

Imputation der fehlenden Vorzeichen bei der Erhebung der Veränderung der tatsächlichen Wochenarbeitszeit in der Task-Zusatzerhebung 2012 (Daniela Rohrbach-Schmidt)

Ausgangslage

In der Task-Zusatzerhebung 2012 (Alda u.a. 2013) wurden im ersten Teil (CATI-Erhebung) wichtige Merkmale des aktuellen Arbeitsplatzes der befragten Erwerbstätigen ca. ein Jahr nach der Haupterhebung (BIBB-/BAUA-ETB 2012, Hall u.a. 2014) erhoben. Beide Datensätze sind als Forschungsdaten (auch SUFs) über das BIBB-FDZ zugänglich; Daten- und Methodenberichte, Fragebögen und weitere Arbeitshilfen können im Metadatensystem des BIBB-FDZ heruntergeladen werden. Im Anhang befindet sich ein Auszug des Fragebogens mit den für die Imputation relevanten Arbeitszeitvariablen.

Zu den wichtigen Variablen der Zusatzbefragung gehören neben der Lohnveränderung auch die Veränderungen in der vertraglich vereinbarten ($F11Std^1$, $F11Min$) und tatsächlichen Wochenarbeitszeit ($F12Std$, $F12Min$). Nach Abschluss der Feldphase wurde deutlich, dass zwar die Beträge, nicht aber die Vorzeichen für die veränderten Wochenarbeitszeiten erhoben wurden (vgl. Tab. 1). Es kann aus den Daten also nicht entnommen werden, ob die Arbeitszeiten um die entsprechenden Beträge zu- oder abgenommen haben. Es liegen aber für alle Befragten der Zusatzbefragung Angaben darüber vor, ob sich das Einkommen gegenüber der Haupterhebung erhöht oder verringert hat (bzw. um wieviel), die durchschnittliche Arbeitszeit zur Hauptbefragung (az), sowie Indikatorvariablen dafür, ob sich die tatsächliche Wochenarbeitszeit gegenüber der Haupterhebung geändert hat ($F12$). Für einige Befragte mit Arbeitszeitveränderung liegen außerdem zusätzlich Angaben darüber vor, wie viele Stunden sie normalerweise im Durchschnitt pro Woche in dieser Tätigkeit zum Zeitpunkt t_2 arbeiten ($F13Std$, $F13Min$). Auf Basis dieser und anderer Merkmale sollen die fehlenden Vorzeichen für die tatsächliche Wochenarbeitszeit² imputiert werden, denn die Wochenarbeitszeitangaben sind wichtige Variablen für Lohnanalysen.

Tab. 1: Übersicht über Daten zu tatsächlichen Wochenarbeitszeit in der Task-Zusatzbefragung 2012

Var.-name	$F12$	$F12Std$	$F12Min$	$F12xx$	$F13Std$	$F13Min$	$F13xx$
Frage	Hat sich Ihre tatsächliche Wochenarbeitszeit geändert?	Um wie viele Wochenstunden...?		$F12Std+$ $F12Min/60$	Und wie viele Stunden arbeiten Sie im Durchschnitt ... ¹		$F13Std+$ $F13Min/60$
Gültige Fälle ¹	Ja: 778 (18,8%)	763	764	761	69 (9% von $F12=Ja$)	n= 69 (9% von $F12=Ja$)	69 (9% von $F12=Ja$)
Mittelwert	-	9,7	2,0	9,8	47,7	0,7	47,7
Range	-	0-85	0-60	0-85	5-96	0-30	5-96

Hinweise: ¹ Die vollständige Frageformulierung lautet: „Und wie viele Stunden arbeiten Sie normalerweise im Durchschnitt pro Woche in dieser Tätigkeit – einschließlich regelmäßig geleisteter Überstunden, Mehrarbeit, Bereitschaftsdienst usw.?“

Statt der Vorzeichen soll die tatsächliche Wochenarbeitszeit für Fälle mit geänderter Wochenarbeitszeit imputiert werden. Auf Basis dieser Variable kann dann die Differenz zur durchschnittlichen Wochenarbeitszeit zum Zeitpunkt der Haupterhebung (az) gebildet werden, woraus schließlich das Vorzeichen für die angegebenen Beträge der Veränderung der tatsächlichen

¹ Die hier genannten Variablennamen (jeweils kursiv) entsprechen denjenigen in den SUFs.

² Im vorliegenden Papier wird das Verfahren zur Imputation der tatsächlichen Arbeitszeit für die Ableitung der Vorzeichen beschrieben. Für den empirisch vergleichsweise selteneren Fall einer Veränderung der vereinbarten Wochenarbeitszeit (n=394) wird keine Imputation durchgeführt. Im Unterschied zur tatsächlichen Wochenarbeitszeit liegen keine Informationen zur vertraglich vereinbarten Wochenarbeitszeit für Personen vor, die angeben, dass sich die vertraglich vereinbarte Wochenarbeitszeit verändert hat.

Wochenarbeitszeit abgeleitet werden kann. Für dieses Vorgehen spricht, dass im anderen Fall der Imputation der Wahrscheinlichkeit eines positiven oder negativen Vorzeichens mittels Maximum-Likelihood eine größere Unsicherheit im Bereich mittlerer Propensity-Scores gegeben ist. Die Imputation der kontinuierlichen Arbeitszeitvariable vermeidet dieses Problem weil die Schätzung (im Anschluss an die Selektionskontrolle) auf einem linearen Regressionsmodell basiert (s.u.). Im Gegensatz zu *F12Std/Min* (Betrag der Arbeitszeitveränderung, keine Vorzeichen) liegen für die tatsächliche Arbeitszeit gültige Fälle und relevante Prädiktoren in den Daten vor. Ein weiterer Vorteil dieser Vorgehensweise ist, dass zusätzliche Plausibilitätsprüfungen zwischen der auf der Basis der imputierten Arbeitszeitvariable berechneten Differenz und den angegebenen Beträgen möglich sind.

Gültige Werte und Ausfallmechanismus

Im ersten Schritt werden die gültigen Angaben für die durchschnittliche Wochenarbeitszeit zu t_2 (*F13Std*, *F13Min*) bzw. die rechnerische Differenz der durchschnittlichen Wochenarbeitszeit von t_2 zu t_1 (*az*) betrachtet. Hiermit lässt sich untersuchen, ob der Betrag der angegebenen veränderten tatsächlichen Arbeitszeit (*F12Std/Min*) mit dem Betrag der berechneten Differenz für alle Fälle übereinstimmt, oder ob für die Variable *F12Std/Min* zusätzlich zum fehlenden Vorzeichen der angegebene Betrag viele Fälle mit implausiblen Werten beinhaltet.

Die normalerweise im Durchschnitt pro Woche geleistete tatsächliche Wochenarbeitszeit (*F13Std*, *F13Min*) sollte für Befragte mit Arbeitszeitveränderung erhoben werden, deren ermittelte tatsächliche Wochenarbeitszeit (Summe *F12xx* und Arbeitszeit aus der Haupterhebung (*az*)) mehr als 70 Stunden umfasst. Die Abfrage von *F13Std*, *F13Min* diente dazu, die ungewöhnlich hohe Wochenarbeitszeit zu validieren³. Bei dem Abgleich der berechneten Differenz und der angegebenen Veränderung in Stunden und Minuten (Betrag) bei den 69 Fällen mit Veränderung⁴ laut *F12* fällt auf, dass in rund einem Drittel der Fälle der Wert der Veränderung der tatsächlichen Arbeitszeit (*F12Std*, *F12Min*) mit der tatsächlichen Arbeitszeit (*F13Std*, *F13Min*) übereinstimmt (siehe beispielhaft die Fälle I_1 und I_5 im Datensatzauszug in Tab. I im Anhang). In weiteren Fällen unterscheiden sich beide Werte nur um wenige Stunden bzw. Minuten. Insbesondere bei hohen Werten liegt die Vermutung nahe, dass die Befragten hier anstelle der Differenz die Wochenarbeitszeit insgesamt angegeben haben.

Fälle, für die dies zutrifft, hatten eine höhere Wahrscheinlichkeit nach der durchschnittlichen Wochenarbeitszeit gefragt zu werden, insofern, als dass sie leichter den Schwellenwert von über 70 Stunden erreichten. Ebenso steigt die Wahrscheinlichkeit hierfür mit steigendem Betrag der angegebenen Veränderung und ist höher für Erwerbstätige mit hoher Arbeitszeit zum Zeitpunkt der Haupterhebung. Ein Vergleich der Gruppen mit und ohne Angaben zur tatsächlichen Arbeitszeit zu t_2 im Hinblick auf diese und weitere Variablen zeigt, dass die beiden Gruppen keine Zufallsstichprobe der gleichen Grundgesamtheit darstellen und wir somit von einem MAR-Mechanismus ausgehen müssen. Weiter unten wird anhand eines logistischen Regressionsmodells die MAR-Annahme für den Ausfallmechanismus für diejenigen mit veränderter Wochenarbeitszeit geprüft.

Zunächst wird eine Indikatorvariable *miss* für Fälle mit Veränderung der Arbeitszeit und (fehlenden) Angaben zur Arbeitszeit zum Zeitpunkt t_2 gebildet. Die Arbeitszeit zum Zeitpunkt t_2 ist für diejenigen

³ Durch einen Filterfehler wurden unter denjenigen ohne Veränderung in der tatsächlichen Arbeitszeit zu Beginn der Feldphase alle nach der durchschnittlichen Wochenarbeitszeit gefragt. Nachdem dies aufgefallen war, wurde der Filter entsprechend korrigiert. Die 767 Fälle ohne Arbeitszeitveränderung aber mit Angabe zur durchschnittlichen Wochenarbeitszeit wurden bei der Imputation nicht berücksichtigt.

⁴ Bei denjenigen ohne Veränderung (*F12*=nein) sollte die Differenz zwischen der durchschnittlichen Arbeitszeit von t_2 zu t_1 Null sein. In 57% der Fälle weichen die Angaben zur wöchentlichen Arbeitszeit zu beiden Zeitpunkten maximal um eine Stunde nach oben oder unten ab. In den übrigen 43% der Fälle sind die Abweichungen größer als eine Stunde.

Fälle gleich der Arbeitszeit zum Zeitpunkt t_1 , die angeben, dass sich ihre Arbeitszeit gegenüber der Haupterhebung nicht geändert hat ($F12=2$). Von den insgesamt 4.356 Fällen im Datensatz der Nachbefragung (Teil 1), geben 3.359 an, dass sich ihre Arbeitszeit nicht geändert hat. 778 Fälle geben an, dass sich die tatsächliche Arbeitszeit geändert hat. 213 Fälle sind zum Zeitpunkt der Nachbefragung nicht mehr erwerbstätig (arbeitslos, in Rente, erwerbsunfähig, in Elternzeit oder Hausfrau, -mann). 6 Personen geben ‚weiß nicht‘ an oder ‚keine Angabe‘. Für die 778 Fälle mit veränderter Wochenarbeitszeit liegen für 69 Fälle Informationen zur Arbeitszeit zu t_2 vor ($miss=0$), für 709 Fälle fehlen diese Angaben ($miss=1$).

Tabelle 2 zeigt die statistischen Kennwerte der Arbeitszeitvariablen zum Zeitpunkt t_1 und t_2 für Erwerbstätige ohne und mit veränderter Wochenarbeitszeit bzw. $miss$ gleich Null und eins. Während diejenigen mit veränderter Wochenarbeitszeit und fehlenden Angaben zur Arbeitszeit zu t_2 eine leicht niedrigere mittlere Arbeitszeit zu t_1 aufweisen (37,69 Stunden mit einem Maximum von 70 Stunden) als diejenigen mit unveränderter Arbeitszeit (37,84 Stunden, maximal 99 Stunden), zeichnen sich diejenigen mit Veränderung und Angaben zu t_2 durch eine deutlich höhere Wochenarbeitszeit zum Zeitpunkt der Haupterhebung aus (53,17 Stunden).

Tab. 2: Verteilungsmaße der Arbeitszeitvariablen in der Tasks- -Zusatzbefragung 2012

	Tatsächliche Arbeitszeit t_1 (az)				Tatsächliche Arbeitszeit t_2 ($F13xx$)			
	Mittelwert	Standardabw.	Min.	Max.	Mittelwert	Standardabw.	Min.	Max.
az unverändert ($n=3.351^1/3.359$)	37,84	11,97	10	99	= t_1	= t_1	= t_1	= t_1
az verändert & $miss=0$ ($n=69$)	53,17	13,35	30	80	47,74	17,96	5	96
az verändert & $miss=1$ ($n=707^1/709$)	37,69	11,42	10	70	-	-	-	-
az verändert ($n=776^1/778$)	39,07	12,41	10	80	-	-	-	-
gesamt ($n=4.127/4.137$)	38,07	12,06	10	99	-	-	-	-

Es ist bekannt, dass das Fehlen der Arbeitszeit zu t_2 nicht zufällig und aufgrund der Filterführung abhängig von den beobachtbaren Merkmalen Arbeitszeit zu t_1 und dem Betrag der Veränderung der tatsächlichen Arbeitszeit ist. Diese Merkmale sind wiederum mit anderen Merkmalen konfundiert, die den Ausfall- (Selektions-)mechanismus determinieren. Mithilfe einer logistischen Regressionsanalyse sollen diejenigen (zusätzlichen) Kovariaten identifiziert werden, die den Ausfallmechanismus der unbeobachteten Daten determinieren.

Tabelle 3 gibt die Ergebnisse des logistischen Regressionsmodells für den Ausfall wieder, bei der die Gruppe derjenigen mit Arbeitszeitveränderung und mit Angabe zu Arbeitszeit zu t_2 die Referenzkategorie bilden (fallweiser Ausschluss). Neben den Logitkoeffizienten enthält die Tabelle für signifikante Schätzer auch die durchschnittlichen marginalen Effekte bzw. den Discrete Change (AME/DC).

Wie erwartet zeigen sich signifikante Effekte der Arbeitszeitvariablen. Darüber hinaus ist die Wahrscheinlichkeit für ein Fehlen der Arbeitszeit zu t_2 geringer bei Erwerbstätigen ohne bzw. mit Berufsausbildung als höchstem Abschluss als bei Erwerbstätigen mit (Fach-)Hochschulabschluss und

Tab. 3: Logistische Regression für das Fehlen der Arbeitszeit zu t_2 ($miss=1$) bei Erwerbstätigen mit veränderter Wochenarbeitszeit ($F12=1$)

	Logit	AME / DC
Geschlecht weiblich (Ref. männlich)	-0,14 (0,21)	
Alte Bundesländer (Ref. neue BL)	0,11 (0,11)	
Alter in Jahren ¹	-1,00 (-0,71)	
Berufserfahrung	0,05 (0,24)	
Berufserfahrung quadriert/10	-0,01 (-0,26)	
Wöchentliche Arbeitszeit zu t_1 in Std. (az)	-0,32*** (-5,84)	-0,006
Veränderung der Arbeitszeit ggüb. t_1 in Std. (Betrag, $F12xx$)	-0,29*** (-6,82)	-0,005
Höchster beruflicher Abschluss (Ref. (Fach)hochschulabschluss): Kein Abschluss	-3,97* (-2,35)	-0,073
Betriebliche/schulische Berufsausbildung	-2,23* (-2,13)	-0,041
Berufliche Stellung (Referenz: Angestellte) ² : Arbeiter	2,13 (1,67)	
Beamte	-0,84 (-0,80)	
Selbständige	1,87 (1,53)	
Freiberufler	5,58** (3,09)	-0,068
Familienstand verheiratet (Ref. unverheiratet/geschieden)	0,68 (0,92)	
Kinder unter 18 Jahre im Haushalt (Ref. keine Kinder)	-1,42 (-1,65)	
Muttersprache deutsch (Ref.: Muttersprache nicht deutsch)	0,13 (0,08)	
Höchster allgemeinbildender Schulabschluss (Ref. Ausländischer/sonstiger Abschluss) ³ : Hauptschulabschluss o.ä.	0,21 (0,10)	
Realschulabschluss/Mittlere Reife o.ä.	0,61 (0,30)	
Hochschulreife o.ä.	-0,38 (-0,20)	
Arbeitgeber- oder Arbeitsplatzwechsel (Ref. Kein Wechsel): Arbeitgeberwechsel	0,67 (0,76)	
Arbeitsplatzwechsel	0,51 (0,49)	
Veränderung Bruttoverdienst (Ref. In etwa gleich geblieben): Verbesserung	-0,58 (-0,84)	
Verschlechterung	-0,44 (-0,45)	
Konstante	39,74 (1,87)	
N	726	
R ²	.793	

Hinweise: ¹Das Modell enthält zudem $Alter^2/10$ und $Alter^3/100$ (jeweils stat. nicht sig.). ²„Freier Mitarbeiter“ (n=4) und „Mithelfende Familienangehörige“ (n=3) haben alle $miss=1$, so dass keine Effekte geschätzt werden können und diese Fälle aus der Analyse herausfallen. ³Erwerbstätige „ohne Schulabschluss“ (n=3) haben alle $miss=1$, so dass hierfür keine Effekte geschätzt werden können und diese Fälle aus der Analyse herausfallen. Die Ergebnisse unterscheiden sich nicht, wenn eine andere Bildungskategorie als Referenzkategorie verwendet wird. t-Statistik in Klammern. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

höher bei Freiberuflern als bei Angestellten⁵. Keine statistisch signifikanten Effekte zeigen sich für das Merkmal, ob die Befragten seit der Haupterhebung ihren Arbeitgeber oder Arbeitsplatz gewechselt haben und ob ihr Einkommen seit der Haupterhebung gestiegen bzw. gesunken ist anstelle gleich geblieben ist⁶, sowie für das Geschlecht, den Migrationshintergrund, den Wohnort (Ost/West), und sonstige, die Arbeitszeit relevante Merkmale (Alter, Berufserfahrung, höchster allgemeinbildender Schulabschluss, Familienstand, das Vorhandensein von Kindern im Haushalt).

Imputation

Für die Imputation wird das zweistufige Verfahren aus Alda und Rohrbach-Schmidt (2011) verwendet, bei der in der ersten Stufe eine Selektionskontrolle erfolgt und in der zweiten die Höhe der tatsächlichen Arbeitszeit zu t_2 geschätzt wird. Als Prädiktoren werden die in der Logitanalyse berücksichtigten Merkmale mit Ausnahme des Betrags der Veränderung der tatsächlichen Arbeitszeit⁷ verwendet.

Es wird dabei angenommen, dass eine gute Annäherung an die tatsächliche, aber nicht beobachtete Arbeitszeit zu t_2 die Addition eines Störwerts e_i mit der Standardabweichung σ zu dem zu einer Regressionsgleichung hervorgehenden Vorhersagewert $x_i b$ ist und sich σ aus der tatsächlichen (beobachtbaren) Verteilung der Arbeitszeitverteilung ergibt. Der Zufallswert e_i wird dabei so gezogen, dass $x_i b + e_i$ größer als (gleich) zehn Stunden ist, da es sich bei den Erwerbstätigen in der Stichprobe der Haupterhebung um Kern-erwerbstätige mit mindestens 10 Wochenstunden handelt. Zudem werden die Werte mehrfach imputiert (tausend Wiederholungen) und es wird der Mittelwert der m -fach imputierten Werte für jede Erhebungseinheit verwendet.

Im Folgenden wird die Verteilung der imputierten Arbeitszeitvariable zu t_2 mit derjenigen für nicht imputierte Werte verglichen. Tabelle 4 gibt den Mittelwert, die Standardabweichung und den Range der Ursprungsvariable und der imputierten Variable wieder.

Tab. 4: Verteilungsmaße der originalen und imputierten Arbeitszeitvariable in der Tasks- - Zusatzbefragung 2012

	Mittelwert	Standardabw.	Min.	Max.
Originale Fälle (n=69)	47,7	17,9	5	96
Imputierte Fälle (n=704 ¹)	39,8	14,4	12,1	113
Imputierte Variable gesamt (n=773 ¹)	40,5	15,0	5	113

Hinweise: Fälle mit fehlenden Angaben auf Prädiktorvariablen wurden auf Basis einer um diese Variablen reduzierten Schätzung ersetzt. Es wurde darauf verzichtet, die fünf Fälle mit fehlenden Angaben auf der Variable „berufliche Stellung“ (n=3) bzw. *az* (n=2) in einem weiter reduzierten Modell zu ersetzen.

Der Vergleich der Mittelwerte zeigt, dass durch die Imputation die durchschnittliche Arbeitszeit deutlich gesunken ist. Das Verfahren ist also in der Lage, die filterbedingt hohen beobachtbaren

⁵ Zur Vorhersage des Ausfalls wird die berufliche Stellung zum Zeitpunkt der Nachbefragung verwendet (gebildet aus *Stib* und *F1*).

⁶ Es werden nur die Indikatorvariablen und nicht das Einkommen zu t_2 als Prädiktoren für den Ausfall bzw. die Arbeitszeit zu t_2 verwendet. Erstens sollen so Endogenitätsprobleme vermieden werden (die Arbeitszeitvariable wird insbesondere für Analysen des Stundenlohns benötigt). Zweitens kann man vermuten, dass unbeobachtete Heterogenität bei der Einkommensvariable (insbesondere im Bereich von hohen Einkommensveränderungen) den linearen Schätzer des Steigungskoeffizienten für den Bruttolohn verzerrt. Die Arbeitszeit zu t_1 geht einfach sowie als Polynom 2. und 3. Ordnung mit ein, weil vermutlich Erwerbstätige mit sehr hohen und sehr niedrigen Arbeitszeiten zu t_1 eher ihre Arbeitszeiten verringern / erhöhen.

⁷ Der Betrag bestimmt denn Ausfall, es ist aber unklar, wie der Betrag mit der Arbeitszeit zu t_2 zusammenhängt (Ab- oder Zunahme). Zudem ergibt sich kein signifikanter Zusammenhang bzw. keine Verbesserung im R^2 .

Arbeitszeiten auszugleichen. Wie aufgrund des Imputationsansatzes zu erwarten, sinken die Standardabweichungen.

Abschließend werden die Ergebnisse von einfachen linearen Regressionsanalysen verglichen, bei denen die imputierten und nicht-imputierten Arbeitszeitvariable zu auf relevante Kovariaten zurückgeführt werden (vgl. Spalten 2 und 3 in Tab. II im Anhang, zu Spalte 4 weiter unten). Die Modelle der Arbeitszeit zu t_2 enthalten auch die Fälle ohne Arbeitszeitveränderung (hier Arbeitszeit t_2 = Arbeitszeit t_1). Das Variablenset ist gut in der Lage, die Varianz in den Arbeitszeitvariablen aufzuklären. In der Mehrzahl der Variablen stimmen Signifikanz, Richtung und Stärke zwischen den Modellen überein. Deutliche Unterschiede zeigen sich im Hinblick auf den Zusammenhang zwischen der Arbeitszeit und der beruflichen Stellung als Beamter einerseits und der Verschlechterung des Bruttoverdienstes gegenüber dem Zeitpunkt der Haupterhebung andererseits⁸.

Mit der imputierten Variablen ist es nun möglich, die Veränderung der Arbeitszeit als Differenz zwischen der Arbeitszeit zum Zeitpunkt der Haupt- und Nachbefragung zu bilden. Damit lässt sich im zweiten Schritt das fehlende Vorzeichen für die Originalvariable ableiten: Hierfür wird die einfache Regel angewendet, dass die Variable im Betrag der Originalvariable entspricht und ein negatives Vorzeichen erhält, wenn die errechnete Differenz aus imputierter Arbeitszeit zu t_2 zur Arbeitszeit zu t_1 negativ ist. Verteilungsmaße der originalen ($F12xx$) und neu gebildeten Variable ($F12xx_neu$) sind in Tab. 6 abgetragen. Schließlich lässt sich hiermit auch - wie ursprünglich anhand der Originaldaten vorgesehen - die Arbeitszeit zum Zeitpunkt t_2 (az_t2) als Summe der Arbeitszeit zu t_1 und der Veränderung laut $F12xx_neu$ berechnen und mit der imputierten Arbeitszeitvariable (az_t2_imp) vergleichen. Geht man so vor, dann weisen sechs Fälle auf der neu berechneten Arbeitszeit zu t_2 Werte gleich oder geringer Null auf. Eine Einzelfallprüfung ergibt, dass in diesen Fällen statt der Veränderung die tatsächliche Arbeitszeit zu t_2 angegeben wurde (siehe Datenauszug in Tab. III im Anhang). Entsprechend wurden diese Fälle auf die Beträge ($F12xx$) umgesetzt⁹. Die Verteilungsmaße der so neu gebildeten Arbeitszeitvariable (az_t2) sind auch Tab. 6 zu entnehmen. Zum Vergleich der Qualität dieser Variablen zur originalen wie imputierten Arbeitszeitvariablen siehe Spalte 4 in Tab. II im Anhang.

Tab. 6: Verteilungsmaße der originalen ($F12xx$) und korrigierten ($F12xx_neu$) Variablen zur Veränderung der Wochenarbeitszeit und der damit neu berechneten Arbeitszeit zu t_2 (az_t2)

	Mittelwert	Standardabw.	Min.	Max.
Originale Variable $F12xx$ (vgl. Tab.1., n=761)	9,8	12,0	0	85
Rekodierte Variable $F12xx_neu$ (n=761)	0,8	15,5	-85	+70
Neu berechnete Arbeitszeit zu t_2 (az_t2): $az + F12xx_neu$ (n=759 ¹) mit 6 Einzelfallkorrekturen	40,0	13,3	1,5	135

Hinweise: ¹ 2 Missing Values auf az .

Die Ergebnisse der Verteilungsmaße und die Regressionsergebnisse sprechen insgesamt für eine gute Qualität der so gebildeten Arbeitszeitvariablen für die Task-Zusatzbefragung.

Der Forschungsdatensatz der Task-Zusatzbefragung enthält zusätzlich zu den originalen Variablen die imputierten bzw. neu gebildeten (Hilfs)Variablen az_t2_imp , $F12xx_neu$, az_t2 und $indic_korr$.

⁸ Eine Lohnverschlechterung kommt deutlich häufiger bei denjenigen mit Arbeitszeitveränderung vor. Diejenigen mit Arbeitszeitwechsel sind im Modell in Spalte 3 (Originaldaten) jedoch unterrepräsentiert.

⁹ Zusätzlich zu der neu gebildeten Variablen mit den sechs rekodierten Werten enthält der Forschungsdatensatz eine Indikatorvariable für diese sechs Fälle ($indic_korr$). Weitere Rekodierungen der neu gebildeten Arbeitszeitvariable sowie der Variablen zur Veränderung der Arbeitszeit sind möglich, bleiben aber den Nutzern überlassen.

Zitierte Forschungsdaten und sonstige Quellen

Alda, Holger; Rohrbach-Schmidt, Daniela; Tiemann, Michael (2013): BIBB Task-Zusatzbefragung 2012 zur Haupterhebung der BIBB/BAuA-Erwerbstätigenbefragung 2012. suf_1.0; Forschungsdatenzentrum im BIBB (Hrsg., Datenzugang); Bonn: Bundesinstitut für Berufsbildung. doi:10.7803/610.12.1.1.10

Alda, Holger; Rohrbach-Schmidt, Daniela (2011): Imputation fehlender Werte für die Einkommensvariable in der BIBB/BAuA-Erwerbstätigenbefragung 2006. BIBB-FDZ Daten- und Methodenberichte 2/2011.

Hall, Anja; Siefer, Anke; Tiemann, Michael (2014): BIBB/BAuA-Erwerbstätigenbefragung 2012 - Arbeit und Beruf im Wandel. Erwerb und Verwertung beruflicher Qualifikationen. suf_3.0; Forschungsdatenzentrum im BIBB (Hrsg.), GESIS Köln (Datenzugang); Bonn: Bundesinstitut für Berufsbildung. doi:10.7803/501.12.1.1.30

Anhang I: Tabellenanhang

Tab. I: Datenauszug 1

	<i>az</i>	<i>F12xx:</i> F12Std+F12Min/60	<i>F13xx:</i> F13Std+F13Min/60	<i>Diff: F13xx-az</i>
Fall	Tatsächliche wöchentl. Arbeitszeit t_2	Veränderung wöch. Arbeitszeit	Tatsächliche wöchentl. Arbeitszeit t_2	Differenz tatsächliche Arbeitszeit $t_2 - t_1$
l_1	43	50	50	+7
l_2	80	12	92	+12
l_3	70	25	45	-25
l_4	70	15	55	-15
l_5	40	45	45	+5
...				

Tab. II: Modelle der imputierten, originalen und neu gebildeten Arbeitszeit-Variablen in der Tasks-Zusatzbefragung (jeweils inklusive der Fälle ohne Arbeitszeitveränderung)

	AZ t ₂ (Imputation)	AZ t ₂ (Originaldaten)	AZ t ₂ (az_t2_neu)
Alter in Jahren ¹	-0,85 (-1,67)	-0,72 (-1,37)	-0,73 (-1,49)
Berufserfahrung	0,65*** (6,35)	0,66*** (6,21)	0,57*** (5,68)
Berufserfahrung ² /100	-0,08*** (-4,40)	-0,07*** (-3,66)	-0,06*** (-3,21)
Höchster berufl. Abschluss (Ref. (Fach)hochschulab.): Kein Abschluss	-8,27*** (-9,89)	-4,91*** (-5,72)	-6,05*** (-7,45)
Betriebliche/schulische Berufsausbildung	-6,18*** (-11,09)	-3,06*** (-5,25)	-4,34*** (-8,02)
Berufliche Stellung (Ref.: Angestellte): Arbeiter	-0,15 (-0,29)	-0,17 (-0,31)	-0,26 (-0,50)
Beamte	1,04 (1,56)	2,38*** (3,52)	1,68*** (2,60)
Selbständige	7,08*** (10,14)	7,22*** (10,18)	6,92*** (10,20)
Freiberufler	9,33*** (7,93)	3,43** (2,70)	5,05*** (4,43)
Freier Mitarbeiter	-4,77 (-1,53)	-3,04 (-0,87)	-3,22 (-1,07)
Mithelfende Familienangehörige	-2,66 (-0,92)	-4,36 (-1,44)	-2,88 (-1,03)
Familienstand verheiratet (Ref. unverh./geschieden)	-1,67*** (-4,26)	-1,39*** (-3,44)	-1,47*** (-3,89)
Kinder unter 18 Jahre im Haushalt (Ref. keine Kinder)	-3,21*** (-7,21)	-3,48*** (-7,54)	-3,40*** (-7,87)
Muttersprache deutsch (Ref.: Muttersp. nicht deutsch)	0,46 (0,53)	-0,76 (-0,83)	-0,41 (-0,48)
Höchster allg.b. Schulabschluss (Ref. ausl./sonstiger): Hauptschulabschluss o.ä.	3,35 (1,15)	4,34 (1,48)	3,91 (1,38)
Realschulabschluss/Mittlere Reife o.ä.	2,57 (0,88)	3,57 (1,22)	3,18 (1,13)
Hochschulreife o.ä.	1,46 (0,50)	4,47 (1,51)	3,13 (1,10)
Kein Abschluss	4,61 (1,32)	5,06 (1,42)	6,78* (2,00)
Arbeitgeber- /Arbeitsplatzwechsel (Ref. Kein Wechsel): Arbeitgeberwechsel	0,02 (0,02)	-0,34 (-0,28)	0,75 (0,83)
Arbeitsplatzwechsel	2,93* (2,64)	2,75* (2,07)	2,82* (2,62)
Veränderung Bruttoverdienst (Ref. i.e. gleich geblieben): Verbesserung	2,73*** (6,49)	1,96*** (4,46)	2,23*** (5,45)
Verschlechterung	-3,96*** (-4,44)	0,65 (0,62)	-2,56** (-2,95)
Alte Bundesländer (Ref. neue BL)	-2,14*** (-4,43)	-2,39*** (-4,84)	-2,28*** (-4,88)
Geschlecht weiblich (Ref. männlich)	-6,52*** (-17,36)	-8,70*** (-22,47)	-7,41*** (-20,36)
Konstante	55,48*** (6,78)	53,17*** (6,27)	53,72*** (6,78)
N	4021	3338	4006

Hinweise: ¹Das Modell enthält zudem Alter²/10 und Alter³/100 (nur Alter³ schwach stat. sig. im Fall von az_t2_neu). t-Statistik in Klammern. * p < 0,05, ** p < 0,01, *** p < 0,001.

Tab. III: Datenauszug 2 (6 Fälle mit miss=1 und Werten \leq null¹ auf neu berechneter Arbeitszeitvariable zu t_2 (az_t2_neu))

	<i>az</i>	<i>az_t2_imp</i> (gerundet)	<i>diff_az_imp</i> (<i>az_t2_imp-az</i> , gerundet)	<i>F12xx</i>	<i>F12xx_neu</i>	<i>az_t2</i> (<i>az+F12xx_neu</i>), unkorrigiert ¹
l_1	20	15,5	-4,5	25	-25	-5
l_2	20	18,6	-1,4	30	-30	-10
l_3	45	36,0	-9,0	60	-60	-15
l_4	20	16,3	-3,7	24	-24	-4
l_5	40	35,8	-4,2	40	-40	0
l_6	20	14,5	-5,5	20	-20	0

Hinweise: ¹ Diese Fälle wurden in einem zweiten Schritt auf die Werte von *F12xx* umgesetzt (s.o.).

Anhang II: Fragebogenauszug Task-Zusatzbefragung 2012

Der vollständige Fragebogen kann im Metadatenportal des BIBB-FDZ heruntergeladen werden

F12	Hat sich Ihre tatsächliche Wochenarbeitszeit (einschließlich regelmäßig geleisteter Überstunden, Mehrarbeit, Bereitschaftsdienst usw.) geändert? INT: auf Nachfrage: in der Haupterhebung waren es [<az>] Stunden
	<1> ja, hat sich geändert → weiter mit F12xx <2> nein, ist unverändert → weiter mit F14 <8> weiß nicht / keine Angabe → weiter mit F13xx
F12xx	Um wie viele Wochenstunden hat sich die tatsächliche Wochenarbeitszeit verändert? INT: Es geht hierbei um die Differenz zwischen der vorherigen tatsächlichen Arbeitszeit und der aktuellen tatsächlichen Arbeitszeit. INT: volle Stunden nicht umrechnen!
wenn F12 = <1>	<input type="text"/> _____ Stunden <F12Std> <input type="text"/> _____ Minuten <F12Min> → wenn F12xx+az>70 weiter mit F13xx <98> weiß nicht <99> keine Angabe → weiter mit F14
F13xx	Und wie viele Stunden arbeiten Sie normalerweise im Durchschnitt pro Woche in dieser Tätigkeit - einschließlich regelmäßig geleisteter Überstunden, Mehrarbeit, Bereitschaftsdienst usw.? INT: volle Stunden nicht umrechnen!
wenn F12xx + Einblendung aus F12 > 70 oder wenn F12 = <8>	<input type="text"/> _____ Stunden <F13Std> <input type="text"/> _____ Minuten <F13Min> <98> weiß nicht <99> keine Angabe