

Daniela Rohrbach-Schmidt

# Imputation der fehlenden Vorzeichen bei der Erhebung der Veränderung der tatsächlichen Wochenarbeitszeit in der Task-Zusatzbefragung 2012

Version 1.0



Heft 2/2015

Daniela Rohrbach-Schmidt

# Imputation der fehlenden Vorzeichen bei der Erhebung der Veränderung der tatsächlichen Wochenarbeitszeit in der Task-Zusatzbefragung 2012

Version 1.0

► Danksagung

Ich danke Holger Alda und Christian Gerhards für hilfreiche Kommentare.

► Datenverfügbarkeit

Die in diesem Methodenbericht dokumentierten Zusatzvariablen sind in den SUF der Task-Zusatzbefragung 2012 eingearbeitet. Der Datenzugang zu dem entsprechenden Forschungsdatensatz mit der doi-Nummer 10.7803/610.12.1.1.20 erfolgt über das BIBB-FDZ.

► Hinweis zur Zählweise bei Versionsnummern des Datensatzes

Änderungen gegenüber der Vorversion des Datensatzes ohne größere inhaltliche Relevanz werden durch fortlaufende Nummern *nach* dem Punkt dokumentiert (zweite Ebene). Inhaltlich relevante Änderungen führen demgegenüber zu einer fortlaufenden Nummerierung auf der ersten Ebene.

## Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

© 2015 by Bundesinstitut für Berufsbildung, Bonn

BIBB-FDZ Daten und Methodenbericht 2/2015

### Herausgeber:

Bundesinstitut für Berufsbildung, Bonn  
Internet: [www.bibb.de](http://www.bibb.de)

### Publikationsmanagement:

AB 1.4. Dr. Thomas Vollmer/Sigrid Koppenhöfer

### Gesamtherstellung:

W. Bertelsmann Verlag GmbH & Co. KG

### Vertriebsadresse:

Bundesinstitut für Berufsbildung  
Arbeitsbereich 1.4 – Publikationsmanagement/Bibliothek  
– Veröffentlichungen –  
Robert-Schuman-Platz 3  
53175 Bonn  
E-Mail: [zentrale@bibb.de](mailto:zentrale@bibb.de)

### Kontakt FDZ:

E-Mail: [fdz@bibb.de](mailto:fdz@bibb.de)  
Tel.: +49 228 107-2041  
Fax: +49 228 107-2020

Printed in Germany

ISSN: 2190-300X



Der Inhalt dieses Werkes steht unter einer Creative-Commons-Lizenz (Lizentyp: Namensnennung – Keine kommerzielle Nutzung – Keine Bearbeitung – 4.0 Deutschland).

Weitere Informationen finden Sie im Internet auf unserer Creative-Commons-Infoseite [www.bibb.de/cc-lizenz](http://www.bibb.de/cc-lizenz).

Diese Netzpublikation wurde bei der Deutschen Nationalbibliothek angemeldet und archiviert.  
URN: [urn:nbn:de:0035-0672-9](http://nbn:de:0035-0672-9)

Internet: [www.bibb.de/veroeffentlichungen](http://www.bibb.de/veroeffentlichungen)

Downloads unter: [www.bibb-fdz.de](http://www.bibb-fdz.de)

# Inhaltsverzeichnis

1	Ausgangslage .....	5
2	Gültige Werte und Ausfallmechanismus .....	7
3	Imputation .....	11
4	Zitierte Forschungsdaten und sonstige Quellen .....	13
	Anhang I Tabellenanhang .....	14
	Anhang II Fragebogenauszug Task-Zusatzbefragung 2012 .....	17

## Tabellen

Tab. 1:	Übersicht über Daten zur tatsächlichen Wochenarbeitszeit in der Task-Zusatzbefragung 2012 .....	6
Tab. 2:	Verteilungsmaße der Arbeitszeitvariablen in der Task-Zusatzbefragung 2012 .....	8
Tab. 3:	Logistische Regression für das Fehlen der Arbeitszeit zu $t_2$ ( $miss=1$ ) bei Erwerbstätigen mit veränderter Wochenarbeitszeit ( $F12=1$ ) .....	9
Tab. 4:	Verteilungsmaße der originalen und imputierten Arbeitszeitvariable in der Task-Zusatzbefragung 2012 .....	11
Tab. 5:	Verteilungsmaße der originalen ( $F12xx$ ) und korrigierten ( $F12xx\_neu$ ) Variablen zur Veränderung der Wochenarbeitszeit und der damit neu berechneten Arbeitszeit zu $t_2$ ( $az\_t2$ ) .....	12
Tab. I:	Datenauszug 1 .....	14
Tab. II:	Modelle der imputierten, originalen und neu gebildeten Arbeitszeit-Variablen in der Task-Zusatzbefragung (jeweils inklusive der Fälle ohne Arbeitszeitveränderung) .....	14
Tab. III:	Datenauszug 2 (6 Fälle mit $miss=1$ und Werten $\leq$ null auf neu berechneter Arbeitszeitvariable zu $t_2$ ( $az\_t2\_neu$ )) .....	16

# 1 Ausgangslage

In der Task-Zusatzbefragung zur Haupterhebung der BIBB/BAuA Erwerbstätigenbefragung 2012 (ALDA u. a. 2015) wurden im ersten Teil (CATI-Erhebung) wichtige Merkmale des aktuellen Arbeitsplatzes der befragten Erwerbstätigen ca. ein Jahr nach der Haupterhebung (HALL u. a. 2014) erhoben. Beide Datensätze sind als Forschungsdaten (auch SUFs) über das BIBB-FDZ zugänglich; Daten- und Methodenberichte (ALDA u. a. 2013; HALL/ROHRBACH-SCHMIDT 2013), Fragebogen und weitere Arbeitshilfen können über das Metadatenportal 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. Tabelle 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 wie viel). Ebenfalls liegen Daten zur durchschnittlichen Arbeitszeit zur Hauptbefragung ( $az$ ) sowie Indikatorvariablen darüber, ob sich die tatsächliche Wochenarbeitszeit gegenüber der Haupterhebung geändert hat ( $F12$ ), vor. Für einige Befragte mit Arbeitszeitveränderung liegen 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 Wochenarbeitszeit<sup>2</sup> imputiert werden, denn die Wochenarbeitszeitangaben sind wichtige Variablen für Lohnanalysen.

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 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.

<sup>1</sup> Die hier genannten Variablennamen (jeweils kursiv) entsprechen denjenigen in den SUFs.

<sup>2</sup> 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.

Tabelle 1

Übersicht über Daten zur 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 ... <sup>1</sup>		$F13Std + F13Min/60$
Gültige Fälle <sup>1</sup>	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: <sup>1</sup>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.“

## 2 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 validieren.<sup>3</sup> Beim Abgleich der berechneten Differenz und der angegebenen Veränderung in Stunden und Minuten (Betrag) bei den 69 Fällen mit Veränderung<sup>4</sup> 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 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<sup>5</sup>

<sup>3</sup> 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.

<sup>4</sup> Bei denjenigen ohne Veränderung ( $F12 = \text{nein}$ ) sollte die Differenz zwischen der durchschnittlichen Arbeitszeit von  $t_2$  zu  $t_1$  null sein. In 57 Prozent der Fälle weichen die Angaben zur wöchentlichen Arbeitszeit zu beiden Zeitpunkten maximal um eine Stunde nach oben oder unten ab. In den übrigen 43 Prozent der Fälle sind die Abweichungen größer als eine Stunde.

<sup>5</sup> Die Werte entsprechen denjenigen in den Originaldaten. Im SUF (<http://dx.doi.org/doi:10.7803/610.12.1.1.20>) wurden alle 220 erwerbslosen Befragten und Auszubildenden in Vollzeit konsequent gefiltert (vgl. Alda u. a. 2013, Abschnitt 6.3).

sind zum Zeitpunkt der Nachbefragung nicht mehr erwerbstätig (arbeitslos, in Rente, erwerbsunfähig, in Elternzeit oder Hausfrau, -mann) oder Auszubildende in Vollzeit. 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).

**Tabelle 2**

Verteilungsmaße der Arbeitszeitvariablen in der Task-Zusatzbefragung 2012

	Tatsächliche Arbeitszeit $t_1$ ( $az$ )				Tatsächliche Arbeitszeit $t_2$ ( $F13xx$ )			
	Mittelwert	Standard-abw.	Min.	Max.	Mittelwert	Standard-abw.	Min.	Max.
$az$ unverändert ( $n = 3.351/3.359$ )	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 & $miss = 1$ ( $n = 707/709$ )	37,69	11,42	10	70	-	-	-	-
$az$ verändert ( $n = 776/778$ )	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 bildet (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-)Hochschulab-

schluss und höher bei Freiberuflern als bei Angestellten.<sup>6</sup> 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, anstatt gleich zu bleiben.<sup>7</sup> Dies gilt auch für das Geschlecht, den Migrationshintergrund, den Wohnort (Ost/West), die Arbeitszeit und sonstige relevante Merkmale (Alter, Berufserfahrung, höchster allgemeinbildender Schulabschluss, Familienstand, das Vorhandensein von Kindern im Haushalt).

**Tabelle 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 <sup>1</sup>	-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 gegenüber $t_1$ 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 (Ref. Angestellte):		
Arbeiter	2,13 (1,67)	

<sup>6</sup> Zur Vorhersage des Ausfalls wird die berufliche Stellung zum Zeitpunkt der Nachbefragung verwendet (gebildet aus  $Stib$  und  $F1$ ).

<sup>7</sup> 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 ein, weil vermutlich Erwerbstätige mit sehr hohen und sehr niedrigen Arbeitszeiten zu  $t_1$  eher ihre Arbeitszeiten verringern/erhöhen.

(Fortsetzung Tab. 3)

	Logit	AME/DC
Beamte	-0,84 (-0,80)	
Selbstständige	1,87 (1,53)	
Freiberufler	5,58** (3,09)	-0,068
Familienstand verheiratet (Ref. unverheiratet/geschieden)	0,68 (0,92)	
Kinder unter 18 Jahren 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) <sup>3</sup> :		
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 <sup>2</sup>	.793	

Hinweise: <sup>1</sup>Das Modell enthält zudem  $\text{Alter}^2/10$  und  $\text{Alter}^3/100$  (jeweils stat. nicht sig.). <sup>2</sup>„Freier Mitarbeiter“ (n=4) und „Mithelfende Familienangehörige“ (n=3) haben alle miss=1, sodass keine Effekte geschätzt werden können und diese Fälle aus der Analyse herausfallen. <sup>3</sup>Erwerbstätige „ohne Schulabschluss“ (n=3) haben alle miss=1, sodass 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. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## 3 Imputation

Für die Imputation wird das zweistufige Verfahren aus ALDA/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 Arbeitszeit<sup>8</sup> 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  $\sigma$  zu dem zu einer Regressionsgleichung hervorgehenden Vorhersagewert  $x_i b$  ist und sich  $\sigma$  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 Kernerwerbstätige 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 Ursprungvariable und der imputierten Variable wieder.

**Tabelle 4**

Verteilungsmaße der originalen und imputierten Arbeitszeitvariable in der Task-Zusatzbefragung 2012

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

*Hinweise:* <sup>1</sup> 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 beobachteten Arbeitszeiten auszugleichen. Wie aufgrund des Imputationsansatzes zu erwarten, sinken die Standardabweichungen.

Abschließend werden die Ergebnisse einfacher linearer Regressionsanalysen verglichen, bei denen die imputierte und nicht imputierte Arbeitszeitvariable auf relevante Kovariaten zurückgeführt werden (vgl. Spalten 2 und 3 in Tabelle II im Anhang, zu Spalte 4 weiter unten). Die Modelle der Arbeitszeit zu  $t_2$  enthalten auch die Fälle ohne Arbeitszeitveränderung (hier Arbeitszeit

<sup>8</sup> 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  $R^2$ .

$t_2 = \text{Arbeitszeit } t_1$ ). 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 andererseits.<sup>9</sup>

Mit der imputierten Variable 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 Tabelle 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 Tabelle III im Anhang). Entsprechend wurden diese Fälle auf die Beträge ( $F12xx$ ) umgesetzt.<sup>10</sup> Die Verteilungsmaße der so neu gebildeten Arbeitszeitvariable ( $az\_t2$ ) sind auch Tabelle 5 zu entnehmen. Zum Vergleich der Qualität dieser Variablen zur originalen wie imputierten Arbeitszeitvariablen siehe Spalte 4 in Tabelle II im Anhang.

**Tabelle 5**

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 $t_2$ ( $az\_t2$ ): $az + F12xx\_neu$ ( $n = 759$ ) mit 6 Einzelfallkorrekturen	40,0	13,3	1,5	135

Hinweise: <sup>1</sup> 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  $indic\_korr$ .

<sup>9</sup> Eine Lohnverschlechterung kommt deutlich häufiger bei denjenigen mit Arbeitszeitveränderung vor. Diejenigen mit Arbeitszeitwechsel sind im Modell in Spalte 3 (Originaldaten) jedoch unterrepräsentiert.

<sup>10</sup> Zusätzlich zu der neu gebildeten Variable 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.

## 4 Zitierte Forschungsdaten und sonstige Quellen

ALDA, Holger; ROHRBACH-SCHMIDT, Daniela; TIEMANN, Michael (2013): BIBB Task-Zusatzbefragung 2012 zur Haupterhebung der BIBB/BAuA-Erwerbstätigenbefragung 2012. BIBB-FDZ Daten- und Methodenbericht 4/2013.

ALDA, Holger; ROHRBACH-SCHMIDT, Daniela; TIEMANN, Michael (2015): BIBB Task-Zusatzbefragung 2012 zur Haupterhebung der BIBB/BAuA-Erwerbstätigenbefragung 2012. suf\_2.0; Forschungsdatenzentrum im BIBB (Hrsg., Datenzugang); Bonn: Bundesinstitut für Berufsbildung. doi: doi:10.7803/610.12.1.1.20.

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; ROHRBACH-SCHMIDT, Daniela (2013): BIBB/BAuA-Erwerbstätigenbefragung 2012. BIBB-FDZ Daten- und Methodenbericht 1/2013.

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

Tabelle I

Datenauszug 1

	<i>az</i>	<i>F12xx:</i> F12Std + F12Min/60	<i>F13xx:</i> F13Std + F13Min/60	<i>Differenz: F13xx – az</i>
Fall	Tatsächliche wöch. Arbeitszeit $t_1$	Veränderung wöch. Arbeitszeit	Tatsächliche wöch. Arbeitszeit $t_2$	Differenz tatsächliche Arbeitszeit $t_2 - t_1$
$I_1$	43	<b>50</b>	<b>50</b>	+7
$I_2$	80	12	92	+12
$I_3$	70	25	45	-25
$I_4$	70	15	55	-15
$I_5$	40	<b>45</b>	<b>45</b>	+5
...				

Tabelle II

Modelle der imputierten, originalen und neu gebildeten Arbeitszeit-Variablen in der Task-Zusatzbefragung (jeweils inklusive der Fälle ohne Arbeitszeitveränderung)

	AZ $t_2$ (Imputation)	AZ $t_2$ (Originaldaten)	AZ $t_2$ ( <i>az_t2_neu</i> )
Alter in Jahren <sup>1</sup>	-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 <sup>2</sup> /100	-0,08*** (-4,40)	-0,07*** (-3,66)	-0,06*** (-3,21)
Höchster berufl. Abschluss (Ref. [Fach-]Hochschulabschluss):			
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)

(Fortsetzung Tab. II)

	AZ $t_2$ (Imputation)	AZ $t_2$ (Originaldaten)	AZ $t_2$ (az_t2_neu)
Selbstständige	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. unverheiratet/geschieden)	-1,67*** (-4,26)	-1,39*** (-3,44)	-1,47*** (-3,89)
Kinder unter 18 Jahren im Haushalt (Ref. keine Kinder)	-3,21*** (-7,21)	-3,48*** (-7,54)	-3,40*** (-7,87)
Muttersprache deutsch (Ref.: Muttersprache nicht deutsch)	0,46 (0,53)	-0,76 (-0,83)	-0,41 (-0,48)
Höchster allgemeinbild. 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. in etwa 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)

(Fortsetzung Tab. II)

	AZ $t_2$ (Imputation)	AZ $t_2$ (Originaldaten)	AZ $t_2$ ( $az\_t2\_neu$ )
Geschlecht weiblich (Ref. männlich)	-6,52***	-8,70***	-7,41***
	(-17,36)	(-22,47)	(-20,36)
Konstante	55,48***	53,17***	53,72***
	(6,78)	(6,27)	(6,78)
N	4.021	3.338	4.006

Hinweise: <sup>1</sup>Das Modell enthält zudem  $Alter^2/10$  und  $Alter^3/100$  (nur  $Alter^3$  schwach statistisch signifikant im Fall von  $az\_t2\_neu$ ).  $t$ -Statistik in Klammern. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Tabelle III

Datenauszug 2 (6 Fälle mit  $miss=1$  und Werten  $\leq null^1$  auf neu berechneter Arbeitszeitvariable zu  $t_2$  ( $az\_t2\_neu$ ))

	$az$	$az\_t2\_imp$ (gerundet)	$diff\_az\_imp$ ( $az\_t2\_imp - az$ , gerundet)	$F12xx$	$F12xx\_neu$	$az\_t2$ ( $az + F12xx\_neu$ ), unkorrigiert <sup>1</sup>
$I_1$	20	15,5	-4,5	25	-25	-5
$I_2$	20	18,6	-1,4	30	-30	-10
$I_3$	45	36,0	-9,0	60	-60	-15
$I_4$	20	16,3	-3,7	24	-24	-4
$I_5$	40	35,8	-4,2	40	-40	0
$I_6$	20	14,5	-5,5	20	-20	0

Hinweise: <sup>1</sup> 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.

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





Bundesinstitut für Berufsbildung  
Robert-Schuman-Platz 3  
53175 Bonn

Telefon (0228) 107-0  
Telefax (0228) 1072976/77

Internet: [www.bibb.de](http://www.bibb.de)  
E-Mail: [zentrale@bibb.de](mailto:zentrale@bibb.de)

Bundesinstitut  
für Berufsbildung **BiBB** ▶

- ▶ Forschen
- ▶ Beraten
- ▶ Zukunft gestalten