftigungsstatus	Durchschnittliche Antwortdauer pro Frage in Sekunden
Unselbstständig beschäftigt	24
Selbstständig beschäftigt	22
Arbeitslos	26
In Pension	27
Sonstiges	23

Quelle: HFCS Austria 2010, OeNB.

Schließlich können langsame Antworten auf komplexe Fragen nicht nur auf die kognitive Kapazität der Respondenten, sondern auch auf das Interviewerverhalten zurückgeführt werden. So zeigen z. B. Olson und Peytchev (2007) oder Olson und Bilgen (2011), dass die Interviewer im Laufe der Feldarbeit mit zunehmender Anzahl von durchgeführten Interviews eine immer kürzere Interviewdauer erzielen. Dies liegt darin begründet, dass sie mit jedem Interview an Erfahrung gewinnen. Es bleibt aber unklar, ob dadurch die Interaktion zwischen Interviewer und Respondent vorteilhaft oder nachteilig beeinflusst wird. So könnte die Ursache für eine kürzere Interviewdauer ein schnelleres Vorlesen der Fragen oder eine Abweichung vom geplanten Interviewablauf sein. Solche geringen Abweichungen können vorteilhaft sein, wenn die Interviewer weniger oft unnötige Erklärungen geben oder irrelevante Gespräche führen. Nachteilig sind sie aber, wenn die Ursache der Verkürzung darin liegt, dass der Interviewer nicht auf den Befragten eingehen will, wenn dieser Feedback benötigen würde.

Wenig wünschenswert ist, wenn der Interviewer im Laufe der Feldarbeit bei seinen Interviews immer schneller wird, weil er verstärkt *Directive Probing* anwendet. Diese Interviewtechnik besteht darin, dass er suggestive Fragen stellt, die dem Befragten bestimmte Antworten nahe legen. (Auf die Frage: „Wie hoch ist der Zinssatz?“ könnte eine mögliche Antwort lauten: „Sehr niedrig“. Die anschließende Frage als *Directive Probe* formuliert würde etwa lauten: „1 %?“, wodurch die Antwort in eine bestimmte Richtung gesteuert würde.) Diese Interviewtechnik generiert zwar schnellere Antworten, schränkt aber die Bandbreite der Antworten ein.

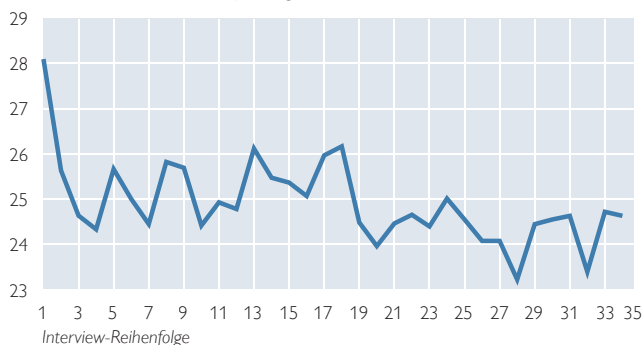
Grafik 3 (linkes Panel) zeigt, dass die Ergebnisse von Olson und Peytchev (2007) auch für das HFCS Austria 2010 zu gelten scheinen. Außerdem zeigt Grafik 3 (rechtes Panel), dass erfahrene Interviewer im Durchschnitt kürzere Interviewzeiten erzielen als ihre noch relativ unerfahrenen Kollegen und dieser Unterschied steigt auch mit der Anzahl der durchgeführten Interviews. Eine tiefer gehende Analyse wäre notwendig, um etwas über die Ursachen sagen zu können. Jedenfalls ist hervorzuheben,

Grafik 3

Zusammenhang zwischen Interview-Reihenfolge und Interviewlänge

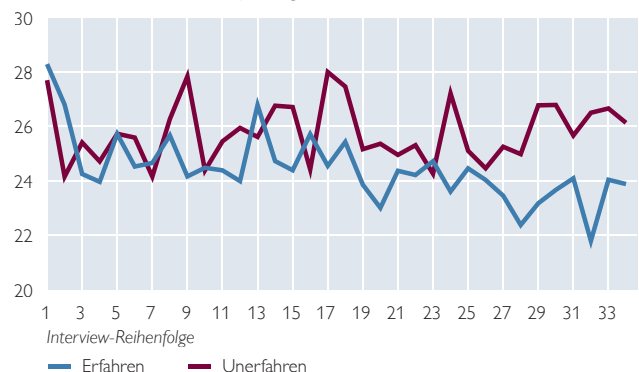
Insgesamt

Durchschnittliche Antwortdauer pro Frage in Sekunden



Nach Interviewer-Erfahrung

Durchschnittliche Antwortdauer pro Frage in Sekunden



Quelle: HFCS Austria 2010, OeNB.

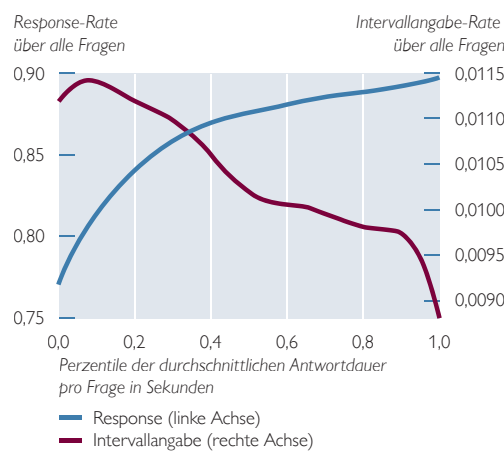
dass die Erfahrung, bereits bei der ersten Welle des HFCS Interviewer gewesen zu sein, auch Nachteile bergen kann.

Schnelle Antworten

Zu rasch gegebene Antworten mindern die Qualität der Interviewergebnisse. Dies wird an der Response-Rate ersichtlich. Je kürzer die Interviewlänge pro Frage ist, desto niedriger ist die gesamte Response-Rate über alle Fragen (siehe Grafik 4). Hinter einem solchen Muster könnte ein Verhalten des Respondenten stehen, das darauf abzielt, so schnell wie möglich das Interview zu beenden, auch um den Preis vermehrter Item-Non-Response. Dieses Verhalten wird in der Literatur *Satisficing* genannt (siehe z. B. Malhotra, 2008): In einer Entscheidungssituation wird nicht nach der optimalen Lösung gesucht, sondern die erstbeste Antwortmöglichkeit als hinreichend betrachtet. Respondenten mit kurzen Interviewlängen pro Frage weisen allerdings auch eine höhere Rate bei Intervallangaben über alle Fragen auf, was wiederum gegen *Satisficing* spricht.

Grafik 4

Korrelation¹ zwischen Beantwortungszeit und Response-Rate



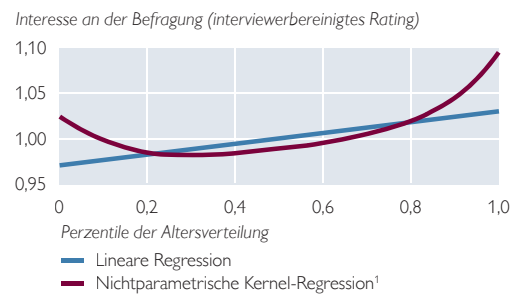
¹ Nichtparametrische Kernel-Regression: Es wird für jeden Datenpunkt eine separate gewichtete Regression geschätzt, wo der jeweilige Datenpunkt das höchste Gewicht bekommt und die weiter entfernten Datenpunkte niedrigere Gewichte.

Interesse des Respondenten an der Befragung

Auch das Ausmaß des Interesses der Respondenten kann auf Messfehler hinweisen. Die Interviewer wurden nach den Befragungen im Rahmen des HFCS Austria 2010 nach ihrer Einschätzung zum Interesse der Respondenten an der Befragung gefragt. Auffallend ist, dass diese Paradaten einen U-förmigen Verlauf beim Alter zeigen (siehe Grafik 5). Das bedeutet, dass das Interesse bei unteren und höheren Altersgruppen größer ist als bei mittleren Altersgruppen.

Grafik 5

Korrelation zwischen Alter der Referenzperson und Interesse an der Befragung



Anmerkung: Rating <1: unterdurchschnittlich hoch, Rating = 1: durchschnittlich hoch, Rating >1: überdurchschnittlich hoch.

¹ Es wird für jeden Datenpunkt eine separate gewichtete Regression geschätzt, wo der jeweilige Datenpunkt das höchste Gewicht bekommt und die weiter entfernten Datenpunkte niedrigere Gewichte.

Tabelle 4

Interesse an der Befragung nach Bildungsabschluss der Referenzperson

Höchster Bildungsabschluss	Interesse an der Befragung
	Interviewbereinigtes Rating
Pflichtschule	1,09
Berufsschule	1,02
Berufsbildende mittlere Schule	0,98
Matura	0,93
Universität	0,92

Quelle: HFCS Austria 2010, OeNB.

Anmerkung: Rating <1: unterdurchschnittlich hoch, Rating = 1: durchschnittlich hoch, Rating >1: überdurchschnittlich hoch.

Die Hypothese, dass gebildete Respondenten ein stärkeres Interesse an der Befragung haben als weniger gebildete, konnte nicht bestätigt werden. Bei Akademikern ist das Interesse am HFCS nur unterdurchschnittlich (siehe Tabelle 4). Trotzdem ist die Item-Non-Response bei dieser Bildungsgruppe unterdurchschnittlich hoch (z. B. bei den Betragsfragen zum Finanzvermögen 12 % unter Akademikern vs. 15 % unter den restlichen Respondenten (diese Werte sind nicht in Tabelle 4 enthalten).

Verwendung von Hilfsmitteln

Gerade bei Betragsfragen etwa zum Einkommen oder Vermögen ist das Heranziehen von Unterlagen (Verträge, Sparbücher usw.) zweckmäßig, um Erinnerungslücken zu schließen. Die Interviewer werden beim Interviewertraining auch dahingehend geschult, die Respondenten zu ermutigen, Unterlagen zu verwenden. Gleichzeitig gilt es aber auch, den Eindruck zu vermeiden, dass den Respondenten misstraut wird. Auch über die Verwendung von Hilfsmitteln beim HFCS Austria 2010 gibt es von den Interviewern Paradaten. Jene Personen, die Unterlagen verwenden, erzielen bessere Response-Raten als die anderen.

Tabelle 5

Verwendung von Unterlagen und Response-/Intervallangaben-Rate

Unterlagen	Response-Rate über Finanzvermögensfragen ¹	Intervallangaben-Rate über Finanzvermögensfragen ¹
	in %	
Ja, häufig	84	7
Ja, manchmal	76	12
Ja, selten	71	13
Nein, nie	62	15

Quelle: HFCS Austria 2010, OeNB.

¹ Unter Finanzvermögensfragen fallen die Betragsfragen zu stillen Unternehmensbeteiligungen, Sichteinlagen, Sparkonten, Investmentfonds, Anleihen, Aktien, treuhänderisch verwalteten Konten, Forderungen an private Haushalte, sonstigen Vermögenswerten, freiwilligen privaten Altersvorsorgeplänen.

Jene, die keine Unterlagen verwenden, nehmen dafür öfter die Möglichkeit von Intervallangaben wahr und gleichen damit potenzielle Messfehler, die durch Item-Non-Response entstehen, teilweise wieder aus.

Misstrauen gegenüber dem HFCS

Man könnte annehmen, dass Respondenten, die nur nach langer Überzeugungsarbeit durch die Interviewer zur Teilnahme an der Befragung bewegt werden können, eventuell weniger akkurate Antworten im Interview liefern. Tabelle 6 zeigt, dass die Einschätzung der Interviewer vor der Befragung zum Misstrauen gegenüber Interviewer bzw. Befragung mit der Response-Rate der Betragsfragen zum Finanzvermögen *ex-post* korrelieren. Vor dem Interview als misstrauisch eingestufte Respondenten antworten während des Interviews seltener und dies schlägt sich in niedrigeren Response-Raten nieder. Misstrauen ist demnach kein Affekt, der sich schnell zerstreuen lässt. Zumindest lässt sich aber ein Teil der durch Misstrauen potenziell verursachten Messfehler durch Intervallangaben reduzieren.

Tabelle 6

Misstrauen bei den Befragten vor der Befragung und Response-/Intervallangaben-Rate

Misstrauen vor Befragung	Response-Rate über Finanzvermögensfragen ¹	Intervallangaben-Rate über Finanzvermögensfragen ¹
	in %	
Nein, überhaupt nicht	70	13
Ja, zumindest teilweise	55	17
Ja, sehr stark	56	17

Quelle: HFCS Austria 2010, OeNB.

¹ Unter Finanzvermögensfragen fallen die Betragsfragen zu stillen Unternehmensbeteiligungen, Sichteinlagen, Sparkonten, Investmentfonds, Anleihen, Aktien, treuhänderisch verwalteten Konten, Forderungen an private Haushalte, sonstigen Vermögenswerten, freiwilligen privaten Altersvorsorgeplänen.

Einschätzung der Interviewer

Neben der Item-Non-Response kann auch die jeweilige Einschätzung des Interviewers zu den Respondenten Hinweise auf Probleme der Haushalte beim Beantworten der Fragen liefern (siehe Tabelle 7). Die Interviewer wurden nach ihrer Beurteilung zum Verständnis der Fragen bei den Respondenten, zur Zuverlässigkeit von deren Angaben und nach möglichen Problemen beim Antworten gefragt. Die Erkenntnis daraus: Was Interviewer von den Respondenten wahrnehmen, spiegelt recht gut die tatsächlichen Response-Raten wider. Dies belegt, dass diese Paradaten wichtige Möglichkeiten eröffnen, Datenqualität zu prüfen.

Beim Verständnis der Fragen sollten sich idealerweise zwischen den Respondenten keine Unterschiede zeigen. Ziel der Erhebung ist es, die Fragen so zu formulieren, dass sie auf keinem Bildungsniveau ein Problem darstellen.

Tatsächlich fanden die Respondenten die Fragen aber im unterschiedlichen Ausmaß schwierig. Jene, die nach Angaben der Interviewer generell Probleme beim Antworten hatten, wiesen auch die niedrigsten Response-Raten bei Betragsangaben zum Finanzvermögen auf. Dies belegt auch, dass die Interviewer eine realistische Einschätzung zu den Problemen der Respondenten geben konnten und dass die Interviewer im Datenerhebungsprozess eine Schlüsselrolle einnehmen.

Interviewerkommentare

Interviewerkommentare, die im Paradaten set der ersten Welle des HFCS Austria 2010 enthalten sind, fallen vielfältig aus. Sie reflektieren in ungefilterter Form, was die Interviewer von ihren Respondenten wahrnehmen und zeigen die enorme Heterogenität bei der Vermögensakkumulation. Da es keine Kriterien für die Kommentare der Interviewer gab, ist in der Auswertung erkennbar, was die Interviewer selbst für wichtig erachten. Es zeigen sich die Wahrnehmungsschemata der Interviewer. Die Interviewer erwähnen etwa einprägsame lebensgeschichtliche Episoden der Respondenten, beschreiben die Lebens- und insbesondere die Wohnsituation der Befragten. Dies hilft beim Verständnis von außergewöhnlichen Haushaltscharakteristika. Handelt es sich um statistische Ausreißer oder um spezielle Haushalte? Die Angaben sind auch ein Indikator dafür, ob beim Haushalt weitere Informationen eingeholt werden müssen (Nachrecherchen durch das durchführende Institut).

Beim HFCS Austria 2010 gab es zwei Fragen, die den Interviewern ermöglichten, freie Kommentare über jedes Interview zu geben. Da sich diese Kommentare für die erste Welle des HFCS Austria (und auch für die Interviewerschulungen für die zweite Welle)

Tabelle 7

Interviewqualität und Response-Rate

Qualitätsindikator	Response-Rate über Finanzvermögensfragen ¹	Intervallangaben-Rate über Finanz- vermögensfragen ¹
	in %	
Verständnis der Fragen		
Ausgezeichnet	68	13
Gut	63	16
Mittelmäßig	55	15
Schlecht	56	13
Zuverlässigkeit der Angaben zu Einkommen und Vermögen		
Sehr zuverlässig	70	13
Einigermaßen zuverlässig	58	18
Unzuverlässig/fehlerhaft	25	20
Probleme beim Antworten		
Sehr leicht	74	11
Relativ leicht	64	15
Teils, teils	55	19
Nicht so leicht	49	22
Überhaupt nicht leicht	25	6

Quelle: HFCS Austria 2010, OeNB.

¹ Unter Finanzvermögensfragen fallen die Betragsfragen zu stillen Unternehmensbeteiligungen, Sichteinlagen, Sparkonten, Investmentfonds, Anleihen, Aktien, treuhänderisch verwalteten Konten, Forderungen an private Haushalte, sonstigen Vermögenswerten, freiwilligen privaten Altersvorsorgeplänen.

als hilfreich erwiesen haben, werden in der zweiten Welle die Möglichkeiten für die Interviewer zur Angabe von Interviewerkommentaren erheblich erweitert werden.

Wie können Messfehler anhand von Paradaten reduziert werden?

Bei der Verminderung von Messfehlern spielen Paradaten eine wichtige Rolle. Zeitstempel, Interviewerbeobachtungen, Interviewerkommentare oder Tastenschläge und Audioaufnahmen geben oft Hinweise auf Probleme bei der Beantwortung von Fragen.

So führt Bassili (1996) vor einer Datenerhebung Pre-Tests durch und untersucht dabei die Beantwortungszeiten zu verschiedenen Fragen, um Probleme aufzuspüren und die Unsicherheit von Respondenten bei bestimmten Antworten zu messen. Ebenfalls betrachtet er dies als einen Indikator für Response-Fehler.

Paradaten können auch während der Datenerhebung für die Verbesserung des Survey-Prozesses verwendet werden. Conrad et al. (2007) verwenden Informationen über Beantwortungszeiten, um während des Interviews jene Respondenten zu identifizieren, die zu lange für die Antworten brauchen, und ihnen dann über das System Erläuterungen zu liefern. Sie finden heraus, dass dies bei diesen Respondenten die Response-Qualität verbesserte. Ähnliche Vorgehensweisen für zu schnell gegebene Antworten liefern ebenfalls positive Ergebnisse bezüglich Datenqualität (Conrad et al. 2011).

Weniger erforscht ist die Verwendung von Paradaten für die Bereinigung von Messfehlern nach der Datenerhebung. So findet Mathiowetz (1998), dass in Imputationsmodellen die Verwendung von Informationen über das Interviewerverhalten wie *Directive Probing* und über das Respondentenverhalten

wie Unsicherheit bei Antworten die Qualität der Imputationen signifikant verbessert.

In der ersten Welle des HFCS Austria 2010 wurden Paradaten hauptsächlich nach der Datenerhebung verwendet, um Messfehler zu reduzieren, vor allem im Editierungsprozess und bei den Imputationen. In der zweiten Welle konnten jedoch durch die Verwendung von Paradaten aus der ersten Welle auch schon vor der Datenerhebung Maßnahmen für die Reduktion von Messfehlern eingeleitet werden. So wurden z. B. Interviewerkommentare für die Verbesserung von Frageformulierungen, für Fallbeispiele in den Interviewerschulungen und für die Erweiterung des Glossars eingesetzt (siehe www.hfcs.at).

Schlussfolgerungen

Der HFCS Austria 2010 enthält viele Paradaten. Zur Identifikation von Messfehlern ist man auf Annäherungen über diese Paradaten angewiesen. Da keine „wahre Datenwirklichkeit“ mit den Angaben der Respondenten verglichen werden kann, ist man gezwungen, auf andere Informationen zurückzugreifen. Das Interpretieren der Paradatenergebnisse folgt fast einer hermeneutischen Methodik. Weder ist klar, wann Antworten zu rasch, noch wann zu langsam gegeben werden. Trotzdem ist die Betrachtung der Beantwortungszeiten eine wichtige Dimension in der statistischen Analyse der erhobenen Mikrodaten. In Haushaltserhebungen muss man auch den Wahrnehmungs- und Interpretations-schemata der Interviewer nachgehen und ihre Einschätzungen mit den Angaben der Respondenten vergleichen, denn die Kommunikation ist kein einseitiger Prozess. Speziell in schwierigen Fragebereichen wie beim Finanzvermögen ist dies wichtig, da dort die Untererfassung bei Erhebungen zu Haushaltsfinanzen massiv ist (siehe Andreasch und Lindner, 2014).

Paradaten werden in der zweiten Welle des HFCS 2014 an Bedeutung gewinnen. Von der Verwendung von insgesamt 13 Zeitstempeln erwarten wir uns eine qualitative Verbesserung hinsichtlich der Prüfung von besonders problematischen Fragen oder Abschnitten. Die Möglichkeiten für freie Interviewerkommentare werden erweitert. Und dies wird Inputs für ein besseres Verständnis der Haushaltsangaben und für kommende Interviewerschulungen liefern.

Literaturverzeichnis

- Albacete, N., P. Lindner, K. Wagner und S. Zottel. 2012.** Household Finance and Consumption Survey des Eurosystems 2010. Methodische Grundlagen für Österreich. Geldpolitik & Wirtschaft, Q3/12 – Addendum, OeNB.
- Albacete, N. und M. Schürz. 2013.** Interviewereffekte beim HFCS Austria 2010. Statistiken – Daten & Analysen Q3/13. OeNB. 57–68.
- Albacete, N. und M. Schürz. 2014.** Paradaten im HFCS Austria 2010 – Teil 1: Evaluierung von Non-Response-Fehlern. Statistiken – Daten & Analysen Q1/14. OeNB. 81–97.
- Andreasch, M. und P. Lindner. 2014.** Micro and Macro Data. A Comparison of the Household Finance and Consumption Survey with Financial Accounts in Austria. ECB Working Paper Series, No 1673/May 2014.
- Bassili, J. 1996.** The How and Why of Response Latency Measurement in Telephone Surveys. In: Schwarz, N. and S. Sudman (Hrsg.). Answering Questions: Methodology for Determining Cognitive and Communicative Process in Survey Research. Jossey-Bass Publishers. 319–346.
- Bricker, J. und G. V. Engelhardt. 2008.** Measurement Error in Earnings Data in the Health and Retirement Study. *Journal of Economic and Social Measurement* 33(1). 39–61.
- Bound, J., C. Brown und N. Mathiowetz. 2001.** Measurement Error in Survey Data. In: J. J. Heckman und E. Leamer (Hrsg.) *Handbook of Econometrics* 5. Chapter 59.
- Conrad, F. G., M. Schober und T. Coiner. 2007.** Bringing Features of Human Dialogue to Web Surveys. *Applied Cognitive Psychology*, 21 (2). 165–187.
- Conrad, F. G., R. Tourangeau, M. P. Couper und C. Zhang. 2011.** Interactive Interventions in Web Surveys Can Increase Response Accuracy. Paper presented at the Annual Conference of the American Association for Public Opinion Research.
- Kapteyn, A. und J. Y. Ypma. 2007.** Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. *Journal of Labor Economics* 25(3). July. 513–551.
- Kreuter, F. 2013.** Improving Surveys with Paradata: Analytic Uses of Process Information. Chapter 1: Improving Surveys with Paradata: Introduction. 1–9.
- Malhotra, N. 2008.** Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly* 72(5). 914–934.
- Mathiowetz, N. A. 1998.** Respondent Expressions of Uncertainty: Data Source for Imputation. *Public Opinion Quarterly* 62(1). 47.
- Olson, K. und I. Bilgen. 2011.** The Role of Interviewer Experience on Acquiescence. *Public Opinion Quarterly* 75(1). 99–114.
- Olson, K. und B. Parhurst. 2013.** Collecting Paradata for Measurement Error Evaluations Sociology Department. University of Nebraska 1-1-2013.
- Olson, K. und A. Peytchev. 2007.** Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes. *Public Opinion Quarterly* 71(2). 273–286.
- Yan, T. und K. Olson. 2013.** Analyzing Paradata to Investigate Measurement Error. University of Nebraska. Sociology Department Faculty Paper 232.